A New Method to Support UMTS/WLAN Integration Using Stream Control Transmission Protocol

by

LI MA

B.Eng, Beijing University of Aeronautics and Astronautics, China, 1991

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

In

THE FACULTY OF GRADUATE STUDIES

Department of Electrical and Computer Engineering

We accept this thesis as conforming to the required standard

THE UNIVERISTY OF BRITISH COLUMBIA

March 2004

Library Authorization

In presenting this thesis in partial fulfillment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Li Ma
Name of Author *(please print)*

22 April 2004
Date (dd/mm/yyyy)

Title of Thesis: A New Method to Support UMTS/WLAN Integration Using Stream Control Transmission protocol

Degree: M.A.Sc

Year: 2004

Department of Electrical and Computer Engineering
The University of British Columbia
Vancouver, BC Canada

Abstract

Third generation (3G) wide area wireless networks and wireless local area networks (WLANs) possess complementary characteristics. The performance and flexibility of wireless services would be dramatically improved if users could seamlessly roam across these two types of networks. In recent years, Stream Control Transmission Protocol (SCTP) developed by the Internet Engineering Task Force (IETF) has gained significant popularity in the studying of next generation IP networks. SCTP has also been selected as a standard signaling protocol for service control in 3G wireless network. Its multi-streaming, multi-homing and partial reliable (PR) data transferring features are especially attractive for applications that have stringent performance and high reliability requirements.

In this research, we propose a new method to facilitate seamless vertical handovers (VHOs) between 3G Universal Mobile Telecommunication System (UMTS) cellular networks and WLANs using SCTP. The multi-homing capability and dynamic address reconfiguration (DAR) extension of SCTP are applied in the UMTS/WLAN overlay architecture to decrease the VHO delay and improve the throughput performance. We develop the UMTS/WLAN bi-directional VHO procedures and study the different scenarios employing a single-homing and a dual-homing fixed server configurations to support the VHO. We compare the performance of these two possible configurations using the ns-2 simulation tool and recommend the better one for our further study. Unlike mobility solutions of network or application layer, SCTP-based vertical handover does not require additional components to be added into the existing networks. Therefore, the proposed scheme provides a network independent solution that is preferred by service providers.

In addition, we propose a new scheme that we call *Sending-buffer Multicast with Fast Retransmission* (SMART-FRX) to allow the sender to enter into the slow start process when *handover loss* (HL) occurs in the WLAN link, and retransmit the *error loss* (EL) caused by multi-path fading over the wireless channel to the same destination IP address. According to the current SCTP link failure detection and recovery process as described in the specification, during a link failure detection period, timeouts and backoffs on the primary link may result in the poor throughput of the whole system. The proposed SMART-FRX scheme multicasts the buffered and new data on both UMTS and WLAN links to subdue the effect of *handover loss* (HL), and avoid unnecessary long delays of retransmitting lost packets due to UMTS transmission errors over the alternate possibly unreachable WLAN destination address. Consequently, the throughput performance is increased significantly.

Moreover, we develop a new analytical model to study the SCTP performance during the VHO. By comparing numerical results for the analytical model with simulation results, we demonstrate that our model is able to accurately predict SCTP throughput. The analytical model provides a useful tool to estimate the SCTP throughput performance.

Table of Contents

List of Tables

List of Figures

Acknowledgements

First and foremost, I would like to express my sincerest gratitude to my advisor Professor Victor Leung, without whose guidance, encouragement and support, it would have been impossible to bring my research work to its logical conclusion. His motivation and drive are the primary reason for arousing my interest in this research. I would also thank him for giving me this unique opportunity to learn and enjoy working in this research. I would like to gratefully acknowledge the support of Telus Mobility, the Advanced Systems Institute of BC and the Canadian Natural Sciences and Engineering Research Council, who support this work under grant CRD247855-01.

I would like to thank Professor Vikram Krishnamurthy and Professor Vincent Wong for serving on my thesis committee. Their comments have enhanced my research in innumerable ways.

I want to express my deep appreciation to Dr. Fei Yu and Dr. Tejinder Randhawa for their insightful suggestions and precious time spending with me to discuss some topics related to this project. Their help motivated and inspired me in many ways. I would like to thank fellow students Chu Zhang and Kassim Olawale for their comments and suggestions on editing of the thesis. Their help, support and encouragement have certainly made this research work memorable.

Finally, I would like to express my special thanks to my husband Henry for his encouragement, love and consistent support during my studies in UBC.

Chapter 1 Introduction

The third generation (3G) cellular networks, e.g., Universal Mobile Telecommunication System (UMTS) networks [1] provide wide area coverage at high mobility and support low to medium data rates. On the other hand, the IEEE 802.11 Wireless Local Area Networks (WLANs) [2] support higher speed data transmissions, but cover only small areas and support limited mobility. The complementary characteristics of UMTS networks and WLANs make it attractive to integrate these two technologies. In this chapter, we describe UMTS/WLAN vertical handover (VHO) in Section 1.1. Sections 1.2 to 1.4 present the motivations, contributions, and organization of the thesis, respectively.

1.1 UMTS/WLAN Vertical Handover

One of the main issues in UMTS/WLAN integration is handover. A handover is a process that occurs when a terminal changes the base station through which it is communicating. There are two types of handovers, the horizontal handover (HHO) and the vertical handover (VHO) [3]. An HHO occurs when a mobile user moves between cells of the same type in terms of coverage, data rates and mobility management mechanisms, such as UMTS to UMTS, or WLAN to WLAN. A VHO occurs when a mobile user moves between cells of different types, such as UMTS to WLAN or vice versa. An efficient UMTS/WLAN handover scheme is crucial to the integration of these two technologies.

The aim of supporting VHO between UMTS and WLAN is to realize real-time multiple network access over these two heterogeneous networks. This enables a user to utilize both WLAN and UMTS networks in parallel. The VHO allows the application services to be seamlessly transferred between different networks. To the end user, this means that devices that have the network access capability over both networks can migrate between the two networks without the need to reboot or restart the devices, and network applications can continue to run during migrations. The ability to use networks in parallel gives the user or service provider the possibility to choose the most economical connection at a specific location.

1.2 Motivations

Existing solutions for mobility management supporting an integration of UMTS and WLAN networks include network layer Mobile IP (MIP) [4] and application layer, e.g. Session Initiation Protocol (SIP)-based techniques [5]. MIP uses a home agent (HA) and a foreign agent (FA) to bind the home address of a mobile host (MH) to a care-of-address at the visited network and provides mobility-transparent packet forwarding when the MH moves between IP subnets. However, MIP suffers from the problems of triangular routing, high handoff latencies and large overheads of tunneling IP packets. Compared with MIP, SIP-based mobility offers attractive benefits for multimedia applications. However, some inherent problems with this approach make the adoption of this scheme difficult [6]. For example, as an application layer solution, it is not likely to be able to overcome the handover latency caused by the lower layers. Thus, it requires the support of a lower layer mobility protocol, e.g., MIP. Another issue is the interoperability of SIP and MIP. The HA and FA registration processes serve the same purpose as the SIP REGISTRATION function. The joint deployment of SIP and MIP becomes problematic.

As MIP and SIP-based approaches have some unappealing characteristics, such as limited performance and additional complexity for the network architecture, people have proposed to solve the mobility problem in the transport layer. The rationale of proposing transport layer mobility is that mobility is an end-to-end issue instead of network component or router issue. An end-to-end problem should be solved in the end systems according to the end-to-end principle [7]. Transport layer mobility leaves the network untouched and still allows roaming between networks. This makes the entire networking architecture simpler by working without additional entities added into the network infrastructure, and without new software states such as Mobile IP's tunnel states introduced into the network. Additionally, with mobility supported by the transport layer, flow and congestion control parameters in the transport layer can be easily and quickly adapted to the new network during and after a VHO.

User Datagram Protocol (UDP) or Transmission Control Protocol (TCP) supports the most popular suite of applications on the Internet today. However, for a long time, the existing UDP and TCP have been found to be inappropriate to support rapidly changing requirements introduced by today's IP networks [8][9]. UDP cannot support reliable data transmissions. For TCP, any time an endpoint's IP address becomes inaccessible due to interface failure, radio channel interference, or an MH moving out of range of its base station, the TCP's connection will timeout, abort and result in a recovery process initiated by the application. This recovery overhead and session delay are usually quite high and are unacceptable in a mobile communication environment.

The limitations of UDP and TCP have motivated the IETF to develop a third transport layer protocol, known as the Stream Control Transmission Protocol (SCTP) [10][11]. Designed with fault tolerance in mind, SCTP is a reliable network-friendly protocol that could co-exist with UDP and TCP in a network. In addition to standard TCP features, SCTP incorporates a number of enhanced features that make it suitable for transferring data over mobile networks and providing mobility support. Firstly, SCTP is the lowest layer to support end-to-end services, which agrees with the end-to-end principle. Secondly, compared with MIP- and SIP-based mobility, an SCTP-based mobility approach has the following advantages:

- No third party other than the end-points participates in the handover.
- It can support concurrent usage of any type of access routers.
- Additional network components or modifications of intermediate routers are not required.

1.3 Contributions

SCTP is a relatively new protocol. Recent research on SCTP over wireless networks mainly aims at exploiting SCTP's current capabilities, or designing new features that can make SCTP well suited for wireless channel characteristics. However, so far there are still no large-scale experimental results available to verify if the current SCTP standard meets the high reliability requirements of the mobility management. Our research in this thesis focuses on the study of SCTP supporting heterogeneous wireless networks integration. The main contributions of this thesis are as follows:

- **Using mobile SCTP (MSCTP) [12][13] to support UMTS/WLAN VHOs [14]**: We propose a new method to support UMTS/WLAN VHOs using MSCTP. Though the latest work on SCTP-based mobility [12][13] proposed a single-homing asymmetric and a dual-homing symmetric configurations to support MSCTP-based handovers, implementation details are lacking. Besides, the existing work on MSCTP is mainly concerned with supporting of general handover managements. To apply the MSCTP technology for seamless UMTS/WLAN VHOs, neither detailed descriptions nor prototype implementations have been presented in the literature. In this thesis, we clarify key concepts on the architecture and develop procedures for the VHO support. We develop an implementation of MSCTP using Network Simulator ns-2 [15][16]. In order to realize efficient UMTS/WLAN bi-directional VHO procedures, we propose to use SCTP bundling and unbundling message technologies to simplify the handover procedures. We perform tests to compare the two possible configurations, the single-homing and the dual-homing configurations, and then recommend the one that has the better delay and throughput performance for further study.
- **Sending-buffer Multicast with Fast Retransmission (SMART-FRX)**: A number of problems may arise from wireless communications during VHOs in a WLAN and UMTS overlay environment, one of which is the timeout and the retransmission caused by sudden long delays because of the mobile user moving out of the coverage of a WLAN. This results in SCTP backoffs and poor throughput performance. We show that two types of packet loss can dominate the SCTP performance during a

WLAN to UMTS forced VHO. One is the consecutive packet dropping on the WLAN link because of the route disconnection during the VHO, so called handover loss (HL); the other one is the high error rate over the wireless channel due to multi-path fading when data are transmitted on the UMTS link, so called error loss (EL). To deal with these two types of packet losses, we propose the SMART-FRX scheme. SMART-FRX is actually composed of the following two sub-schemes that perform different functionalities to assist the WLAN to UMTS forced VHO - *Sending-buffer Multicast-Aided Retransmission* (SMART) sub-scheme and *Fast Retransmission* (FRX) sub-scheme:

1) The SMART sub-scheme multicasts the data buffered in the primary WLAN link on both the UMTS and WLAN links, so that the UMTS link can enter into slow start data transmissions at the beginning of the VHO period instead of at the end of this period.

2) The FRX sub-scheme sends all retransmissions caused by EL to the UMTS destination IP address during the VHO. Thus, unnecessarily long delays of EL retransmissions to the possibly unreachable WLAN destination IP address can be avoided.

- **Modeling SCTP throughput:** Due to the lack of research in analytical modeling of SCTP, in order to study the behavior of SCTP, some best-known proposed analytical models on TCP such as [17] and [18] can be applied. The work [19], which is the only work we can find on analytical modeling of SCTP, proposed to extend the TCP analytic model in [17] to study the data transfer performance in SCTP associations. However, simulation results show that the extended model generally provides a pessimistic estimate of the throughput exhibited in experiments. In order to provide a more reliable and accurate analytical model to predict the SCTP performance, we propose a new analytical model that takes into account the dynamic changes of the congestion window (cwnd), the Round Trip Time (RTT), both the slow start and congestion avoidance processes, the network propagation delay and other factors that may affect the SCTP performance. By validating the analytical model using simulation results based on a wireless random loss model, we show that the theoretical performance is able to match the simulation performance accurately. It provides a useful tool to study the SCTP performance during the VHO.

1.4 Organization of Thesis

The rest of thesis is organized as follows: Chapter 2 presents the related work including an overview of the UMTS/WLAN integration, current multi-homing schemes to support mobility and an introduction of SCTP and MSCTP. Chapter 3 describes an abstract address-handling model, the architecture and the procedures of supporting UMTS/WLAN VHOs using MSCTP. Chapter 4 presents the SMART-FRX scheme to improve the WLAN to UMTS forced handover performance and the proposed SCTP analytical model. In Chapter 5 we describe the simulation methodology, present the simulation results for the proposed schemes and compare with analytical results. Finally, Chapter 6 concludes the thesis with a summary of the presented work and suggests future work.

Chapter 2 Literature Review

In this chapter, Section 2.1 gives an overview of related work on the UMTS/WLAN integration. Section 2.2 gives a brief description on the concept of multi-homing and reviews the related work including a number of existing proposed multi-homing approaches to address the end-host mobility issues. Finally, SCTP, MSCTP and their development status are reviewed in Section 2.3.

2.1 UMTS/WLAN Integration

In this section, we first give an overview of the UMTS/WLAN integration in Section 2.1.1. Then the related work on the UMTS/WLAN integration schemes is reviewed in Section 2.2.2.

2.1.1 Overview

With the recent successful deployment of WLAN in numerous hotspots, WLAN technology will play a key role in future wireless data transmissions. Cellular network operators have recognized this fact and strived to exploit the WLAN technology by integrating this technology into the cellular data networks. For this reason, UMTS/WLAN integration becomes an important research topic for the next generation IP-based wireless networks. An UMTS/WLAN integrated network combines the strengths of both UMTS and WLAN technologies and results in a wide-area system that is capable of providing users with ubiquitous service ranging from low to high data transmission speed in strategic locations. Table 2-1 shows the complementary

Table 2-1 Complementary characteristics of the UMTS and WLAN technologies

	UMTS	**WLAN 802.11**
Throughput	384 kbps and up to 2 Mbps	2 Mbps and up to 54 Mbps
Typical cell range	35 km	50 m
Applications	Full mobility, wide geographical availability, public networks, small devices, support data and voice technologies.	Nomadic usage, limited mobility, small geographical availability, public or private networks, laptops, support data transfer technology.

characteristics of the UMTS and WLAN technologies.

There are two generic approaches to the design of an integrated UMTS/WLAN network architecture as specified by the European Telecommunications Standards Institute (ETSI): so called tight coupling and loose coupling inter-working architectures [20]. Fig. 2-1 shows the generalized architecture for UMTS/WLAN integration. In the tight coupling inter-working architecture, the WLAN is connected to the UMTS core network in the same manner as other UMTS radio access networks. Serving General Packet Radio Service (GPRS) Support Node (SGSN) and Gateway GPRS Support Node GGSN need to be upgraded to be able to handle the much higher bit rate supported by the WLAN. The main advantage of this solution is that the mechanisms for mobility, QoS and security in the UMTS core network can be reused directly. However, tightly coupled solutions will be highly specific to the UMTS technology and cause a larger impact in the

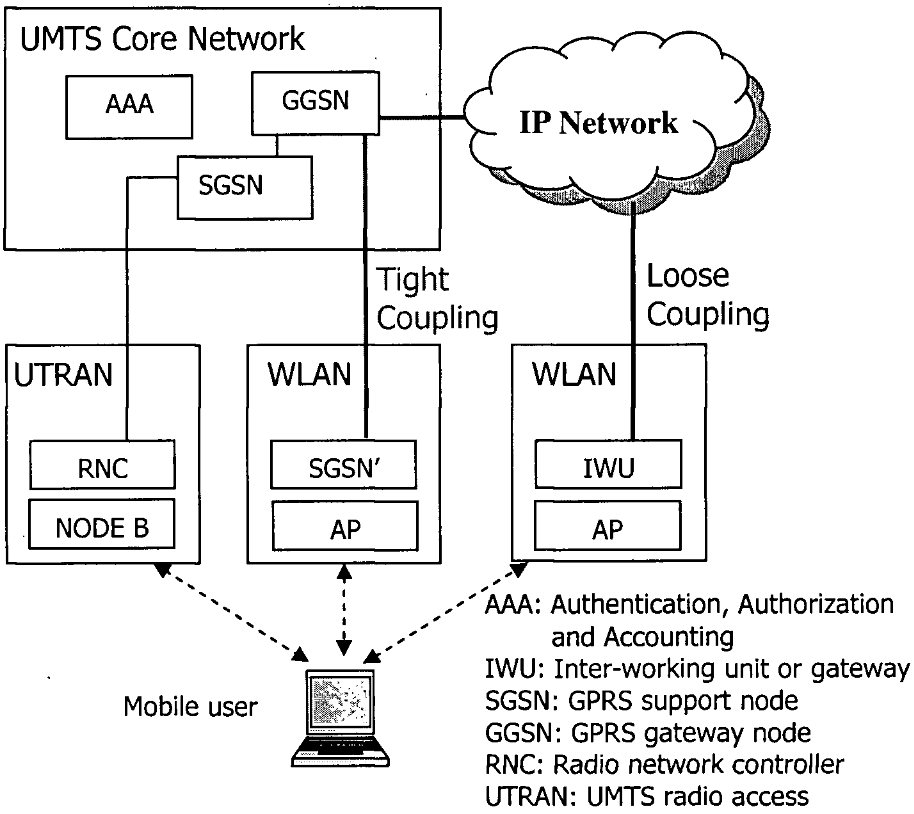


Fig. 2-1 Generalized architecture for UMTS/WLAN integration

need of extensive access interface standardization. Unlike the tight coupling approach, a loose coupling inter-working architecture introduces a new element called Inter-working Unit (IWU) or Gateway in the WLAN, so that the WLAN can be used as a packet-based access network complementary to the UMTS network. The UMTS network and the WLAN share the same Authentication, Authorization and Accounting (AAA) subscriber database for functions such as security, billing and customer management, but the two networks will be peer IP domains. This scheme avoids any impact on the SGSN and the GGSN and allows many network operators and service providers to operate through roaming agreements between the networks.

2.1.2 Related Work on UMTS/WLAN Integration

Varma *et al.* proposed solutions using MIP or SIP for mobility management in integrated GPRS and WLAN [21]. However, they mainly studied the tight coupling architecture. For the loose coupling architecture, they address only location management issues instead of seamless heterogeneous handovers and left handover issues as future work. Although Buddhikot *el al.* focused on the loose coupling architecture [22], they introduced a complicated new network element called the inter-working access gateway (IOTA) in the WLAN, and new service access software on the client devices to integrate the WLAN with the UMTS network. Mobile-IP agent functionality is implemented in the IOTA system to support MIP technology. Tsao *et al.* proposed MIP, Gateway and Emulator approaches for UMTS/WLAN inter-working according to different deployment scenarios [23]. Most of these papers point out that tight coupling approach may achieve a better performance in terms of handover latency. However, the tight coupling approach lacks flexibility. The loose coupling architecture allows the independent deployment and traffic engineering of WLAN and UMTS, and therefore offers several architectural advantages over the tight coupling approach. Moreover, in existing solutions for mobility management that employ either MIP- or SIP-based techniques, changing the IP address could result in the loss of connectivity for upper layers due to long delays. Supporting concurrent usage of access points (APs) or base stations (BSs) to maintain the ongoing session is a challenging issue. It results in additional complexities in GGSN as well as IWU that requires further study.

2.2 Related Work on Multi-homing and Transport Layer Protocol Supporting Mobility

In the IP terminology, a communication host or endpoint is called multi-homed if it can be addressed by (and thus "owns") multiple IP addresses. Multi-homing refers to a situation where an end-point has several parallel communication paths (or links) that it can use. Usually multi-homing is a result of either the host having several network interfaces or a single network interface with multiple IP addresses assigned. The concept of multi-homing allows more than one paths to coexist between different endpoints and concurrent usage of different APs (or BSs) when a mobile user is moving between different cells. Thus, it opens new horizons to people to develop new solutions to support inter-network mobility.

In this section, we review a number of existing network and transport layer proposed approaches, other than MSCTP, to address issues of multi-homing supporting end-host mobility. This includes Location Independent Network for IPv6 (LIN6) in Section 2.2.1 and Host Identity Protocol (HIP) in Section 2.2.2. As a part of related work of transport layer protocol supporting mobility, though Migrate-TCP (M-TCP) did not use the concept of multi-homing, it was one of the most important approaches of transport layer protocol supporting mobility before the use of SCTP was proposed. We introduce this technique in Section 2.2.3.

2.2.1 Location Independent Network for IPv6 (LIN6)

LIN6 [24] provides an IP layer solution for multi-homing supporting mobility by using LIN6 generalized identities (GIs). The scheme proposed a Location Independent Network Architecture (LINA) that is based on the idea of separating the identifier and the locator of a node. The network layer is divided into two sub-layers: the identification sub-layer and the delivery sub-layer. LIN6 is a protocol that supports mobility by applying LINA to IPv6. For practical purposes, LIN6 is carefully designed to maintain compatibility with conventional IPv6 so that there is a minimal impact on the existing IPv6 infrastructure. A host can be uniquely named by its LIN6 identity (node identifier) with the LIN6 prefix (location identifier), resulting in what is called a LIN6 GI. A Mapping Agent (MA) performs the mapping between a conventional IPv6 address to an LIN6 GI. LIN6 supports multi-homing by allowing a single GI, which acts as a fixed server (FS) or a corresponding host (CH) to be associated with several GIs that can be used by a mobile host (MH). Fig. 2-2 shows the mechanism of multi-homing to support a mobile client (MC)'s handover. In this figure, the steps of LIN6 multi-homing supporting of handover are as follows:

1) An MC whose node identifier is ID-2 issues a Mapping Update Request to MA so that it can get a new location identifier prefix-B.

2) The MC sends a Mapping Refresh Request to FS or CH whose node identifier is ID-1 to notify the changing of the location identifier.

3) The FS or CH sends a Query to MA and MA replies ID-1 with the new location identifier prefix-B of ID-2.

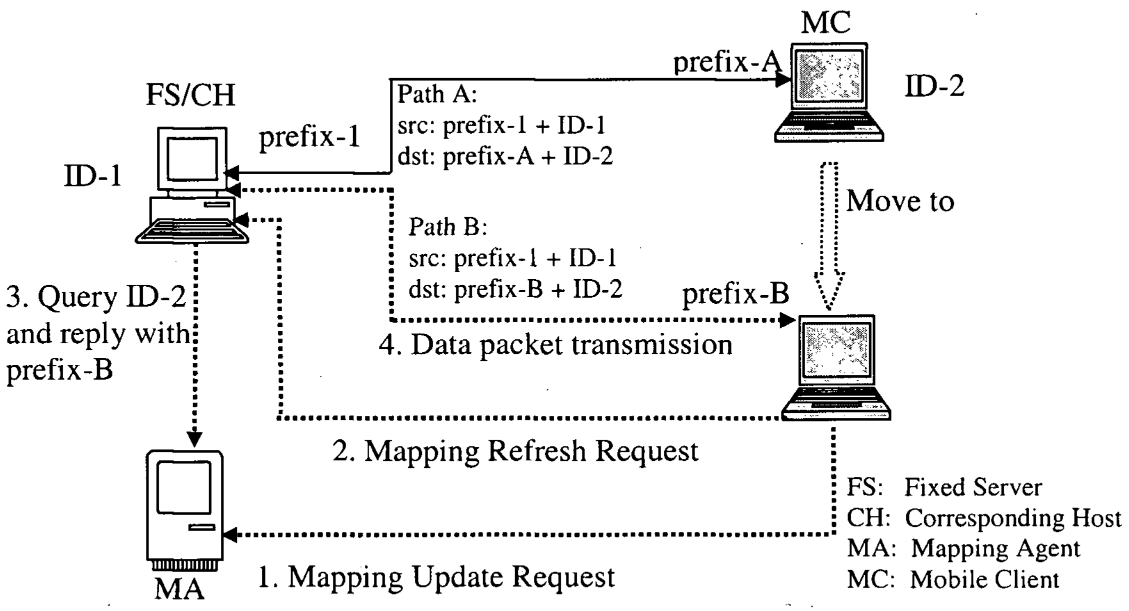


Fig. 2-2 LIN6 multi-homing supporting of handover

4) The FS or CH starts to transmit data packet on the new path: prefix-B+ID-2.

The disadvantage of this approach is that, it is a proprietary solution; e.g., to use this approach, an IPv6 address is translated to a GI address, but the format of the GI is defined by individual service operator and there is no open standard so far. The second disadvantage of this scheme is, it splits the network layer to identification sub-layer and delivery sub-layer and introduces an additional component, Mapping Agent, into the network for the mobility and address management. Thus, there is a relatively high implementation complexity. The development of this technology is still in progress by Keio University, University of Tokyo, and Sony Computer Science Laboratory.

2.2.2 Host Identity Protocol (HIP)

The HIP [25] architecture provides simple and efficient means for endpoints to communicate while being multi-homed, mobile, or simultaneously mobile and multi-homed. HIP defines a new layer called Host Identity Layer (HIL) that decouples the transport layer from the network layer, and introduces a new host identity namespace. The transport layer sockets are no longer named with IP addresses but with separate host identifiers (HIs). The host identity layer translates the HI into IP address by mapping between HI and IP addresses. The mappings from HIs to IP addresses can be extended from a static one-to-one into a dynamic one-to-many mapping. In order to support mobility, the multi-homing capability enables a moving end-host to be in a simultaneous one-to-many relationship. A Forwarding Agent is used as a rendezvous server that provides MIP Home Agent like functionality for HIP enabled mobility. HIP specifies a mechanism that allows a host to update its address(es) to its peers. The address update is implemented with a new HIP packet type, the HIP Readdress (REA) packet. Due to the danger of flooding attacks [26], the peer must always check the reachability of the node before it can use a new address. The reachability check is implemented with a pair of new HIP packet types, HIP Address Check (AC) and HIP Address Check Reply (ACR). In addition to initiating and reachability checking, an AC packet may also act as an acknowledgement for a preceding REA packet.

The HIP approach adds a new layer, HIL, between IP and transport layers and proposes to solve multi-homing and mobility problems by introducing Forward Agents into the IP networks, which act as Rendezvous servers and perform MIP HA functionality. This results in a high implementation complexity on the current IP network. HIP is developed by Ericsson Research Nomadic Laboratory in July 2003 and the research currently is in progress.

2.2.3 Current Transport Layer Protocol Supporting Mobility

In early research of transport layer protocol supporting mobility, there was no multi-homing concept involved. One of the most important proposals is a TCP extension scheme called Migrate TCP (M-TCP), which was proposed by Snoeren and Balakrishnan [27]. M-TCP allows the TCP end-points to migrate from one IP address to another based on dynamic Domain Name System (DNS) updates. In M-TCP, whenever a mobile client moves to a new network, it obtains a new IP address and updates the DNS mapping for its host name. The separation of identity and location is achieved by using the host name and the IP address. In TCP, changing the mobile host's address would disrupt any ongoing TCP connection whenever a handover occurs. This disruption may result in the closing of the existing TCP connection. To overcome this problem, M-TCP is designed to support the secure migration of an established TCP connection across an IP address change. The basic idea of M-TCP is, a TCP peer can suspend an open connection and reactivate it from another IP address, which is transparent to an application that expects uninterrupted reliable communication with the peer. In a TCP connection migration process, the mobile host activates a previously established TCP connection from a new address by sending a special Migrate SYN packet that contains a token identifying this as the old connection. The FS will then re-synchronize the connection with the mobile client using the new address. The consumed time of a TCP connection migration process is called TCP migration delay. The main advantage of this approach is that it has no need for a third party (HA or FA) for handover.

However, M-TCP does not have multi-homing function. On the other hand, the M-TCP architecture may incurs similar or even worse handover delays compared with

MIP because it relies on dynamic DNS for initiation of a connection. In a later paper [28], the work is continued towards a more generic mobile session layer concept. The authors concentrated on dealing with long-lasting but transient outages in connections, and suggested that in addition to solving the initial connection and mobility tracking problems, proper mobility architectures should also address the problems of graceful disconnections, recoveries from peer hibernations and the need of fast reconnections. These issues have become the motivations that led researchers turn to study SCTP supporting mobility, because these problems can be solved by the SCTP multi-homing, association graceful shutdown and failure detection and recovery functions.

2.3 SCTP and MSCTP

In this section, we introduce the protocol of SCTP in Section 2.3.1. We review MSCTP and its related work in Section 2.3.2.

2.3.1 SCTP Overview

SCTP was originally developed to carry telephony signaling messages over IP networks [29]. Due to its attractive features such as multi-homing and multi-streaming, SCTP has received much attention from the research community and industry. Until today, SCTP has gradually evolved into a general purpose transport protocol and may eventually replace TCP and perhaps also UDP in the future. Fig. 2-3 shows where SCTP fits in the IP architecture. Today SCTP has more than twenty-six implementations in over a dozen operating systems. A total of six conformance testing groups are working to detect and correct implementation inconsistencies in order to make SCTP ready for

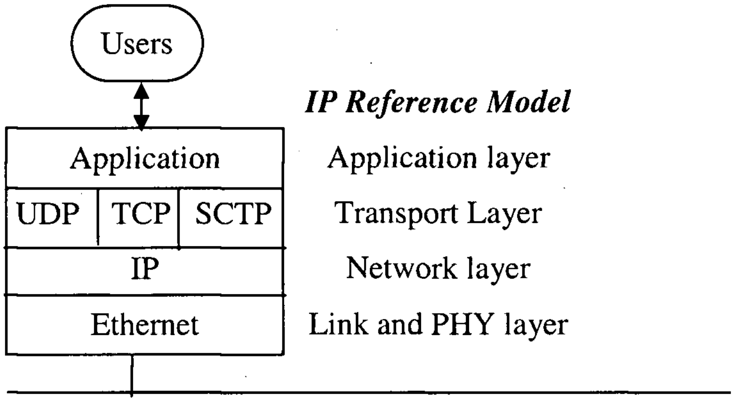


Fig. 2-3 SCTP in the IP reference model

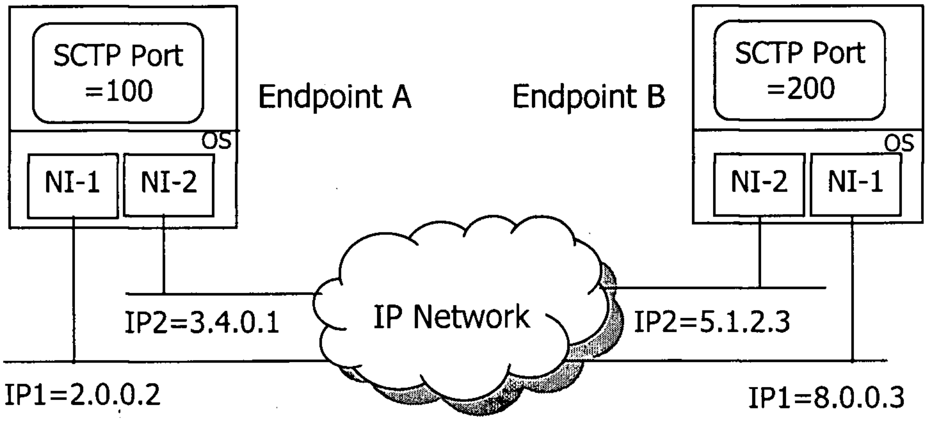
deployment. Table 2-2 summarizes the services and features of SCTP in comparison with TCP and UDP.

Like TCP, SCTP provides a reliable full-duplex connection called association between endpoints, and employs mechanisms to control congestion in the network. SCTP is connection-oriented in nature, but the SCTP association is a broader concept than the TCP connection. SCTP provides the means for each SCTP endpoint to provide the other endpoint (during association startup) with a list of transport addresses (i.e., multiple IP addresses in combination with an SCTP port) through which that endpoint can be reached and from which it will originate SCTP packets. The association spans over all of the possible source/destination combinations that may be generated from each endpoint's lists. Unlike TCP or UDP, SCTP offers an advanced delivery option – the partially reliable data transfer (PR) [30]. PR lets a user specify a reliability level on a per-message basis. The reliability level defines how persistent an SCTP sender should be in attempting to communicate a message to the receiver - for example, never retransmit, retransmit up to a certain times, retransmit until lifetime expires, or retransmit until the association aborts. SCTP PR offers the flexibility to provide intermediate reliability levels, in

Table 2-2 Comparison between SCTP, TCP and UDP [29]

Services/Features	**SCTP**	**TCP**	**UDP**
Connection-oriented	Yes	Yes	No
Full duplex	Yes	Yes	Yes
Reliable data transfer	Yes	Yes	No
Partial-reliable data transfer	Optional	No	No
Unordered data delivery	Yes	No	Yes
Flow control	Yes	Yes	No
Congestion control	Yes	Yes	No
Explicit Congestion Notification (ECN) capable	Yes	Yes	No
Selective Acknowledgement (SACK)	Yes	Optional	No
Path Maximum Transmission Unit (PMTU) discovery	Yes	Yes	No
Fragmentation of user messages	Yes	Yes	No
Bundling and unbundling of user messages	Yes	Yes	No
Multi-streaming	Yes	No	No
Multi-homing	Yes	No	No
Reachability checking	Yes	No	No
Failure detection and recovery	Yes	No	No

addition to the two extremes that UDP and TCP currently provide. Therefore, it is particularly desirable for real-time telephony signaling and multimedia applications. The benefits of this option can be observed when real-time traffic is transferred during periods of poor quality of service, heavy congestion, and path failures within the network. SCTP's four-way handshake for association establishment makes it resistant to blind



NI: Network Interface
OS: Operating System

Fig. 2-4 SCTP with multi-homing configuration

denial-of-service attacks, thus increasing overall protocol security.

A TCP connection allows for only one IP address at each endpoint. However, an SCTP endpoint is identified by an SCTP transport address, which is composed of multiple IP addresses and an SCTP port number, i.e., an SCTP endpoint *A* can be expressed as,

endpoint $A = [\ IP1, IP2, \ldots IPn : SCTP\ port\ number]$.

Therefore, a single SCTP association allows multiple connections coexisting between two endpoints, so called SCTP multi-homing. Such multi-homing provides redundancy at the path level, thus increasing association survivability in the case of a network path failure. Fig. 2-4 shows an example of the network level fault tolerant SCTP multi-homing configuration. In this set up, there are four possible paths between endpoint *A* and endpoint *B*.

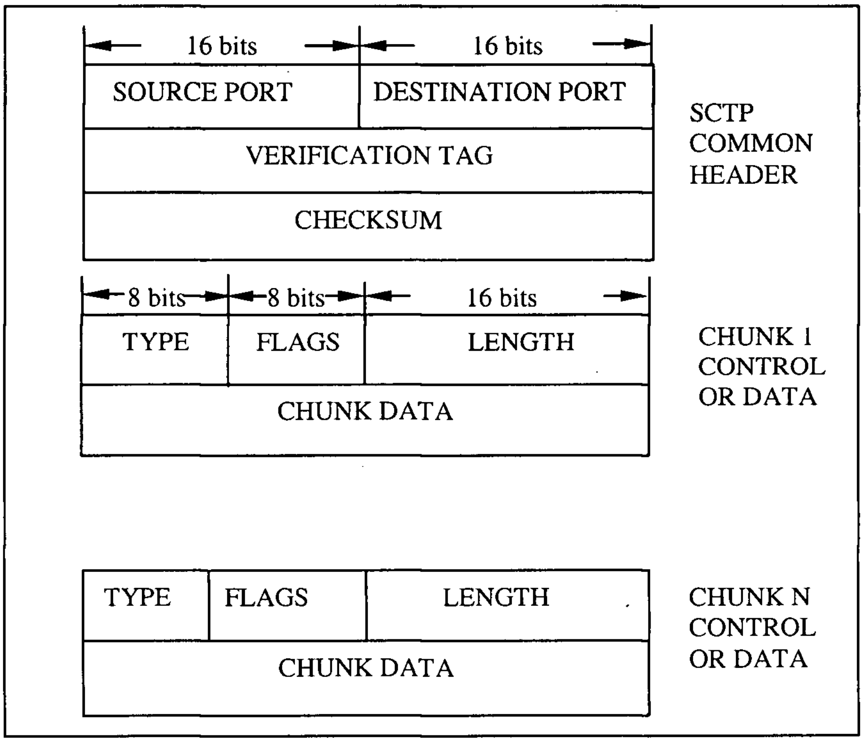


Fig. 2-5 SCTP packet format

Unlike TCP's byte-stream service, SCTP multi-streaming allows data from upper layer application to be multiplexed into one association. SCTP packets are structured to provide a message-oriented service and allow flexible message bundling. Fig. 2-5 illustrates a general SCTP packet format, which always begins with an SCTP common header. The common header provides the following three basic functions:

1) With the source and destination port numbers, together with IP addresses in the IP header, identify the association to which an SCTP packet belongs.

2) The verification tag ensures that the SCTP packet belong to the current incarnation of an association.

3) Data integrity of the entire packet is maintained with the checksum.

The remainder of an SCTP packet consists of one or more concatenated building blocks called chunks. Chunks fall into two basic types: those that carry user information are called data chunks, and those that carry control information are called control chunks. This characteristic of SCTP is different from both TCP and UDP packets, which have control information in the header and a single optional data field. Currently fourteen different control chunks have been defined for association establishment, association termination, data Selective Acknowledgement (SACK), destination failure detection and recovery, Explicit Congestion Notification (ECN) and error reporting, which leaves an additional 240 new chunk types that may be defined in the future by the IETF. Each chunk is delineated by a chunk header, which identifies a chunk's type, special flags needed by a given chunk type and the length of the chunk. SCTP has the flexibility of concatenating or bundling different chunk types into a single SCTP packet[1]. The only restriction is that the size of a packet should not exceed the Path Maximum Transmission Unit (PMTU).

2.3.2 Mobile SCTP (MSCTP)

SCTP introduces the idea of multi-homing [9]-[11][29] where a host can support multiple connections with different interfaces and IP addresses simultaneously in an association. Although the primary goal of this feature is error resilience, it can provide a simple and powerful framework for mobility support in IP networks. To support multi-homing, SCTP endpoints exchange lists of IP addresses during the initiation of an

[1] In the rest of this thesis, we only use bundle technology for control chunks. We are not going to bundle data chunks, i.e., each SCTP packet carries at most one data chunk.

association. Changing an SCTP endpoint's address list in a running association using the recently proposed SCTP DAR extension [31] makes mobility and seamless handover possible. SCTP DAR enables the SCTP endpoints to add and delete IP addresses, and change the primary IP address dynamically in an active association using Address Configuration (ASCONF) and Address Configuration Acknowledgement (ASCONF_ACK) chunks. SCTP with its DAR extension is called MSCTP. Since the typical use of SCTP is for reliable data stream transmissions between a server and its client(s), in this thesis, without loss of generality, we use a client-server model to apply the MSCTP technology and discuss a seamless handover at a Mobile Client (MC) between UMTS and WLANs, where the other endpoint is a Fixed Server (FS). IP mobility can be divided into issues of location management and handover management [13]. There are two types of sessions considered in MSCTP: (1) sessions originated at MCs towards FSs; and (2) sessions originated at FSs towards MCs. Note that only sessions in (2) require the additional location management functionality for the session originator to find the current location of an MC. Since our research focuses on discussion of handover management, we only consider sessions in scenario (1), which fits well in the client-server model. The handover configuration is shown in Fig. 2-6. The only requirement for supporting the handover based on SCTP is that the MC and the FS hosts are equipped with the MSCTP implementation, i.e., SCTP with DAR extension. The basic assumption for seamless handover is that the MC is able to obtain a new IP address in the new location with the help of Dynamic Host Configuration Protocol (DHCP) in IPv4/IPv6 or stateless address auto-configuration in IPv6. The number of IP addresses used by the FS can be either one or two, respectively resulting in two types of

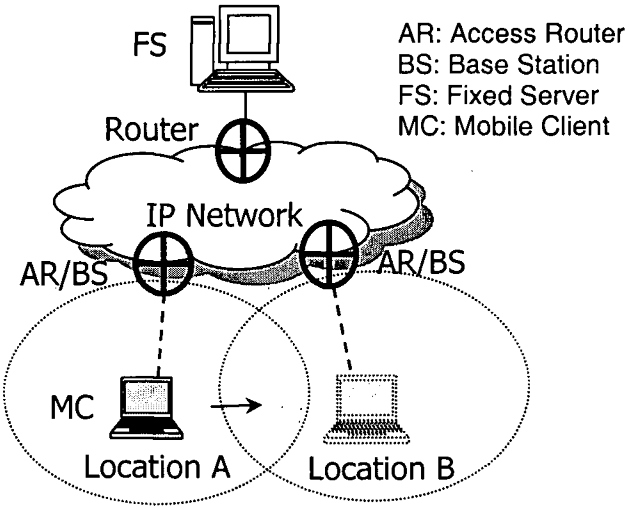


Fig. 2-6 SCTP support of seamless handovers

configurations for the MSCTP handover: a single homing asymmetric configuration [13] or a dual homing symmetric configuration [12]. For both of these configurations, the procedure for the MSCTP handover is:

1) Before a handover, an SCTP association is established between an MC and FS.

2) The MC moves to a new location and obtains a new IP address. The MC adds the new IP address to the SCTP association using the ASCONF and ASCONF_ACK chunks described in [31]. Thus, the MC becomes multi-homed and is now reachable by two different IP addresses.

3) The MC continues to move towards the new location. The MC should always play the active role to trigger an MSCTP handover because only the MC knows the movement of itself and the signal strength from the old and new BSs or APs. The method to realize a handover is by changing the MC's primary IP address. The possible rules for triggering a handover by the MC are: a) as soon as a new IP address is detected; b) by using an explicit indication from the underlying layer.

4) When the MC confirms the failure of the old link by detecting the signal strength from the physical layer, the MC informs its peer that its old IP address is now no longer reachable and can be removed from the SCTP association.

5) MC repeats the handover procedure whenever it moves to a new location, until the SCTP association is closed.

Although Riegel and Tuexen proposed to use SCTP to support mobility management in [12] and Koh *el al.* further describe the architecture and handover procedure in [13], the following problems are found in the work.

- The handover procedures described in both [12] and [13] are quite vague. Both proposed to use SCTP DAR extension to achieve the handover. However, these proposals lack detailed descriptions of the procedures. The parameter setting in the address configuration chunk (ASCONF) is not clear either.

- Authors of [12] and [13] have different ideas regarding the configuration used to support MSCTP. Authors of [12] strongly suggest that a server must use multiple IP addresses to provide MC with multiple paths in order to fully take advantage of the existence of a second interface at the MC for fault resilience. However, authors of [13] argue that it is natural to consider FS supporting handover with only one IP address since a fixed host should not add new IP addresses dynamically. If it is unavoidable to use two IP addresses, FS probably need to be in the dual homing state from the association initiating stage.

- Neither of these documents address implementations of simulation models or prototypes.

Authors in [32] presents a prototype implementation for mobile SCTP (so called M-SCTP in [32]) with the configuration described in [13]. However, the experimental results show that there is a long interruption time during an SCTP handover.

2.4 A Summary of Related Work

Table 2-3 summarize the protocols supporting handover that we have reviewed. The disadvantages of the network layer mobility can be summarized as follows:

- As router solutions, most network layer solutions require additional hardware network components to be introduced into the network. Thus, these solutions have a high impact on the existing networks.
- The overall network architecture is generally complicated for network layer solutions.
- Mapping between IP addresses and node/location identifiers increases the traffic load on the endpoints and may result in unexpected delays in handover management.
- It is difficult to implement flexible user interface selection policies. It may only be possible to explicitly defining static rules for interface changing.

We have also reviewed the related work on transport layer mobility. We summarize the advantages of the transport layer mobility as follows:

- Mobility is an end-to-end issue instead of a router or a network component issue. Solving mobility issues with transport layer protocols in endpoint systems make the entire network architecture simpler.
- Transport layer protocol-based mobility can adapt flow and congestion control parameters quickly and easily to the new network during and after the handover.

Table 2-3 Summary of protocols supporting handover

Protocol	**IP Version**	**Layer**	**Endpoint ID**	**Home Agent functionality**	**Interface selection policy**
MIP	IPv4/v6	Network	IP address(es)	HA	Not defined
LIN6	IPv6	Network	Generalized ID (GI)	Mapping Agent	Not defined
HIP	IPv4/v6	Between L3 and L4	Host Identity (HI)	Forward Agent	Not defined
M-TCP	IPv4/v6	Transport	Transport address	None	API
SCTP	IPv4/v6	Transport	SCTP transport address	None	API
SIP	IPv4/v6	Application	IP address(es)	SIP servers	Support

- From the hardware point of view, transport layer mobility does not need additional components or mobility enabled routers incorporated into the network. Therefore, it has a low impact on the existing networks. From the software point of view, there is no need to introduce new software states such as MIP's tunnel states into the network.
- Transport Layer Application Programming Interface (API) makes it easy to implement user interface selection policies.

SCTP is a relatively new protocol. Though recently researchers are investigating mechanisms that exploit SCTP's novel features to improve performance in wireless and mobile environments, there is still no large-scale experimental results available to verify if the current SCTP standard meets each of the high reliability requirements of the mobility managements in wireless networks or not. We believe that not only SCTP is a promising solution to support mobility, but also a good candidate for UMTS/WLAN integration in the next generation IP networks. MSCTP-based UMTS/WLAN integration requires neither hardware modifications to the current networks, nor software modifications to applications. It is independent of access technologies too. Therefore, it provides a network-independent solution that should be preferred by service providers.

Chapter 3 UMTS/WLAN Unforced VHO

This chapter explains the key design issues involved in MSCTP supporting UMTS/WLAN VHO. In Section 3.1, we present the architecture. In Section 3.2, we develop the VHO procedures in two possible configurations, the single-homing and the dual-homing configurations based on an abstract address-handling model, which is described in details in this section as well.

3.1 System Architecture

The objective of designing an UMTS/WLAN integration scheme is to make the VHO between the two types of networks as seamless and efficient as possible. The rationale behind the proposed scheme of supporting UMTS/WLAN VHOs using SCTP is that, due to the SCTP multi-homing feature, from the association point of view, it does not matter whether an endpoint's network interfaces belong to the same technology or not. As long as it is possible for an interface to establish a connection to the Internet via an IP address, the interface can be added into the current association via its IP address. MSCTP provides an end-to-end soft handover solution for mobility management. Therefore, introducing MSCTP to support UMTS/WLAN integrations makes the entire network architecture simpler. Fig. 3-1 shows the architecture of UMTS/WLAN VHO using MSCTP. The architecture can be either tight coupling or loose coupling. The basic assumption for the seamless VHO between UMTS and WLAN cell is that the MC is able to obtain a new IP address when it moves into a WLAN cell, via either DHCP or stateless address auto-configuration in IPv6 network. The general requirements for SCTP to

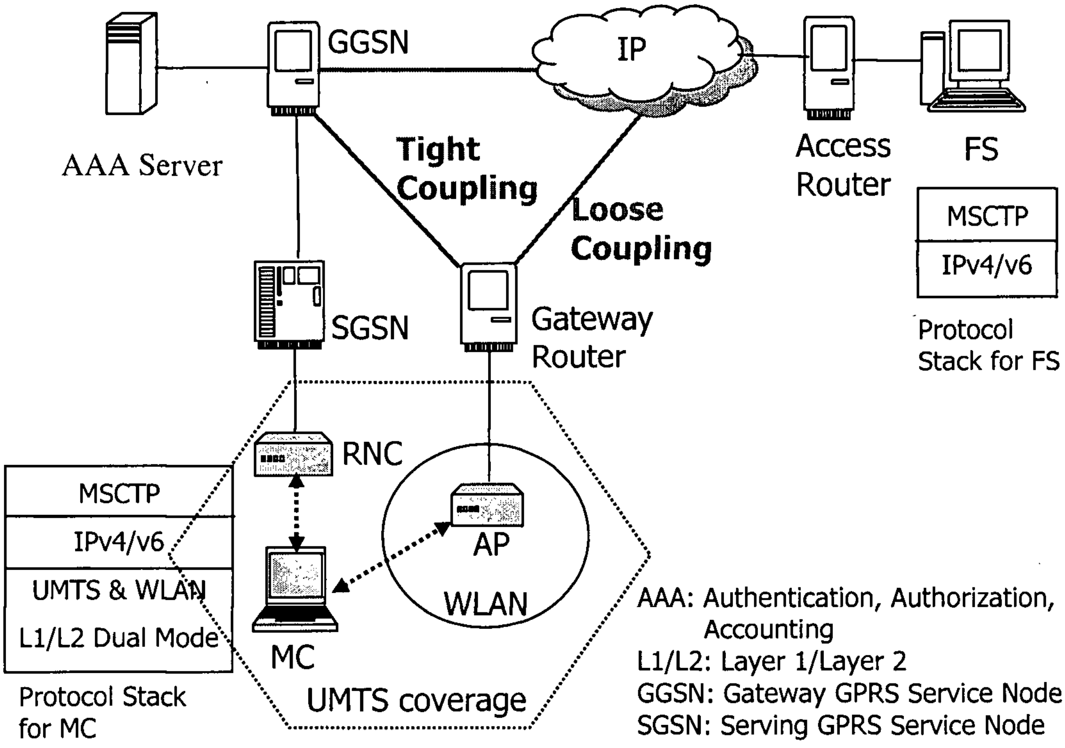


Fig. 3-1 Architecture of UMTS/WLAN integration using MSCTP

support seamless VHO are as follows:

- Both MC and FS are equipped with MSCTP implementation, i.e., SCTP with DAR extension;
- Dual-mode support of UMTS and WLAN is available at the physical and data link layers of the MC;
- Issues such as AAA, subscriber identification and QoS provisioning due to change of access network have been resolved through roaming agreements between UMTS and WLAN under one or more operator(s).

It should be noticed that the mobility between UMTS and WLAN is not completely symmetric. To take advantage of the higher speed of WLAN, the VHO from

UMTS to WLAN may take place as soon as the WLAN coverage is available. This is an *unforced VHO process*. However, in the reverse direction, the VHO from WLAN can be either a *forced VHO process* or not. This is because the WLAN to UMTS VHO can be triggered either by: 1) the MC leaving the WLAN coverage and losing the WLAN signal (a forced VHO process); or 2) the user's preference to perform a VHO from WLAN to UMTS under certain rules when both of WLAN and UMTS networks are available (an unforced VHO process). For example, during the WLAN peak traffic period, the user may not be satisfied with the quality of the WLAN service and may prefer to use UMTS instead. In this chapter, we only consider the VHO that is not a forced handover process in both directions. The case of WLAN to UMTS forced VHO will be discussed in Chapter 4. The reason we separate these two cases is that, for the forced handover process, SCTP failure detection and recovery functionalities are activated and result in a different handover procedure than that of an unforced one.

3.2 Unforced VHO Procedure

UMTS/WLAN unforced VHO means that the application or user's agent in the MC explicitly requests a VHO based on the user's preference when both of the UMTS and WLAN coverage are available. In this subsection, we firstly present an abstract address-handling model in Section 3.2.1. Using this model definition makes our discussion on address handling and handover processing easier to follow. Then we introduce the proposed UMTS/WLAN VHO procedures for MC and FS in both the single-homing and the dual-homing configurations in Section 3.2.2. In Section 3.2.3, by comparing the vertical handover delay and other characteristics of these two configurations, we recommend the preferred configuration, on which the remainder of this thesis will be based.

3.2.1 Abstract Address-handling Model Description

An SCTP association is the transport layer connection that SCTP provides to the endpoints. There are two endpoints and two address sets known by the peers [29] in an association. An association A between two endpoints $E1$ and $E2$ is given by:

$A = [E1\text{: } Addr\ (E1)\ ;\ E2\text{: } Addr\ (E2)]$.

In IP implementations, the outgoing interface of a multi-homed host is often determined by the destination IP address. The mapping of an outgoing source IP address and a destination address is done by a lookup in the host routing table maintained by the operating system [33]. Assume that an endpoint E has m source IP addresses: s_IP1, ..., s_IPm and its peer has n destinationIP addresses: d_IP1,..., d_IPn in the association. The primary path in the association is the connection between s_IP1 and d_IP1. Then the host routing table of the endpoint E can be expressed as:

$RT(E) =[(s\_IP1,d\_IP1);(s\_IP2,d\_IP2),\ldots,(s\_IPm,d\_IPn)]$.

Note that the primary IP address pair is separated from the secondary IP address pairs with a ";". When $m \neq n$, the address mappings are not in one-to-one correspondence, and the routing table may allow a maximum of $m \times n$ possible paths between the two endpoints.

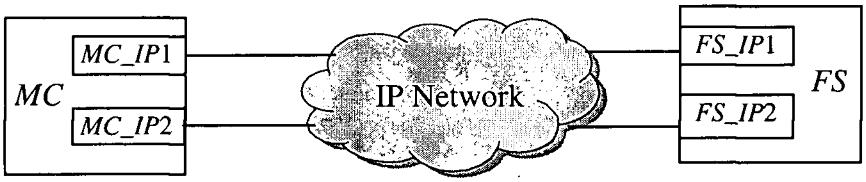


Fig. 3-2 Simplified architecture of dual-homing configuration

We now apply this model to summarize the key address handling features implemented in SCTP DAR extension. We use the configuration in Fig. 3-2 as an example, where an association A is given by:

$$A = [MC\text{: } Addr(MC);\ FS : Addr(FS)].$$

1) Add IP Address:

- Before Add IP Address process:

$$Addr_{old}(MC) = \{MC\_IP1\}$$

$$Addr(FS) = \{FS\_IP1\}$$

$$RT_{old}(MC) = [(MC\_IP1, FS\_IP1)]$$

$$RT_{old}(FS) = [(FS\_IP1, MC\_IP1)]$$

- After Add IP Address process (MC's new IP address MC_IP2 is added to the association):

$$Addr_{new}(MC) = Addr_{old}(MC) \cup \{MC\_IP2\} = \{MC\_IP1, MC\_IP2\}$$

$$RT_{new}(MC) = [(MC\_IP1, FS\_IP1)]$$

$$RT_{new}(FS) = [(FS\_IP1, MC\_IP1); (FS\_IP1, MC\_IP2)]$$

2) Delete IP Address:

- Before Delete IP Address process:

$Addr_{old}(MC) = \{MC\_IP1, MC\_IP2\}$

$Addr(FS) = \{FS\_IP1\}$

$RT_{old}(MC) = [(MC\_IP1, FS\_IP1)]$

$RT_{old}(FS) = [(FS\_IP1, MC\_IP1); (FS\_IP1, MC\_IP2)]$

- After Delete IP Address process (*MC*'s address *MC_IP2* is deleted from the association):

$Addr_{new}(MC) = Addr_{old}(MC) - \{MC\_IP2\} = \{MC\_IP1\}$

$RT_{new}(MC) = [(MC\_IP1, FS\_IP1)]$

$RT_{new}(FS) = [(FS\_IP1, MC\_IP1)]$

3) Set Primary Address:

- Before Set Primary Address process:

$Addr(MC) = \{MC\_IP1, MC\_IP2\}$

$Addr(FS) = \{FS\_IP1, FS\_IP2\}$

$RT_{old}(MC) = [(MC\_IP1, FS\_IP1); (MC\_IP2, FS\_IP2)]$

$RT_{old}(FS) = [(FS\_IP1, MC\_IP1); (FS\_IP2, MC\_IP2)]$

- After Set Primary Address process (*MC* requests *FS* to set address *MC_IP2* as the primary address):

$RT_{new}(MC) = [(MC\_IP2, FS\_IP2); (MC\_IP1, FS\_IP1)]$

$RT_{new}(FS) = [(FS\_IP2, MC\_IP2); (FS\_IP1, MC\_IP1)]$

Note that in the dual-homing configuration shown in Fig. 3-2, though there are four possible paths between the two endpoints, for reasons of considerations on path diversity, policy-based routing, load balancing, *etc.*, the routes between *MC* and *FS* are

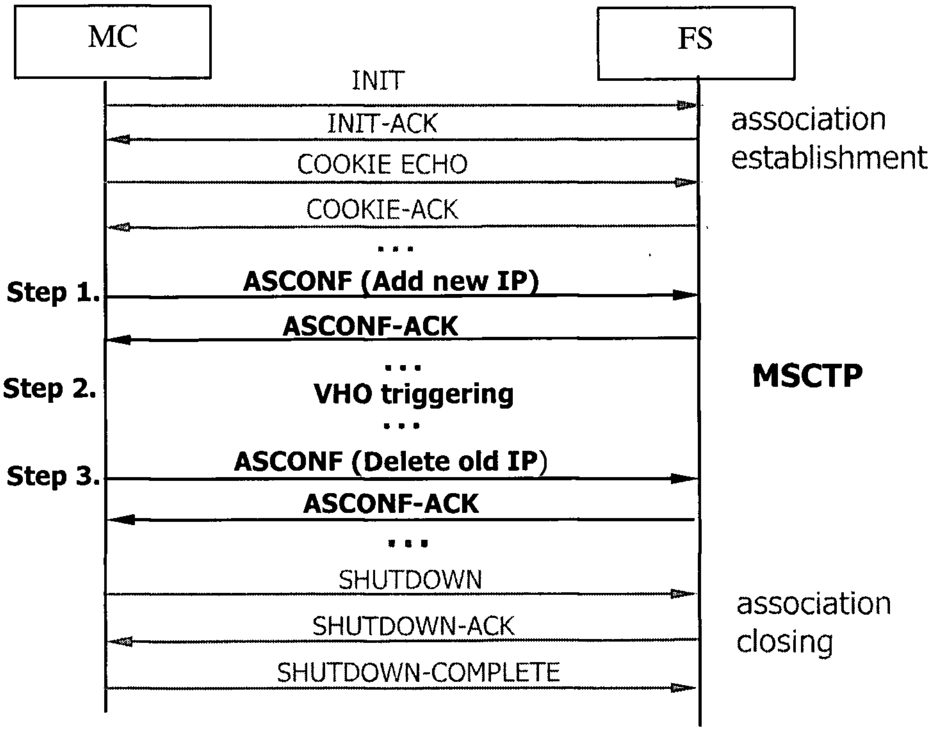


Fig. 3-3 Three basic steps in VHO procedure

set to be only two in the routing table, i.e. path (*MC_IP*1, *FS_IP*1) and path (*MC_IP*2, *FS_IP*2). This configuration simplifies the abstract address-handling model and our further discussions on VHO procedure.

3.2.2 The Proposed VHO Scheme and Procedures

To support VHOs, the FS may be configured for: (1) single homing, i.e., the FS provides only one IP address to support the handover; or (2) dual homing, i.e., the FS allows more than one (usually two) IP addresses to support the MC's mobility. For each of these configurations, the handover procedure has three basic steps between association establishment and closing processes as shown in Fig. 3-3:

- Step 1, *Add IP Address process*: MC moves into a new location, gets a new IP address and sends an ASCONF to request FS to add the new IP address to the current association.
- Step 2, *VHO Triggering process*: MC triggers a VHO by either sending a "Set Primary IP Address" request with an ASCONF or directly setting the remote primary destination IP address.
- Step 3, *Delete IP Address process*: MC is unable to detect signals from the old BS or AP and sends a "Delete IP Address" request with an ASCONF to FS to delete the MC's old IP address from the association.

1) The single-homing FS:

The VHO procedure for FS in the single-homing configuration is shown in Fig. 3-4. Note that UMTS to WLAN VHO is shown in the upper part and the VHO in the reverse direction is in the lower part of the figure. When an MC moves into a WLAN cell covered by a UMTS cell, it gets a new IP address *WLAN_IP*. The "Add IP Address" process allows MC to inform the FS its new IP address.

The "VHO Triggering" process allows MC to trigger a VHO based on some decision rules. The VHO from UMTS to WLAN is triggered by MC sending an ASCONF with parameter type set to "Set Primary Address". This is a handshake process. After that, FS sets the MC's new IP address as its primary destination address and sends an ASCONF_ACK to confirm the operation success. In this process, assuming that the data transmission rate is txmn_rate, because of the handshake process, the VHO delay can be calculated by:

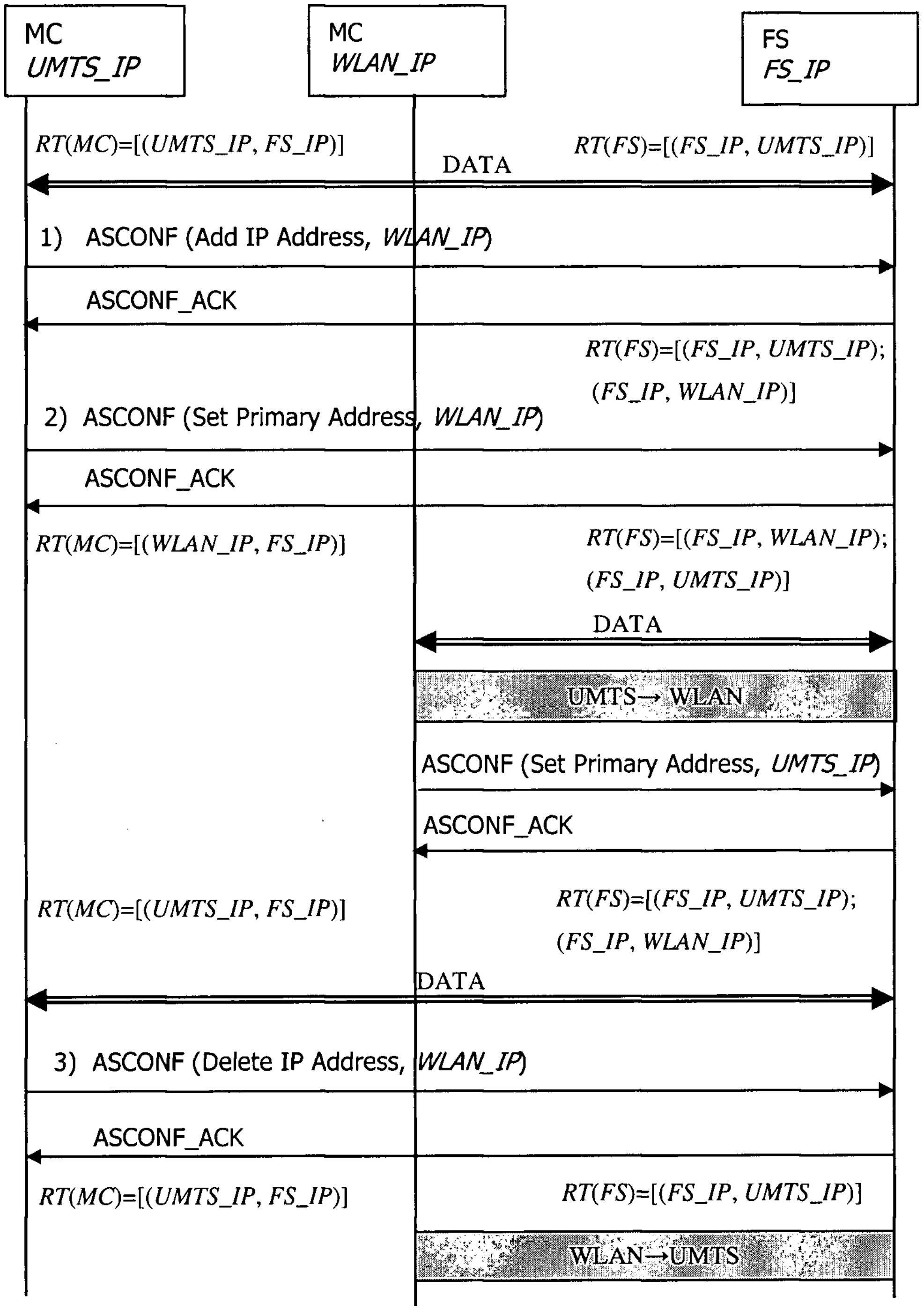

MC
UMTS_IP
MC
WLAN_IP
FS
FS_IP
RT(MC)=[(UMTS_IP, FS_IP)]
DATA
RT(FS)=[(FS_IP, UMTS_IP)]
1) ASCONF (Add IP Address, WLAN_IP)
ASCONF_ACK
RT(FS)=[(FS_IP, UMTS_IP);
(FS_IP, WLAN_IP)]
2) ASCONF (Set Primary Address, WLAN_IP)
ASCONF_ACK
RT(MC)=[(WLAN_IP, FS_IP)]
RT(FS)=[(FS_IP, WLAN_IP);
(FS_IP, UMTS_IP)]
DATA
UMTS→ WLAN
ASCONF (Set Primary Address, UMTS_IP)
ASCONF_ACK
RT(MC)=[(UMTS_IP, FS_IP)]
RT(FS)=[(FS_IP, UMTS_IP);
(FS_IP, WLAN_IP)]
DATA
3) ASCONF (Delete IP Address, WLAN_IP)
ASCONF_ACK
RT(MC)=[(UMTS_IP, FS_IP)]
RT(FS)=[(FS_IP, UMTS_IP)]
WLAN→UMTS

Fig. 3-4 VHO procedure (FS is in the single-homing configuration)

$$Delay_{\text{overall}} = T_{\text{ASCONF}} + T_{\text{handover}} \,, \tag{3.1}$$

where T_{ASCONF} is the transmission time of ASCONF and ASCONF_ACK chunks:

T_{ASCONF} = (ASCONF_chunk_size/txmn_rate + propagation_delay) × 2, and

T_{handover} = changeover command delay + buffered data transferring time.

After the WLAN to UMTS VHO Triggering process, if MC loses the signal from the WLAN cell, it enters in the "Delete IP Address" process. MC sends an ASCONF with parameter type set to "Delete IP Address" to request FS to release the address *WLAN_IP* from its host routing table. When MC receives the ASCONF_ACK from FS, MC deletes *WLAN_IP* from its address list and *WLAN_IP* is released from the association.

2) The dual-homing FS:

In SCTP, an SCTP packet is designed to carry multiple chunks, including both control and user data chunks. When small outbound messages are queued for transmissions, the sending endpoint can bundle as many small messages together as the current Path Maximum Transmission Unit (PMTU) allows and then transmits them in a single IP datagram [29]. An SCTP data receiver uses the length field of a chunk to determine the end of the data in each chunk, and then unbundles this bundled user message into individual messages. This is called the SCTP User Message Bundling service. In this way, the SCTP bundling message service not only reduces the overhead of additional SCTP and IP headers, it can also substantially reduce the number of SACKs needed by the data receiver. The bundling service is not a mandatory feature for implementation of SCTP, but we found it helpful to simplify the VHO procedure for FS in the dual-homing configuration.

The VHO procedure for FS in the dual-homing configuration is shown in Fig. 3-5. There are two differences between this procedure and the one for FS in the single-homing

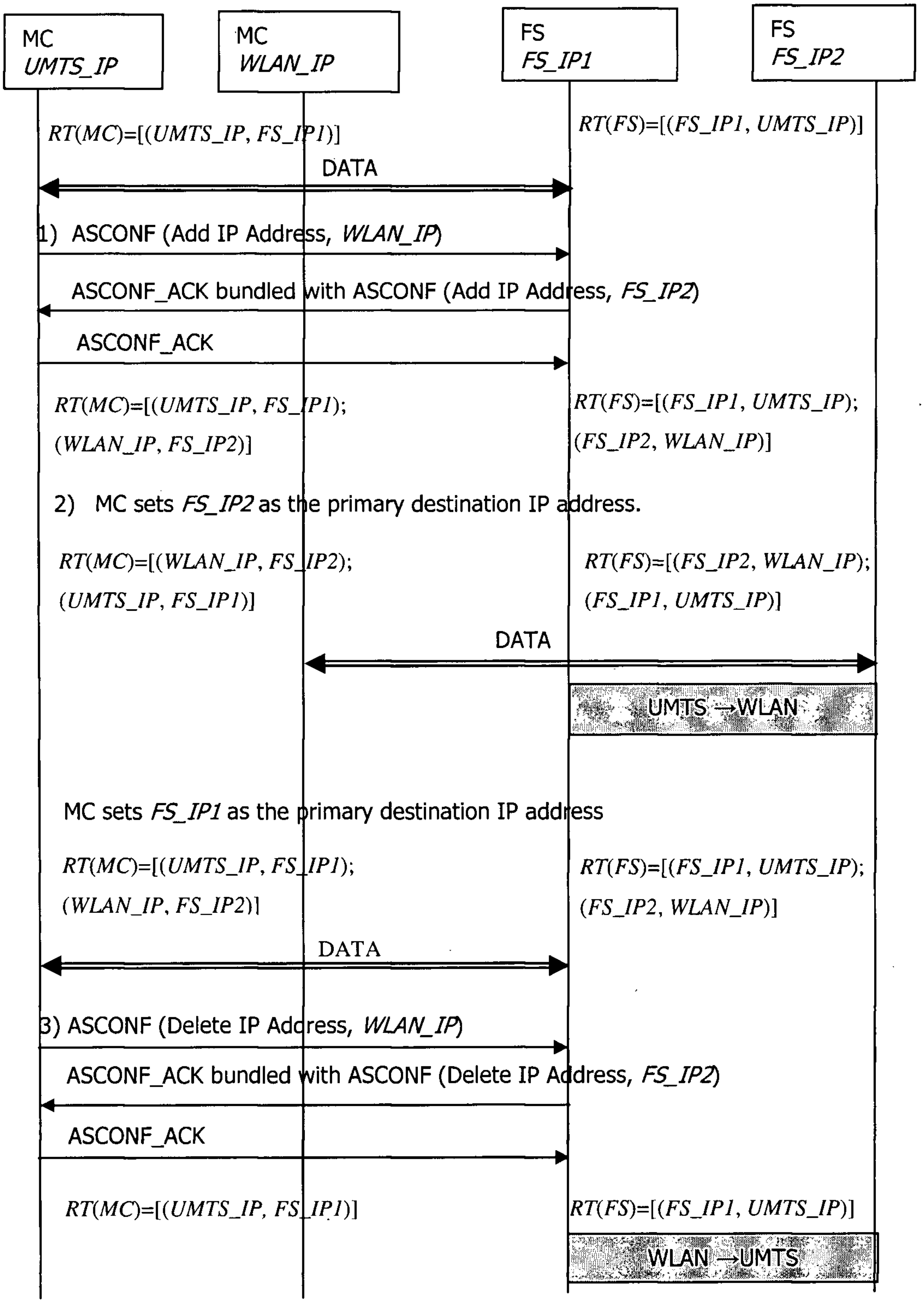

MC
UMTS_IP
MC
WLAN_IP
FS
FS_IP1
FS
FS_IP2
RT(MC)=[(UMTS_IP, FS_IP1)]
RT(FS)=[(FS_IP1, UMTS_IP)]
DATA
1) ASCONF (Add IP Address, WLAN_IP)
ASCONF_ACK bundled with ASCONF (Add IP Address, FS_IP2)
ASCONF_ACK
RT(MC)=[(UMTS_IP, FS_IP1);
(WLAN_IP, FS_IP2)]
RT(FS)=[(FS_IP1, UMTS_IP);
(FS_IP2, WLAN_IP)]
2) MC sets FS_IP2 as the primary destination IP address.
RT(MC)=[(WLAN_IP, FS_IP2);
(UMTS_IP, FS_IP1)]
RT(FS)=[(FS_IP2, WLAN_IP);
(FS_IP1, UMTS_IP)]
DATA
UMTS →WLAN
MC sets FS_IP1 as the primary destination IP address
RT(MC)=[(UMTS_IP, FS_IP1);
(WLAN_IP, FS_IP2)]
RT(FS)=[(FS_IP1, UMTS_IP);
(FS_IP2, WLAN_IP)]
DATA
3) ASCONF (Delete IP Address, WLAN_IP)
ASCONF_ACK bundled with ASCONF (Delete IP Address, FS_IP2)
ASCONF_ACK
RT(MC)=[(UMTS_IP, FS_IP1)]
RT(FS)=[(FS_IP1, UMTS_IP)]
WLAN →UMTS

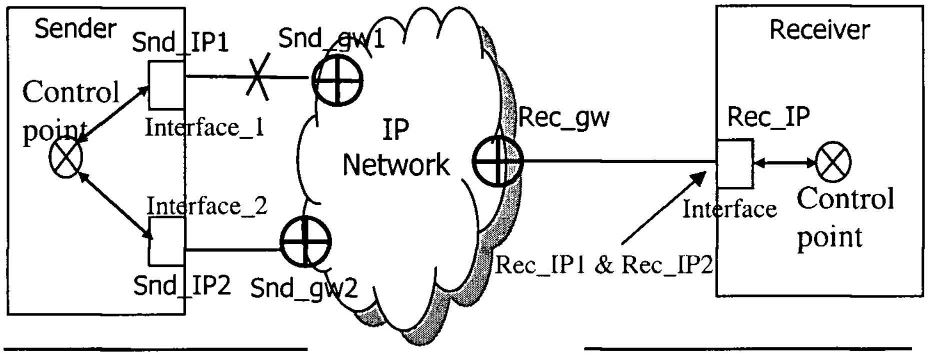
Fig. 3-5 VHO procedure (FS is in the dual-homing configuration)

configuration. The first difference is the "Add/Delete IP" Address process. In the dual homing configuration, when FS responds to MC's adding/deleting IP address request with an Acknowledgement (ACK), FS bundles an ASCONF to request MC to add/delete the FS's secondary IP address into/from the association. The MC sends ASCONF_ACK to confirm the completion of Add/Delete IP Address process. The second difference is in the VHO Triggering process. Since both MC and FS are in the dual-homing configuration. MC can directly set the FS's secondary address as the primary destination in its host routing table and start to send data on the new link. In this case, there is no handshake process and the VHO delay becomes:

$$Delay_{overall} = T_{\text{handover}}. \tag{3.2}$$

3.2.3 Comparison of Single-homing and Dual-homing Configurations

In practice, network layouts may be less than ideal. The endpoints engaged in communications may have an asymmetric number of network addresses assigned to them. Fig. 3-6 shows an example in which an SCTP association is set up between a multi-homed endpoint sender and a single-homed endpoint receiver. In such a case, the redundant network addresses at the sender will help very little in providing fault resilience to the communication because of the routing tables shown in the figure. Due to the lack of redundancy, if the path between gateway router Snd_gw1 and Snd_IP1 breaks in either direction, the association will not survive. This is because of the fact that the number of possible different paths that can be used to route data between two endpoints can never be larger than the minimum number of IP addresses used by the endpoint. Therefore, we see from the example that if FS and MC are in one-to-two asymmetric



Sender

Destination	Gateway
Rec_IP	Snd_gw1

Receiver

Destination	Gateway
Snd_IP1	Rec_gw
Snd_IP2	Rec_gw

Fig. 3-6 An asymmetric SCTP association

multi-homing configuration, a single path failure in the network may cause the association failure even when a backup path exists.

The best solution of this problem is to let the FS have two IP addresses (even with only one physical interface), thus enabling both MC and FS to be in symmetric dual-homing configuration. This two-to-two configuration enables easy distinction of the two links at the MC. Unlike the single-homing configuration, the dual-homing configuration allows SCTP to represent different paths by different entries in the host routing table. It is beneficial in terms of fault resilience for both MC and FS to use all the IP addresses available to them when VHO is performed.

The second advantage of the FS in dual-homing configuration over the single-homing configuration is, as we have seen in Section 3.2.2, the VHO delay for the FS in dual-homing configuration, T_{handover}, is less than that of the single-homing configuration, $(T_{\text{ASCONF}} + T_{\text{handover}})$.

Because of these two advantages, we recommend the dual-homing configuration to support VHOs in the UMTS/WLAN integration environment.

Chapter 4 WLAN to UMTS Forced VHO

In Chapter 3, we developed the UMTS/WLAN unforced VHO procedure in both directions. However, in an UMTS/WLAN overlay environment, besides the user application explicitly requesting VHO, the most likely reason of triggering a WLAN to UMTS VHO is that the mobile user leaves the WLAN coverage and losses the signal from the WLAN AP. This is referred as the WLAN to UMTS forced handover process. In this section, we specifically focus on studying the SCTP behaviors when an MC leaves the WLAN coverage and goes into a WLAN to UMTS forced VHO period.

The rest of this chapter is organized as follows. Section 4.1 describes the problem. Section 4.2 presents a new scheme that we call SMART-FRX to assist the data transmission during the WLAN possibly unreachable period. In Section 4.3, we present a comprehensive analytical model by modeling the SCTP throughput during the WLAN to UMTS forced VHO.

4.1 Problem Descriptions

In this section, we divide the packet losses in the wireless environment into three types in Section 4.1.1. We describe problems that may be encountered on both WLAN and UMTS links during a WLAN to UMTS forced VHO in Sections 4.1.2 and 4.1.3.

4.1.1 Transport Layer Protocol in the Wireless Network

In a wireless Internet environment, most packet losses are caused by multi-path fading over wireless channels, not by congestion [34] [35]. Mobile handovers as well as

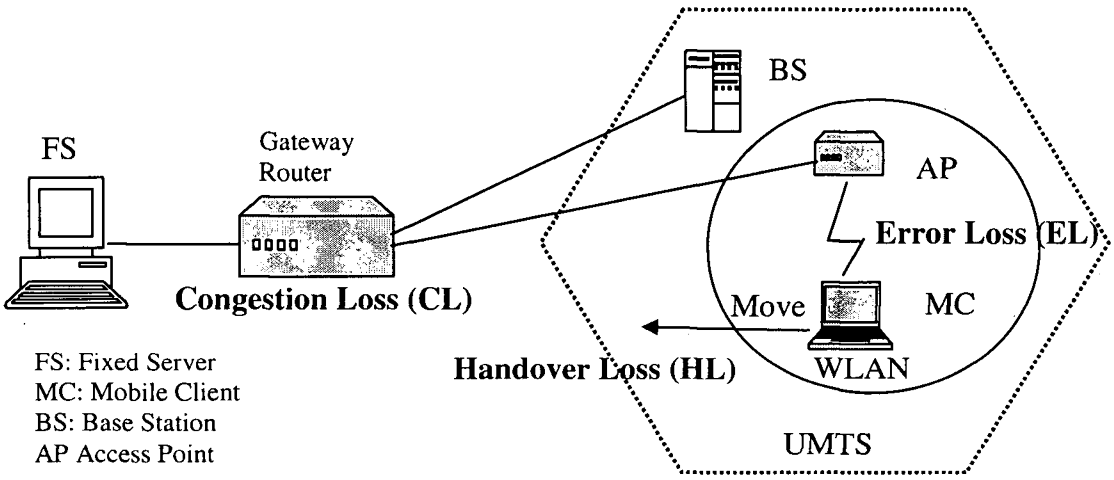


Fig. 4-1 Three types of packet loss affecting SCTP in mobile network

transmission errors in wireless Internet access result in the following three kinds of losses that will dominate the SCTP flow and congestion control behavior in WLAN/UMTS VHO situations. We illustrate the locations of these three losses in Fig. 4-1:

- *Congestion Loss* (CL): CL happens in the routers or Gateway routers in wireless networks because of bandwidth limitations.

- *Error Loss* (EL): This type of loss occurs in the wireless channel mainly due to multi-path fading. EL is the main packet losses when an MC communicates with a fixed server via a wireless link.

- *Handover Loss* (HL): This type of loss is unrelated to network congestion and comes from handovers between cells when packets should be re-routed to and from the mobile client in a new location. Packets may be lost due to futile transmissions over the wireless network when an MC moves out of the old cell (e.g. WLAN coverage) and enters into a new cell (e.g. UMTS coverage).

Among the above-mentioned three types of loss, two types of loss can dominate the SCTP performance in the WLAN to UMTS forced VHO process. One is consecutive packet dropping because of route disconnection from handovers, i.e, HL; the other is high error rate caused by multi-path fading over radio channels, i.e., EL. Same as TCP, SCTP uses fast retransmission and timeout retransmission mechanisms to recover packet losses. Because HL is mainly caused by an MC at which the Received Signal Strength (RSS) from the old coverage has fallen below threshold for at least a period of the minimum Retransmission Timeout (RTO_{min}), it generally results in the SCTP Timeout (TO) retransmission. On the other hand, when we model data transfer subject to EL, we assume that packet losses happen only in the direction from sender to receiver. This assumption is acceptable because it is very rare that wireless fading may cause consecutive packet losses in both directions longer than RTO_{min} (1-second time). Because of the SCTP fast retransmission algorithm, EL can be detected by a sender receiving duplicate SACKs [29]. In SCTP, a retransmission that is triggered by the sender receiving Four Duplicate SACKs (FD) is called the FD retransmission. In this chapter, we use HL to model the packet loss on the WLAN link and EL to model the packet loss on the UMTS link during the WLAN to UMTS forced VHO process.

4.1.2 Data Transmission Subject to HL

When an endpoint's peer is multi-homed, the general rules for data transmissions and retransmissions specified in [11] can be summarized as the following:

- By default, an endpoint should always transmit via the primary path.
- A sender should try to retransmit a chunk to an alternative active destination address.

- An SCTP endpoint uses a dynamic changing Timer 3 for Retransmission (T3-rtx) to ensure data delivery in the absence of any feedback from its peer. The initial value of this timer is referred as RTO.

- Each time the T3-rtx expires for a destination address, an error counter of that destination address will be incremented and the value of the T3-rtx is doubled. An upper bound of this timer $T3\text{-}rtx_{\max}$ is decided by the Path.Max.Retrans threshold. When the value of the error counter exceeds the Path.Max.Retrans, the endpoint should mark the destination address as inactive. Therefore, the Path.Max.Retrans threshold is used to detect the failure of an individual address of the peer. Assume that the Path.Max.Retrans is *PMR*, the first T3-rtx is *RTO*, the total time taken to detect the failure of a destination IP address *addr* is

$$Time_{failure}(addr) = RTO + 2 \times RTO + 2^2 \times RTO + \ldots + 2^{PMR} \times RTO.$$

We use the examples shown in Figs. 4-2 to 4-3 to illustrate the above-mentioned rules. If *PMR* is five as recommended in the specification [11], link failure occurs when there are six consecutive timeouts detected. Figs. 4-2 and 4-3 show the SCTP transmission when the WLAN link is in possibly unreachable period and the WLAN link is in failure, respectively. If we assume that *RTO* is set to be $\text{RTO}_{\min}$, which is 1 s as specified in the specification [11], the minimum total time for the WLAN link failure detection is:

$$Time_{failure}(addr_{WLAN}) = RTO + \ldots + 2^5 \times RTO = 63 \text{ s}.$$

In other words, as shown in Fig. 4-3, when an MC leaves the WLAN core area and losses the signal from the WLAN coverage, in a 63-second WLAN possibly

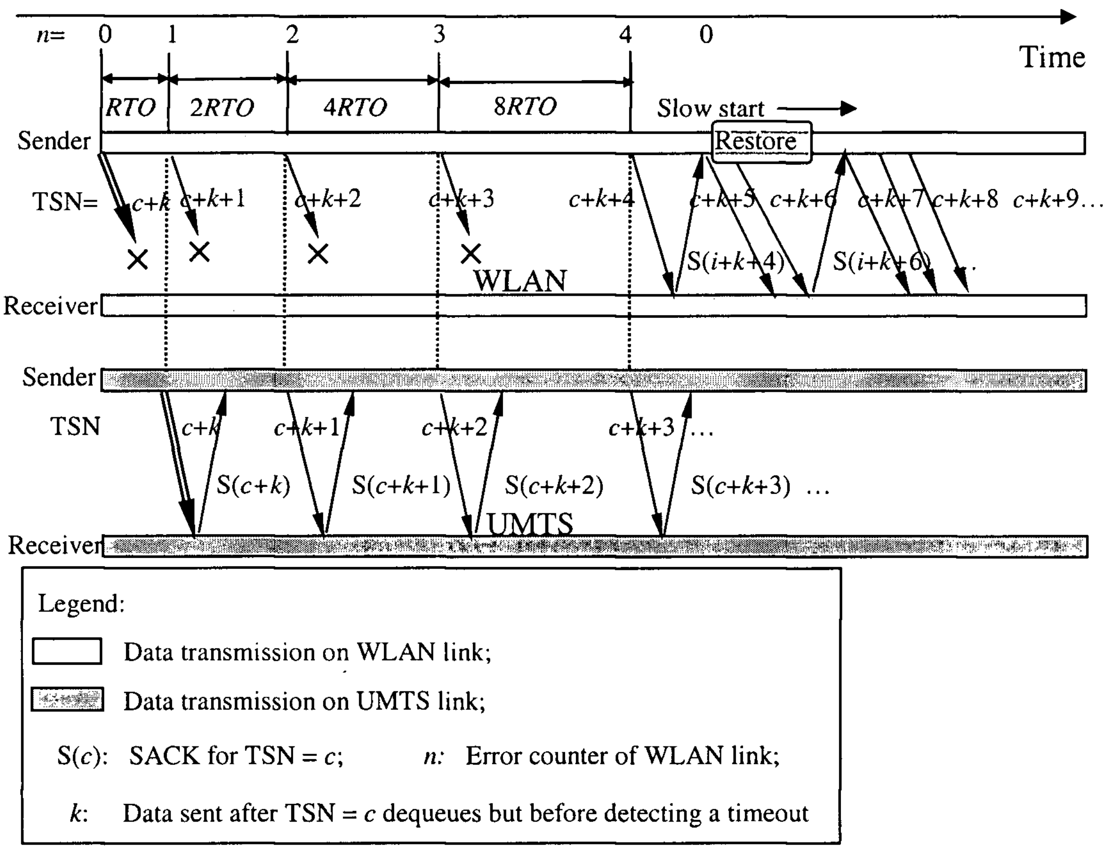


Fig. 4-2 SCTP transmissions when the WLAN link is in possibly unreachable period

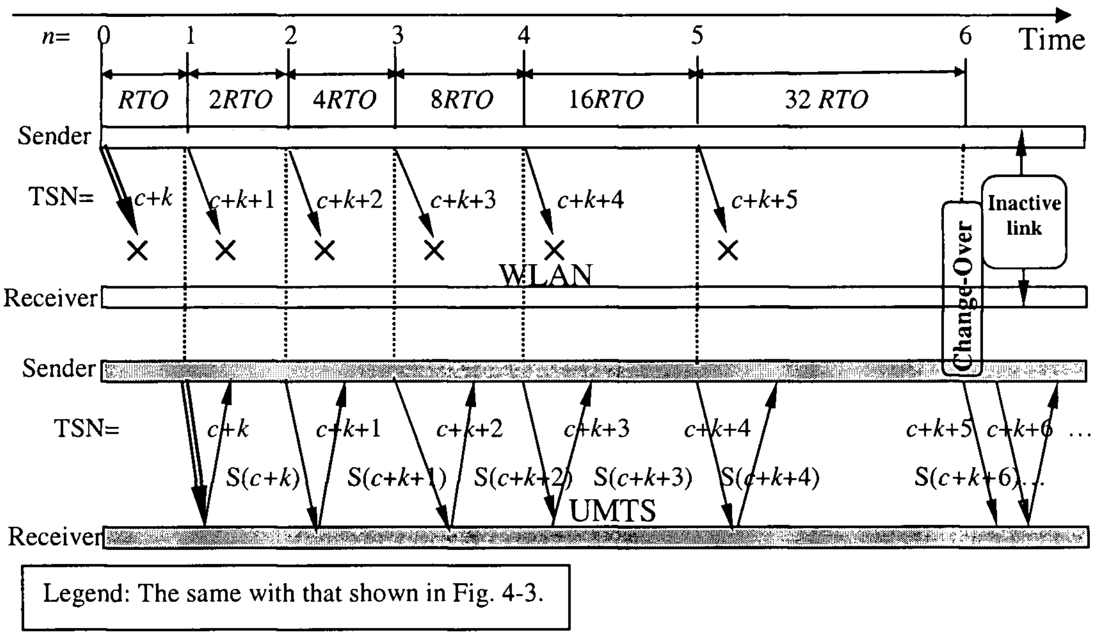


Fig. 4-3 SCTP transmissions when the WLAN link is in failure

unreachable period, only (*k*+6) packets are successfully transmitted. In the following we explain how *k* is determined. We assume that a data chunk whose Transmission Sequence Number (TSN) is *c* encounters a timeout. When the chunk *c* dequeues from the sending buffer of the WLAN link, a T3-rtx equals to *RTO* is started. Before T3-rtx expires, the sender does not know that the chunk *c* will encounter a timeout. Therefore, the sender continuously sends data on the link, and *k* is the number of data chunks sent in the time period, which is from the time that chunk *c* dequeues from the WLAN sending buffer until the sender detects its timeout. In all cases, the condition of a TO retransmission being triggered is the complete loss of a block of data, TSNs from *c* to (*c*+*k*) and their corresponding SACKs. If the sender can receive SACKs during this block of data transmissions, the SCTP fast retransmission is activated instead.

4.1.3 Data Transmissions Subject to EL

In this section, we study the SCTP behavior when transmissions are subject to EL. In the SCTP dual-homing configuration, according to the fault resilience and routing mechanism, when both primary and secondary links are not stable, data retransmissions on these two links follow Round Robin scheduling. Assuming that there are only two paths between two endpoints, a lost packet on the primary link is retransmitted on the secondary link; if the packet is found to be lost again on the secondary link, either detected by TO or by FD, the packet is next retransmitted on the primary link. This process repeats until the association error counter reaches the value of Maximum Association Retransmissions (Association.Max.Retrans). If the association error counter exceeds the limit indicated in this protocol parameter, the local endpoint marks the peer endpoint as unreachable and closes the association.

In our case, when the WLAN link is in the possibly unreachable period, all TO data on the WLAN link are retransmitted on the UMTS link. If any of the packets are lost on the UMTS link, these packets are retransmitted via the WLAN link. However, because the WLAN is currently in the possibly unreachable period, the packet is thus waiting in the WLAN sending-buffer. Even if the packet can be dequeued, because HL is in effect on the WLAN link, it is unlikely that the packet can be successfully transmitted. While the UMTS link is unable to receive any information about this retransmission for a long time, it considers the packet as lost and a TO retransmission is triggered. This results in both WLAN and UMTS links being in TO retransmission processes. It would greatly reduce the data transmission rates and easily cause a shutting down of the association.

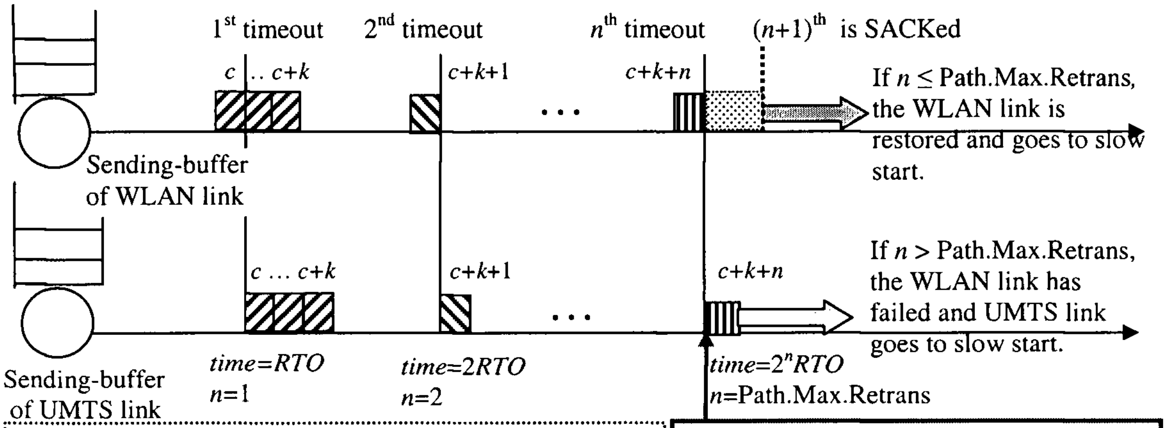
4.2 The Proposed Scheme

From the problem descriptions, we see that under the current SCTP specification, when a primary link is in the failure detection period, SCTP packet TOs may cause serious slow down of data transmissions. This is because in the current SCTP specification, under the assumption that failures are temporary and can be repaired quickly, the SCTP's failure detection and recovery mechanism acts conservatively to ensure that associations do not prematurely and incorrectly assume link failures. However, a drawback of this conservative approach is that the performance unnecessarily suffers during the failure detection period. We show that the current SCTP failure detection and recovery mechanism described in the specification is too conservative to fully exploit SCTP's multi-homing capability to maintain seamless communications between the endpoints. We propose a novel scheme called SMART-FRX, a solution that does not wait until the WLAN link is restored or is determined to be in failure before initiating the recovery process. The SMART sub-scheme allows the data buffered for transmissions on the WLAN link to be multicasted on both WLAN and UMTS links during the time the WLAN destination IP address is possibly unreachable. If there is any packet loss caused by multi-path fading when the data is being multicasted, the FRX sub-scheme sends the retransmissions to the same destination IP address, without alternating retransmissions between the primary and secondary links. We argue that the proposed scheme fully takes advantage of the SCTP's multi-homing feature and avoids unnecessarily long delays caused by the retransmissions to the possibly unreachable alternate IP address. Therefore, it can improve the overall throughput significantly.

In this section, we describe the SMART sub-scheme in Section 4.2.1 and the FRX sub-scheme in Section 4.2.2.

4.2.1 The SMART Sub-scheme

We propose a new queue management scheme we call SMART to deal with HL on the WLAN link so that the overall throughput can be increased during the WLAN possibly unreachable period. The WLAN possibly unreachable period is defined as the period that begins at detecting of the first TO, i.e., the WLAN link error counter is set to be one, and ends at the time of either the WLAN link is restored or the link is set to be inactive. If the WLAN link error counter resets to zero before exceeding the Path.Max.Retrans, then the link is being restored. If the WLAN link error counter


1st timeout
2nd timeout
nth timeout
(n+1)th is SACKed
c .. c+k
c+k+1
c+k+n
If n ≤ Path.Max.Retrans, the WLAN link is restored and goes to slow start.
Sending-buffer of WLAN link
c ... c+k
c+k+1
c+k+n
If n > Path.Max.Retrans, the WLAN link has failed and UMTS link goes to slow start.
Sending-buffer of UMTS link
time=RTO
n=1
time=2RTO
n=2
time=2^nRTO
n=Path.Max.Retrans

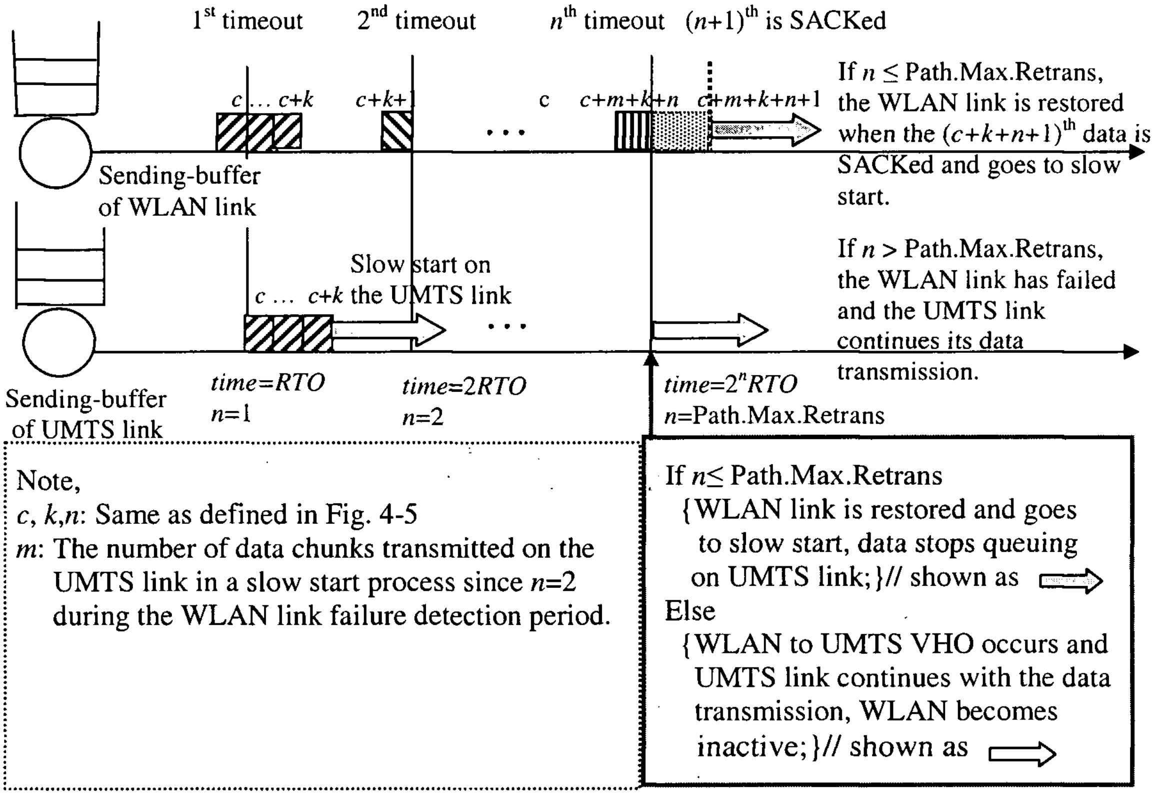
Note,
c: TSN of data chunk that encounters the 1st timeout
k: The number of data chunks transmitted between the chunk *c* dequeued and detection of its timeout
n: Error counter of the WLAN link, i.e., the numbers of consecutive timeout detected.

```
If n≤ Path.Max.Retrans
  {WLAN link is restored and goes
   to slow start;}// shown as ⇨
Else
  {WLAN to UMTS VHO occurs
   and UMTS link goes to slow
   start;}// shown as ⇨
```

Fig. 4-4 Data transmissions before application of the SMART scheme

exceeds the Path.Max.Retrans, then the WLAN link is set to be inactive. In SCTP, if a sender does not receive any responses from the peer in a given period, the sender cannot wait forever. The phrase "possibly unreachable" emphasizes that, if a sender notices that all expected responses are missing for consecutive packets sent to a destination address, the sender will then have to conclude that the particular destination address of the peer is very likely unreachable now. This leads to the algorithm of the TO failure detection in SCTP. Fig. 4-4 illustrates the data transmission before application of SMART.

Our proposed SMART mechanism is to copy all the buffered data from the WLAN link to the UMTS link so that the data in sending-buffers of both WLAN and UMTS links are completely the same. New data are queued at the tails of both of these two sending-buffers. As shown in Fig. 4-5, because of the copying and multicasting actions, instead of waiting and only retransmitting the TO data from the WLAN link, the


1st timeout
2nd timeout
nth timeout
(n+1)th is SACKed
c ... c+k
c+k+1
c
c+m+k+n
c+m+k+n+1
Sending-buffer of WLAN link
If n ≤ Path.Max.Retrans, the WLAN link is restored when the (c+k+n+1)th data is SACKed and goes to slow start.
Slow start on the UMTS link
c ... c+k
If n > Path.Max.Retrans, the WLAN link has failed and the UMTS link continues its data transmission.
Sending-buffer of UMTS link
time=RTO
n=1
time=2RTO
n=2
time=2^nRTO
n=Path.Max.Retrans

Note,
c, *k*,*n*: Same as defined in Fig. 4-5
m: The number of data chunks transmitted on the UMTS link in a slow start process since *n*=2 during the WLAN link failure detection period.

```
If n≤ Path.Max.Retrans
  {WLAN link is restored and goes
   to slow start, data stops queuing
   on UMTS link;}// shown as ⇨
Else
  {WLAN to UMTS VHO occurs and
   UMTS link continues with the data
   transmission, WLAN becomes
   inactive;}// shown as ⇨
```

Fig. 4-5 Data transmissions with application of the SMART scheme

UMTS link enters into a slow start process since the time of the first TO is detected on the WLAN link. This duplicate queuing action ends when the data sender detects the end of the WLAN possibly unreachable period. After this, 1) if the WLAN link is restored, new data are queued only in the a slow start process since the time of the first TO is detected on the WLAN link. This duplicate queuing action ends when the data sender detects the end of the WLAN possibly unreachable period. After this, 1) if the WLAN link is restored, new data are queued only in the WLAN sending-buffer and the WLAN link remains as the primary connection. The UMTS link becomes idle after all the buffered data have been sent out; or 2) if finally the WLAN link error counter exceeds the Path.Max.Retrans, a WLAN to UMTS forced VHO is triggered. As a result, the UMTS link becomes the primary connection and new data is queued only in the UMTS sending-buffer, while the WLAN link becomes the secondary connection and goes to the inactive state. In any case, the slow start data transmission process on the UMTS link starts at the beginning of the WLAN possibly unreachable period, instead of at the end of this period.

4.2.2 The FRX Sub-scheme

The difference between HL and EL is that HL is usually detected by a TO, while EL can be detected by a sender receiving duplicate SACKs. Both TCP and SCTP detect losses in two ways, the retransmission timer timeouts and duplicate SACKs (or ACKs), but with a minor difference. In SCTP, four instead of three duplicate SACKs (or ACKs) are needed to trigger a retransmission. Except for this, SCTP bases its congestion control on the TCP congestion control principles [36].

In order to avoid EL on the UMTS link being unnecessarily retransmitted on the WLAN link when buffered data are multicasted under SMART, we propose to retransmit packet losses caused by EL to the same destination. According to the current SCTP fast retransmit algorithm, a single missed data chunk can be retransmitted quickly on the alternative link before the retransmission timer expires. Because of the shorter loss detection time of FD compared with that of TO, the data transmission rate when recovering from EL is generally better than that from HL. The idea of "all retransmissions are sent to the same destination as their original transmissions" was first proposed by a group of researchers in University of Delaware in [37]. Here, we extend this idea to deal with different losses caused by different reasons. If a loss is detected by the TO, SMART is activated to multicast the buffered and new data on both primary and alternate IP addresses. If FD detects a loss, the SCTP fast retransmission mechanism is triggered and the retransmission should be sent to the same destination IP address. The later scheme is what we call the FRX. With the proposed FRX solution, during the WLAN to UMTS forced handover, the unnecessary FD retransmissions on the possibly unreachable WLAN link and long waiting delays on the UMTS link can be avoided.

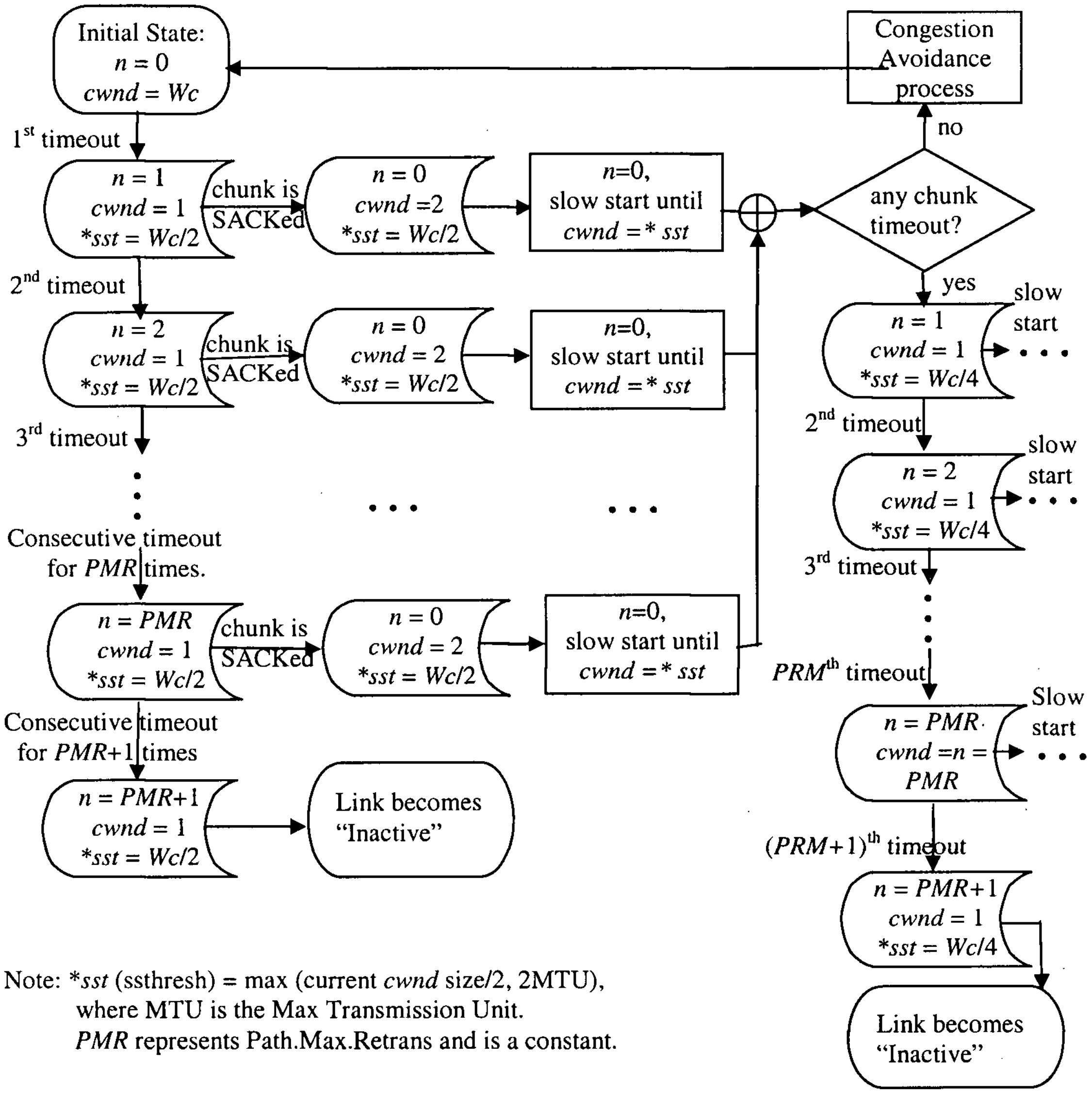
4.3 The Proposed Analytical Model

Same as TCP, SCTP is a window-based transport protocol that provides reliable end-to-end data communications. It performs flow control and congestion control by regulating its sending window. SCTP includes several mechanisms, among which the timeout, the slow start, the congestion avoidance and the fast retransmission can have a significant effect on the traffic pattern. The congestion window, or *cwnd*, is a variable maintained and dynamically updated by the data sender. It is an indication of how much more data traffic, in number of bytes or chunks, the sender can inject into the path between the sender and receiver before causing path congestions. It is noted that when the data receiver is multi-homed, the data sender needs to maintain a separate cwnd for each of the destination IP addresses of the data receiver. In this thesis, we use chunk or packet as the unit of cwnd.

In recent years, some analytical models have been proposed to characterize TCP performance in terms of round-trip delay and packet loss rate. The best-known analytical models for the steady state performance of TCP [17][18] captured the essence of TCP's congestion avoidance behavior by taking into account of fast retransmissions, timeouts and the impact of window limitations. Authors of [19] examined experimentally the application of TCP analytical model [17] to SCTP sources. However, the experiment results show that the model generally provide a pessimistic estimate of the throughput compared with that measured in a corresponding experiment. This is because in the TCP analytic model [17], the packet error rate p is not in fact the loss rate but the probability that a packet will be lost within a TCP "round" given that there has been no loss so far in the round. Once such a loss occurs, the remainders of packets sent in the current round are also considered to be lost. Thus, in the TCP analytical model [17], packet losses occurring in groups result in an actually higher per-packet loss rate. Other than this, there is very little work addressing the analytical modeling of SCTP.

In order to theoretically analyze the SCTP performance in the WLAN to UMTS forced VHO process, in this chapter, we propose and derive an analytical SCTP performance analytical model. We consider an SCTP flow starting at time $t = 0$ s, for any given time t>0, the variable *pkt* is defined to be the number of packets successfully transmitted in the interval [0,t]. Then the throughput *Thr* during the interval [0, t] is defined as *Thr* = *pkt* / *t*. Note that *pkt* is the number of packets that has been SACKed. Unlike the work of [17]-[19], in calculations of the throughput, our model captures the dynamic changes of cwnd for the slow start, congestion avoidance and fast retransmission processes, and the changes of RTT with the delay in each round.

The remainder of this section is organized as follows. In Section 4.3.1, we present how we model data transmissions subject to HL on the WLAN link. In Section 4.3.2, we develop an analytical model to study the data transmissions subject to EL on the UMTS link.



Note: *sst (ssthresh) = max (current *cwnd* size/2, 2MTU),
where MTU is the Max Transmission Unit.
PMR represents Path.Max.Retrans and is a constant.

Fig. 4-6 Transitions of cwnd size and error counter subject to HL

4.3.1 Modeling Data Transmission Subject to HL on the WLAN Link

Fig. 4-6 is a state transition diagram that describes the behavior of *n* (error counter), *cwnd* (cwnd) and *sst* (slow start threshold (ssthresh)) on the WLAN link in the presence of HL. *Wc* is the current window size, *PMR* represents the Path.Max.Retrans, *n* is the value of the error counter or the times of consecutive TOs detected and $n \leq (PMR+1)$. Then,

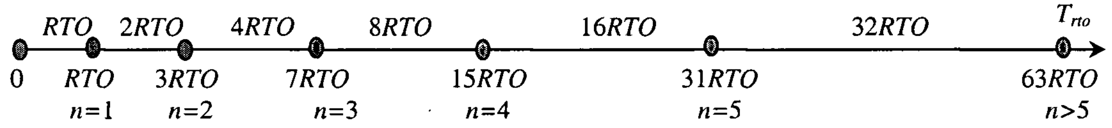


Fig. 4-7 The relation between the error counter n and T_{rto}

1) In case of a timeout, all data that are in flight are marked for retransmissions: The *cwnd* will be set to one MTU (Message Transmission Unit), and the *sst* will be set to either one-half the current window size *Wc* or two MTU (whichever is greater).

2) If n reaches (PMR + 1), then the WLAN to UMTS forced VHO is triggered. The UMTS link becomes the primary connection. The WLAN link becomes the secondary connection and enters to the state of "Inactive".

We now model the SCTP data transmission subject to HL on the WLAN link. As shown in Fig. 4-7, we define the duration of an MC staying outside of the WLAN coverage area as T_{rto}. RTO equals to RTO$_{\min}$ (one-second time) as recommended in specification. Then the throughput Thr_{HL} (in the unit of packets/second) is given by,

$$Thr_{HL} = \begin{cases} 0, & \text{if } T_{rto} \in (0,1\,\text{s}], \text{i.e.}, n \le 1 \\ (k+2)/T_{rto}, & \text{if } T_{rto} \in (1\,\text{s},3\,\text{s}], \text{i.e.}, n = 2 \\ (k+3)/T_{rto}, & \text{if } T_{rto} \in (3\,\text{s},7\,\text{s}], \text{i.e.}, n = 3 \\ (k+4)/T_{rto}, & \text{if } T_{rto} \in (7\,\text{s},15\,\text{s}], \text{i.e.}, n = 4 \\ (k+5)/T_{rto}, & \text{if } T_{rto} \in (15\,\text{s},31\,\text{s}], \text{i.e.}, n = 5 \\ (k+6)/T_{rto}, & \text{if } T_{rto} \in (31\,\text{s},63\,\text{s}], \text{i.e.}, n = 6 \end{cases} \tag{4.1}$$

where k is the number of packets on the way to the destination during the time from the first TO chunk dequeuing until the TO being detected. For details of the value k, please refer to Section 4.1.2.

4.3.2 Modeling Data Transmission Subject to EL on the UMTS Link

Fig. 4-8 illustrates the transitions of cwnd size and probabilities on the UMTS link in the presence of EL. The definitions such as *cwnd*, *sst*, *Wc*, *PMR* and T_{rto} are the same as in Sections 4.3.1 and 4.3.2. Besides, we set the initial value of *sst* to 65536 bytes (or 45 packets) with the packet size set to 1468 bytes. We define p to be the packet loss rate on the UMTS link and $q = (1 - p)$ is the packet successful transmission rate. Then, we have:

1) When there is no packet loss, the slow start process stops when *cwnd* reaches 45 packets. This is because, when the *cwnd* equals to or is longer than the initial ssthresh, the system enters into congestion avoidance process from slow start. The SCTP congestion avoidance process continues after *cwnd* reaches 45.

2) In the proposed SMART-FRX scheme, the retransmissions of EL packets on the UMTS link are sent to the same destination address. Whenever the sender receives duplicate SACKs, the sender reduces its cwnd to one-half of the current window size, i.e., *cwnd* is set to be *Wc*/2 and *sst* is set to either *Wc/2* or two MTU (whichever is greater).

Before application of the SMART-FRX, since HL on the WLAN link dominates the SCTP performance, we ignore any EL that may happen on both the WLAN and UMTS links. Then, we can obtain the system throughput from equation (4.1).

After SMART-FRX is applied, the UMTS link enters into slow start during $1 \text{ s} \leq T_{rto} \leq 63$ s. In this case, EL on the UMTS link may dominate the SCTP performance and cannot be ignored.

To develop an analytical model to study the SCTP performance on the UMTS link with EL, we assume that packet losses happen only in the direction from the sender to the receiver. We model SCTP behavior in terms of "rounds", where a round starts when the sender begins the transmission of a window of packets and ends when the sender receives the last acknowledgement for packets in this window. The duration of a round is a function of window size *cwnd.* SCTP essentially doubles its window size for slow start stage and increases the cwnd size linearly for congestion avoidance stage if none of the packets in the previous window gets lost during the previous RTT. If there were one or more than one packet in the previous window that were lost, the window size is set to half of the current window size. We use a recursive function to calculate the SCTP throughput at the i-th round.

We use arrays $cwnd(i)$ and $qq(i)$ to denote the congestion window sizes and the packet successful transmission probabilities after the i-th round, respectively. We use the variables $pkt(i)$ and $RTT(i)$ to denote the accumulative number of successful packet transmissions and the accumulative RTT until the i-th round, respectively. Assuming that we already know the following parameters: the network propagation delay is a constant given by *delay*, the chunk size is given by *Chunk_size*, the data SACK chunk size is given by *SACK_size*, the data transmission rate is *txmn_rate* and the loss rate is p. The transitions of the array $cwnd(i)$ and $qq(i)$ are illustrated in Fig. 4-8.

1) Before the data transmission starts, i.e., $i = 0$, there is only one element in $cwnd(i)$ and $qq(i)$. Since the initial cwnd size is 2 as specified in [11], we set the following initial values:

   $cwnd(0) = (2)$, and $qq(0) = (1)$.

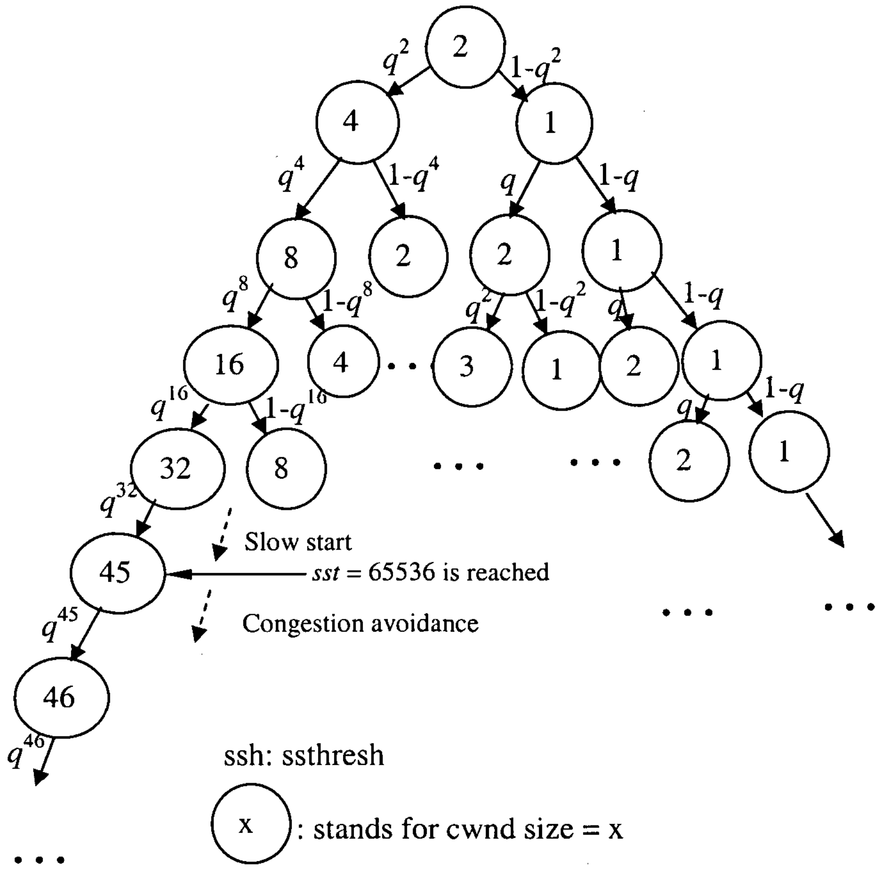


Fig. 4-8 Transitions of cwnd size and probability subject to EL

2) During the 1st round of data transmissions , we have

a) The number of successful packet transmissions in the 1st round is given by,

$pkt(1) = cwnd(0) \times q \times qq(0) = 2q$.

b) The RTT of the 1st round data transmission is,

$RTT(1) = rtt(cwnd(0)) \times qq(0) + delay = rtt(2) + delay$,

where *rtt* is a function of *cwnd*, which is the round trip time for *cwnd* packets transmission without delay,given by,

$$rtt(cwnd) = \frac{[cwnd \cdot Chunk\_size + (cwnd/2) \cdot SACK\_size]}{txmn\_rate}. \quad (4.2)$$

3) After the 1st round, the round index i increases to 1, the arrays $cwnd(0)$ and $qq(0)$ are split into two elements. If all the packets in $cwnd(0)$ are successfully transmitted, cwnd size becomes $cwnd(0)\times 2 = 4$, for slow start. If there is one or more than one packet loss, cwnd size becomes half of the current window size, i.e., $cwnd(0)/2 = 1$. Since the current window size is 2, the probability that cwnd size transits from 2 to 4 is $q^{cwnd(0)} = q^2$; the probability that cwnd size transits from 2 to 1 is $1 - q^{cwnd(0)} = 1 - q^2$. That is,

$cwnd(1) = (cwnd(0)\times 2 \;\; cwnd(0)\times 0.5) = (4 \;\; 1)$, and

$qq(1) = (qq(0)\times q^{cwnd(0)} \;\; qq(0)\times(1-q^{cwnd(0)})) = (q^2 \;\; 1-q^2)$

4) During the second round of data transmissions, we have

a) The number of accumulative successful packet transmissions until the second round is given by,

$pkt(2) = pkt(1)$ + number of successful packet transmissions at the 2nd round

$= pkt(1) + cwnd(1) \times q \times qq(1)^T = 2q + 4q \times q^2 + 1q \times (1-q^2) = 3q^3 + 3q$

b) The accumulative RTT until the 2nd round is given by,

$RTT(2) = RTT(1)$ + the round trip time of the 2nd round

$= RTT(1) + rtt(cwnd(1) \times qq(1)^T + delay$

$= RTT(1) + rtt(4) \times q^2 + rtt(1) \times (1-q^2) + delay$

We apply (4.2) to this expression and get,

$RTT(2) = rtt(3) + rtt(3)\, q^2 + 2delay.$

5) After the 2nd round, the round index i increases to 2, the arrays $cwnd(1)$ and $qq(1)$ are split into four elements. They are

$cwnd(2) = (4\times2 \;\; 4/2 \;\; 1+1 \;\; \lceil 1/2 \rceil) = (8 \;\; 2 \;\; 2 \;\; 1)$, and

$$qq(2) = (q^2\times q^4 \quad q^2\times(1-q^4) \quad (1-q^2)\times q \quad (1-q^2)\times(1-q))$$
$$= (q^6 \quad q^2-q^6 \quad q-q^3 \quad 1-q-q^2+q^3)$$

6) During the 3rd round of data transmissions, we have

a) The number of accumulative successful packet transmissions until the second round is given by,

$pkt(3) = pkt(2)$ + number of successful packet transmissions at the 3rd round

$= pkt(2) + cwnd(2) \times q \times qq(2)^{\mathrm{T}}$

b) The accumulative *RTT* until the 3rd round is given by,

$RTT(3) = RTT(2)+$ the round trip time of 3rd round

$= RTT(2) + rtt(cwnd(2)) \times qq(2)^{\mathrm{T}} + delay$

......

After the (i−1)-th round, the round index increases to (i-1). As a result of the calculations in round (i−1), we get the value of variable $pkt(i-1)$ and $RTT(i-1)$, and arrays $cwnd(i-1)$ and $qq(i-1)$. During the i-th round, we have,

a) The number of accumulative successful packet transmissions until the i-th round is,

$pkt(i) = pkt(i-1)$ + number of successful packet transmissions at the i-th round

$= pkt(i-1) + cwnd(i-1) \times q \times qq(i-1)^{\mathrm{T}}$,

b) The accumulative *RTT* until the i-th round is,

$RTT(i) = RTT(i-1)+$ the round trip time of the i-th round

$= RTT(i-1) + rtt(cwnd(i-1)) \times qq(i-1)^{\mathrm{T}} + delay$

After the i-th round, the $2^{(i-1)}$ elements in the arrays $cwnd(i-1)$ and $qq(i-1)$ are split into 2^i elements for $cwnd(i)$ and $qq(i)$, as illustrated in Fig. 4-8. For any cwnd size $cwnd$ with transition probability value of qq in $(i-1)$-th, qq splits to $qq \times q^{cwnd}$ and $qq \times (1-q^{cwnd})$ in i-th round; $cwnd$ splits to $(cwnd \times 2)$ (for slow start) or to $(cwnd+1)$ (for congestion avoidance) and $cwnd/2$ in i-th round. We have,

a) The number of accumulative successful packet transmissions until the $(i+1)$-th round is,

$pkt(i+1) = pkt(i)$ + number of successful packet transmissions at the $(i+1)$-th round

$$= pkt(i) + cwnd(i) \times q \times qq(i)^{\mathrm{T}},$$

b) The accumulative RTT until the $(i+1)$-th round is,

$RTT((i+1) = RTT(i)+$ the round trip time of the $(i+1)$-th round

$$= RTT(i) + rtt(cwnd(i)) \times qq(i)^{\mathrm{T}} + delay$$

We repeat this recursive function until a predefined round r is reached. Then, at the end of r-th round, the expected throughput $throughput(p,r)$ is equal to the cumulative number of successful packet transmissions $pkt(r)$ divided by the cumulative time taken $RTT(r)$, i.e.,

$throughput(p,r)= pkt(r)/RTT(r)$, if r is the final round number.

In summary, given initial congestion window size $cwnd(0)$, loss rate p, propagation delay, chunk and SACK size, data transmission rate and ssthresh, we are able to calculate the link throughput until the certain round.

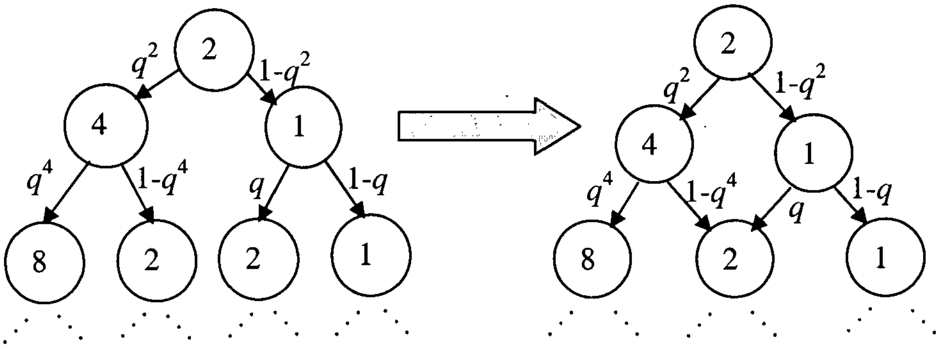


Fig. 4-9 Combine the nodes with the same *cwnd* in the same round

However, in this algorithm, for r-th round, there are 2^r elements in the arrays *cwnd*(r) and *qq*(r). Therefore, the computation work load increases exponentially with r. We propose to combine the nodes with the same cwnd size in the same round together, so that the algorithm can be improved while preserving the accuracy of the results. Fig. 4-9 illustrates this method. At the end of the second round, we combine the nodes of *cwnd* 2 together. After the combination, At the end of the second round, the array *cwnd*(2) = (8 2 1) instead of before the combination (8 2 2 1); the array $qq(2) = (q^6 \;\; q+q^2-q^3-q^6 \;\; 1-q-q^2+q^3)$ instead of $(q^6 \;\; q^2\text{-}q^6 \;\; q\text{-}q^3 \;\; 1-q-q^2+q^3)$. In this way, the number of elements in the arrays *cwnd*(i) and *qq*(i) are constrained to increase closely linearly instead of exponentially with increasing round number i. This method makes it possible to calculate the throughput for large number of rounds.

Chapter 5 Simulation Results and Discussions

In this chapter, we describe the simulation methodology in Section 5.1. We present and discuss the simulation results in Section 5.2.

5.1 Simulation Methodology

We evaluate our proposed MSCTP-based UMTS/WLAN VHO, SMART-FRX scheme and validate the developed analytical model using the ns-2 network simulator [15] with the SCTP module [16] developed by the Protocol Engineering Lab at the University of Delaware. The simulation work consists of the following steps:

1) Design and implement the proposed schemes, set up the simulation configuration.

2) Design the simulation study, and select the performance metrics used to evaluate the simulation results.

3) Perform the simulations and verification tests.

4) Analyze and verify the simulation results by evaluating and comparing different scenarios and different parameter settings.

In this section, we explain the simulation methodology in details. We give a brief overview of our simulation and computation tools, ns-2, its SCTP module, and Matlab in Section 5.1.1. Section 5.1.2 presents the design and implementation on the simulation model of MSCTP-based VHO. Section 5.1.3 describes the implementation of the proposed SMART-FRX scheme. In Section 5.1.4, we present the algorithm that we develop to implement the analytical model on Matlab. We describe the ns-2 experimental setting used to verify the proposed analytical model in this section too.

5.1.1 Overview of Simulation Tools

Network Simulator – ns-2 [15], which is available on several platforms such as FreeBSD, Linux, SunOS, Solaris and Windows, is an object-oriented, discrete event driven network simulator targeted at computer networking and protocol research. It has become one of the most popular research tools in the networking area. Ns-2 provides substantial capabilities from graphically dealing with network traffic to simulation of routing, multicast and other IP protocols. It is an open-source simulation tool primarily useful for simulating local and wide area networks. It includes implementation of network protocols such as TCP, UDP, SCTP over wired and wireless (local and satellite) networks, traffic source behaviors such as FTP (File Transfer Protocol), TELNET (TCP/IP Terminal Emulation Protocol) and HTTP (Hypertext Transfer Protocol), queuing algorithms including fair queuing, RR (round-robin) and FIFO (First In First Out), router queue management mechanisms, and routing algorithms, *etc.* Its advantages have made it a useful tool that can be used extensively in network simulations.

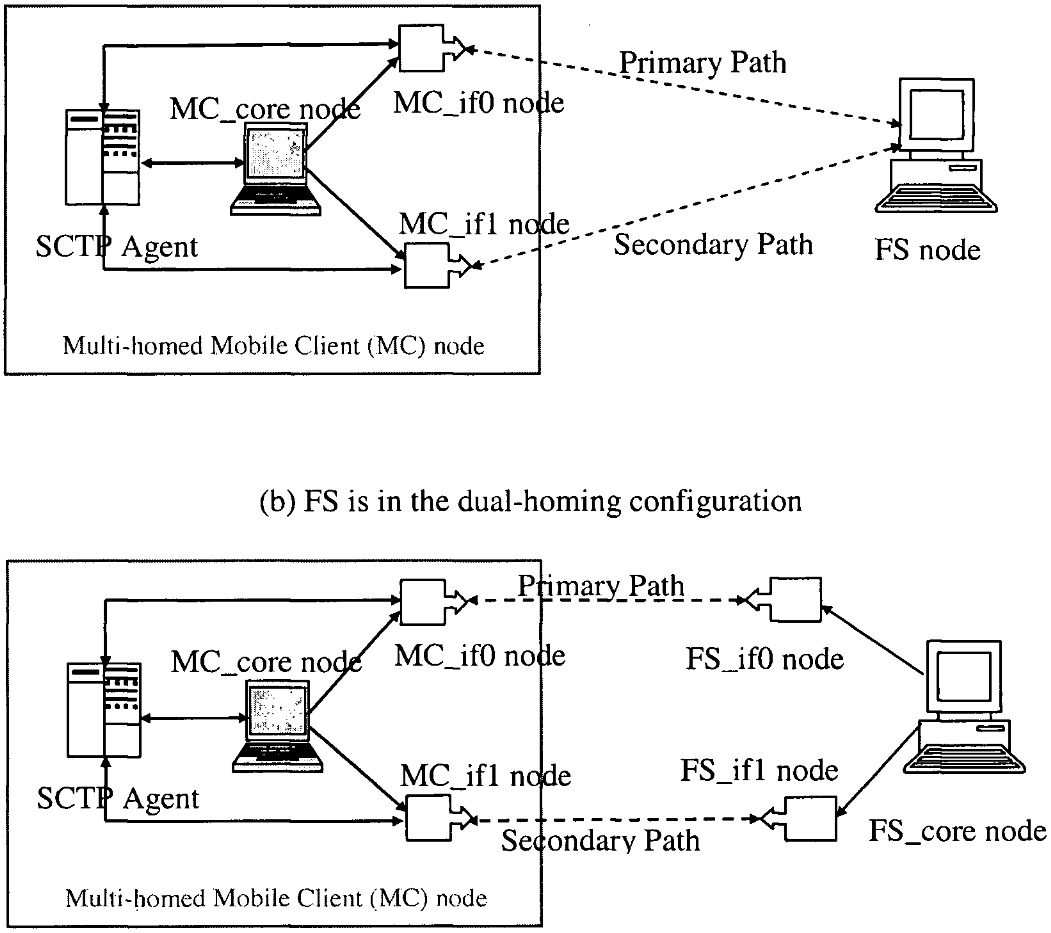
The ns-2 SCTP module [16] developed by the Protocol Engineering Lab at the University of Delaware currently supports the core features in [11] and [29]. These features with the referenced section numbers in [29] are summarized in Table 5-1. The SCTP agent establishes an association using a four-way handshake, but the handshake process is kept simple and does not strictly conform to the specification. For example, the handshake does not update RTT. Instead, RTT estimation begins with the first data

Table 5-1 Features included in existing ns-2 SCTP module [39]

Section in	**Feature description**
5.1	Normal establishment of an association
6.1	Transmission data chunks
6.2	Acknowledgement of reception of data chunks
6.3	Management retransmission timer
6.4	Multi-homed SCTP endpoints
6.5	Stream identifier and stream sequence number
6.6	Ordered and unordered delivery
6.7	Report gaps in received data TSNs
7.2	SCTP slow-start and congestion avoidance
8.1	Endpoint failure detection
8.2	Path failure detection
8.3	HEARTBEAT function without upper layer control

chunk. SCTP packets generated by the SCTP agent and destined to a peer SCTP agent over a traced link generate a trace file.

Matlab is an interactive program for numerical computation and data visualization; it is used extensively for analysis and design in research and engineering. We use the Matlab 6.5 to analyze and display the simulation results graphically. We implement our analytical model with the Matlab and validate the analytical model against ns-2 simulations.


(a) FS is in the single-homing configuration

(b) FS is in the dual-homing configuration

Fig. 5-1 MSCTP-based VHO configurations

5.1.2 The Simulation Model of UMTS/WLAN VHO Using MSCTP

The simulation configuration to test MSCTP-based UMTS/WLAN VHO is shown in Fig. 5-1. Fig. 5-1(a) shows the single-homing configuration and Fig. 5-1(b) shows the dual-homing configuration. Due to limitations of ns-2's architecture, a node cannot actually be multi-homed. To get around this limitation, each multi-homed node is actually made up of more than one nodes. As shown in Fig. 5-1(a) and (b), a multi-homed node has a "core node" and multiple "interface nodes" to simulate the interfaces. The SCTP agent resides in all the nodes, but traffic only goes to and from the interface nodes. The core node is used for routing and is connected to each interface node via a uni-directional link towards the interface node, but traffic does not traverse this link. Instead, these links are used to dynamically determine which outgoing interface node to use for sending to a particular destination. The SCTP agents are two-way agents, which means that they are symmetric in the sense that they represent both a sender and a receiver.

MSCTP is SCTP with implementation of DAR [31]. It is an optional extension that is introduced to provide SCTP with the ability to reconfigure IP address information on an existing association. In details, the functions of DAR are: 1) A graceful method to add or delete interfaces or IP addresses to an existing association; and 2) A method for an endpoint to request its peer to set the primary destination address. SCTP together with the DAR extension makes it possible that an endpoint changes its one or multiple addresses dynamically without affecting the established association. Ns-2 SCTP module [16] supports multi-homing shown in Table 5-1, but it does not support DAR. In order to perform the simulations and obtain the results reported in this thesis, we extend the SCTP module so that the multi-homing feature can work over wireless links. As shown in Fig. 5-1, a multi-homed MC actually consists of two MC_if nodes that perform as the two interfaces of the MC, and an MC_core node that is a wired node and performs as a control part of the MC to determine the packet route. All these nodes are connected to SCTP agent to trigger SCTP traffic as the source and handle the incoming SCTP traffic as the destination. The design of the MC_if node is based on the nodes called *MobileNode* and *BaseSation* in the current ns-2 wireless communication module. The existing *MIPBSAgent* and *MIPMHAgent* are modified to support a new type of routing

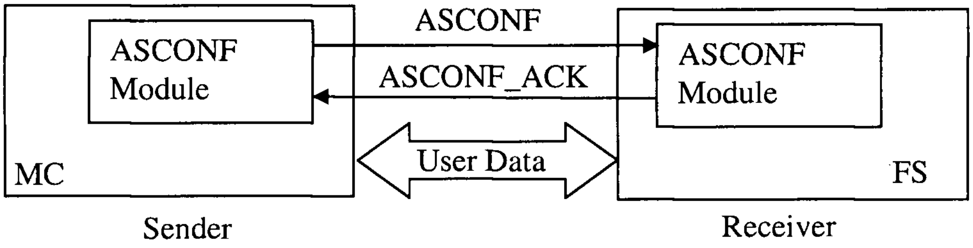


Fig. 5-2 MC and FS interaction with ASCONF and ASCONF_ACK chunks

agent called Non-Ad-Hoc routing agent (NOAH) [40]. After introducing this new agent, the MC_if interface nodes act as the SCTP traffic source that can be controlled by the MC_core node. The ns-2 IEEE 802.11 WLAN module is used to support the MAC (Medium Access Control) layer. The bandwidths (data transmission rates) are set to be 384 Kbps for the UMTS link and 2 Mbps for the WLAN link.

We implement two new messages that have the same functions as ASCONF and ASCONF_ACK chunks so that we can add or delete an IP address dynamically into or from the current active association. Fig. 5-2 shows the communication between the MC and FS endpoints with the ASCONF and ASCONF_ACK chunks. As the address reconfiguration requester, the MC creates an ASCONF chunk with a parameter of "Add" or "Delete" or "Set Primary Address". When the endpoint FS receives an ASCONF chunk from the remote MC peer, the processing of the ASCONF chunk begins:

- If the Address Parameter in the ASCONF chunk is "Add" or "Delete", the specified address is added to or deleted from the destination list on the FS.
- If the Address Parameter in the ASCONF chunk is "Set Primary Address", the specified address in the chunk is set to be the primary address by FS.

With the implementation of the MSCTP function in the test network, after an SCTP association is initiated, under the coordination of SCTP agent and routing agent, FS starts to send FTP traffic to MC on the primary link. In the test network, we set the network propagation delay to be 120 ms, chunk size to be 1468 bytes, MTU to be 1500 bytes. We set the sending queue limits and the receiver window size to be large enough to ensure that they do not interfere the SCTP performance during a VHO. We initiate an association between the FS and MC endpoints at 0 s and trigger the FTP traffic on the primary WLAN (or UMTS) link from the FS to the MC at 1 s. To measure the MSCTP supporting VHO performance, we trigger a VHO at 5 s so that an ASCONF message is sent from the MC to request to switch the primary link to the secondary link. We record the TSN for each of the data chunks and monitor the overall throughput on both links at the receiver MC endpoint. The overall throughput is defined as the total number of distinct data chunks received from both links divided by the total elapsed time. We measure and analyze the VHO delay and the throughput at MC for the single-homing and the dual-homing configurations, and study the impact of different configurations on the VHO performance.

5.1.3 Implementation of SMART-FRX

Fig. 5-3 shows the implementation architecture of SCTP. The SCTP instance module is composed of two basic functions: the SCTP sending function and the SCTP receiving function. Both these functions are implemented in the SCTP Agent. Above the sending module is the possible applications that use the services provided by SCTP such as FTP, TELNET, HTTP, *etc.* The SCTP data transmission mechanisms, retransmission

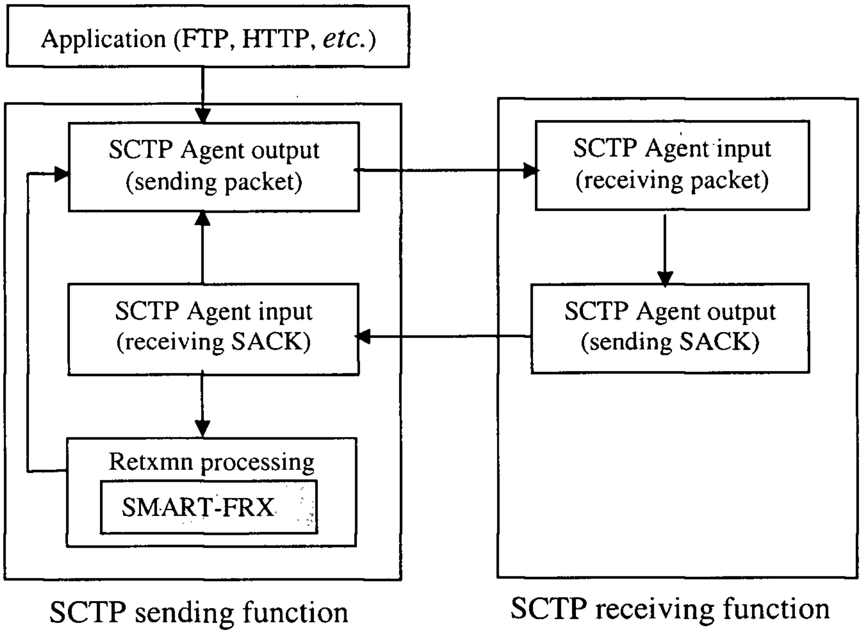

Application (FTP, HTTP, *etc.*)
SCTP Agent output (sending packet)
SCTP Agent input (receiving SACK)
Retxmn processing
SMART-FRX
SCTP Agent input (receiving packet)
SCTP Agent output (sending SACK)
SCTP sending function
SCTP receiving function

Fig. 5-3 Implementation architecture of SCTP

timers and timeout processes are part of the sending function. The SCTP specification does not allow a sender to multicast data on multiple paths. An SCTP sender maintains a primary destination to which all new data are sent. Only retransmissions are sent to the alternate destinations. However, we have seen that this mechanism is not suitable in the WLAN to UMTS forced VHO situation when there are consecutive packet losses and timeouts due to HL. This is the motivation behind the proposed SMART-FRX scheme. To implement the proposed Sending-buffer Multicast Aided Retransmission (SMART-FRX) scheme, we assume that:

- Multiple (usually two) independent paths exist between the MC and the FS.
- Receiver's advertised window (rwnd) at receiver side (the MC) is large enough not to constrain the sender.

- There is no queue length limitation for the sending-buffers of both WLAN and UMTS links.

These assumptions enable us to study the changes in cwnd at the sender, the cmTSN (cumulative TSN) and the TSN Gaps in SACK [29] and other congestion control parameters independently for each destination.

The SMART-FRX module is implemented in the retransmission processing module of the SCTP implementation architecture shown in Fig. 5-3. The entry point to activate the SMART-FRX module is that the WLAN link encounters a transmission TO that is marked by the WLAN error counter n being set to one. Without the SMART-FRX scheme, new traffic is always queued on the primary link and only the retransmitted data are queued on the secondary link. The SMART-FRX module enables the function that once the sender detects the WLAN link entering a possibly unreachable period, the sender copies the data in the sending buffer of the primary connection over the WLAN link to that of the secondary connection over the UMTS link, and enqueues new data chunks in both sending-buffers. Thus, the buffered data as well as any new data are able to multicast on both the primary and secondary links during the WLAN possibly unreachable period. Though the bandwidth of both links are different, the multicasting increases the data-sending rate that would have been reduced because of TO process on the primary connection. On the other hand, at anytime, if there is any packet loss detected by FD on the UMTS link, the retransmission is forced to be sent to the same destination. When the WLAN link possibly unreachable period ends, which is determined by either the WLAN link error counter n resetting to zero or reaching Path.Max.Retrans, the sender enqueues new data chunks to the current primary link only.

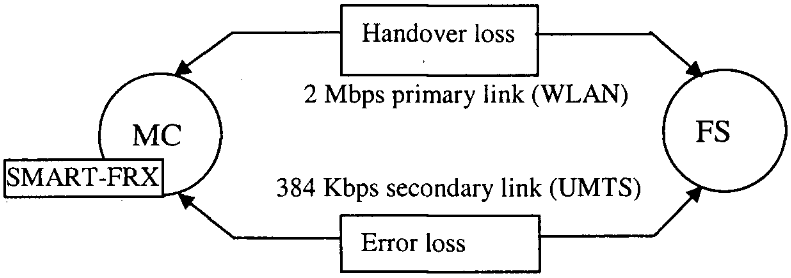


Fig. 5-4 SMART-FRX simulation configuration

The simulation configuration to verify the effectiveness of the SMART-FRX scheme is shown in Fig. 5-4. Two packet-drop modules are inserted to the WLAN and the UMTS links, respectively. The parameters of the network propagation delay, the chunk size and the MTU size are set to be the same as that of Section 5.1.2. On the WLAN link, the packet-drop module drops consecutive packets so that the WLAN link encounters a TO and the error counter n becomes 1,2,...6 (Path.Max.Retrans), respectively. We monitor the TSN of each data chunks and the throughput performance for each the error counter value at the receiver with and without application of the SMART-FRX schemes. In order to make sure the correctness of the results, each scenario is tested for 10 times. We compare the delay and the throughput performance with and without the proposed scheme to ensure the effectiveness of the proposed scheme.

5.1.4 Validation of Analytical Model

In Chapter 4, we developed an analytical model in order to study the throughput performance on the UMTS link with EL when the WLAN link is in possibly unreachable period. In this chapter, we use ns-2 to validate this analytical model.

The pseudo code to implement the algorithm on Matlab 6.5 is attached as Appendix A. The simulation configuration on ns-2 is shown in Fig. 5-4. The simulation is with the SMART-FRX implemented. The implementation of the error loss module on the UMTS link would be critical to the relevance of any results obtained from the experimental set-up. We use the random loss error model in ns-2 to simulate the transmission of fixed size packets over a fading channel. The parameters of packet transmission error rates *Pe* are given by 0, 0.01.0.025, 0.05, 0.075, 0.1 and 0.11. For each of the packet loss rate, we run the simulation for 10 times and calculate the average throughput to observe and plot the relationship between the throughput and the packet loss rates. We change the test scenario by setting different network propagation delays as 0 ms, 50 ms, 120 ms, 150 ms, 250 ms and 400 ms. The experimental results are compared with the analytical predictions generated by the Matlab. The average error is given by:

$$Error_{avg} = \frac{\sum_{simulations} \dfrac{| throughput_{analytical} - throughput_{simulation} |}{throughput_{simulation}}}{\text{number of simulations}} \tag{5.1}$$

The average error shows the mean difference between the analytical results and the simulation results. A smaller average error indicates a better model accuracy. In this way, the accuracy of the analytical model can be validated.

5.2 Results and Discussions

We show our simulation results in this section. We analyze, compare and verify the simulation results with different scenarios and parameter settings. The objectives of the simulations are to evaluate the performance of our proposed schemes and validate the developed analytical model. In Section 5.2.1 we study the simulation results of the MSCTP-based UMTS/WLAN unforced VHO scheme. We analyze two critical parameters, the UMTS/WLAN VHO delay and the throughput for both FS in single-homing and dual-homing configurations. In Section 5.2.2, we observe the WLAN link TO and the UMTS link retransmission situation when a mobile user leaves the WLAN coverage and triggers a forced WLAN to UMTS VHO. We illustrate and compare the throughput performance before and after application of the SMART-FRX scheme and show the effectiveness of this scheme. Finally, in Section 5.2.3, we show the results of using ns-2 random loss error model to validate and verify the developed analytical model.

5.2.1 UMTS/WLAN Unforced VHO

In this experiment, the bandwidth between the MC and the FS endpoints are set to be 384 Kbps for the UMTS link and 2 Mbps for the WLAN link. The network propagation delay is set to 100 ms. FTP traffic is triggered from MC at time 1 s. VHO Triggering process is activated at time 5 s. Data transmission is from MC to FS.

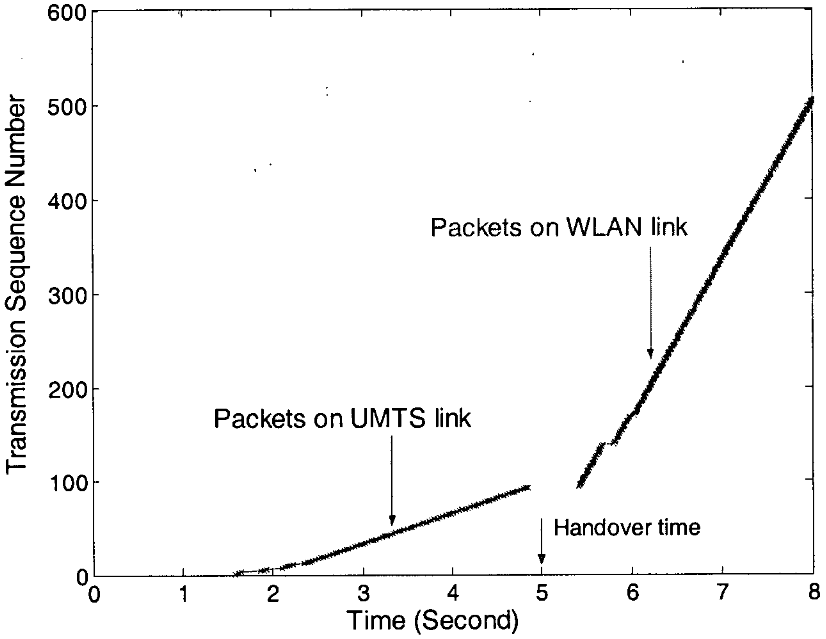
Figs. 5-5 and 5-6 show the delay performance for the VHO from UMTS to WLAN and for the reverse direction, respectively. For FS in single-homing configuration, we define the VHO delay as the time difference between FS receives the first packet on the new primary link and the last packet on the old primary link. For FS in dual-homing configuration, assuming the first packet received by FS is TSN *first*, the VHO delay is defined as the time difference between FS receives TSN *first* on both links. According to the simulation results, when the FS is in the single-homing configuration, the UMTS to WLAN handover delay is 533 ms in Fig. 5-5(a) and WLAN to UMTS handover delay is 513 ms in Fig. 5-6(a). When FS is in the dual-homing configuration, these two handover delays are reduced to 234 ms in Fig. 5-5(b) and 212 ms in Fig. 5-6(b), respectively. This is because when the FS is in the single-homing configuration, the MC sends a "Set Primary Address" request to trigger a handover, thus increasing the overall delay with a handshake processing time. However, when the FS is in the dual-homing configuration, the MC can trigger a handover by directly setting the FS's secondary address. Therefore the handover delays in both directions are reduced significantly.

Fig. 5-7 shows the throughput performance for VHO in both directions. We can see that the throughput (number of bits received/time) of FS in dual-homing configuration is much higher than that of FS in the single-homing configuration. This is because, besides the delay advantage, a dual-homing FS allows both MC and FS to operate in a symmetric multi-homed configuration. This configuration enables easy distinction of the two paths between the MC and the FS, so that the redundant path can help to provide the fault tolerance to the data transmissions during handover. In the simulations, the buffered data are sent at both old and new connections when a changeover occurs. In this way, packet loss and retransmission delay can be avoided. Duplicate packets can be dropped by the receiver, and different strategies may be employed by the sender and receiver to adapt flow, congestion and other QoS control parameters easily and quickly during and after the handover. In Fig. 5-7, we also observe that SCTP readily copes with the sudden change of the link bandwidth during a handover.

5.2.2 SMART-FRX Scheme

The objective of these simulations is to evaluate the potential performance improvement of the proposed SMART-FRX scheme. We show the effectiveness of the proposed scheme by comparing the throughput performance with and without the

(a) FS is in the single-homing configuration



(b) FS is in the dual-homing configuration

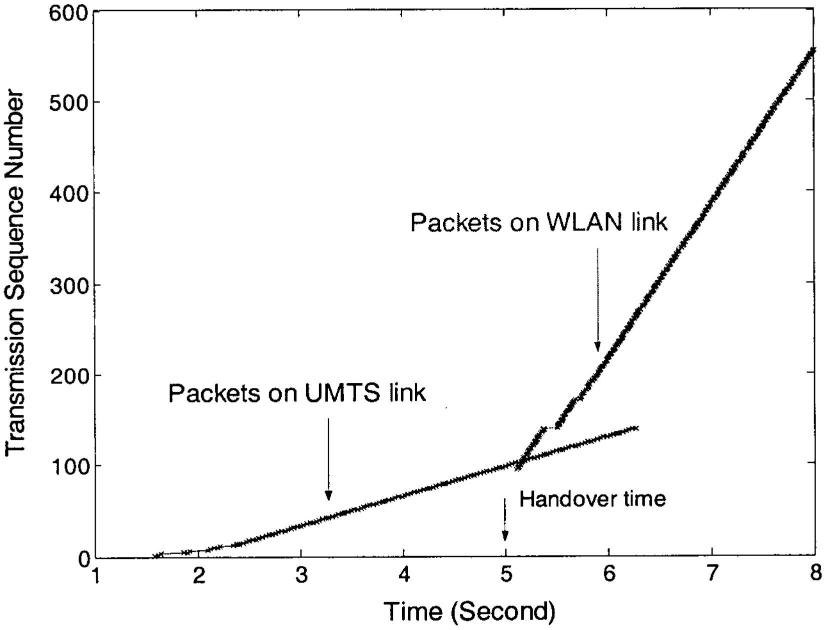
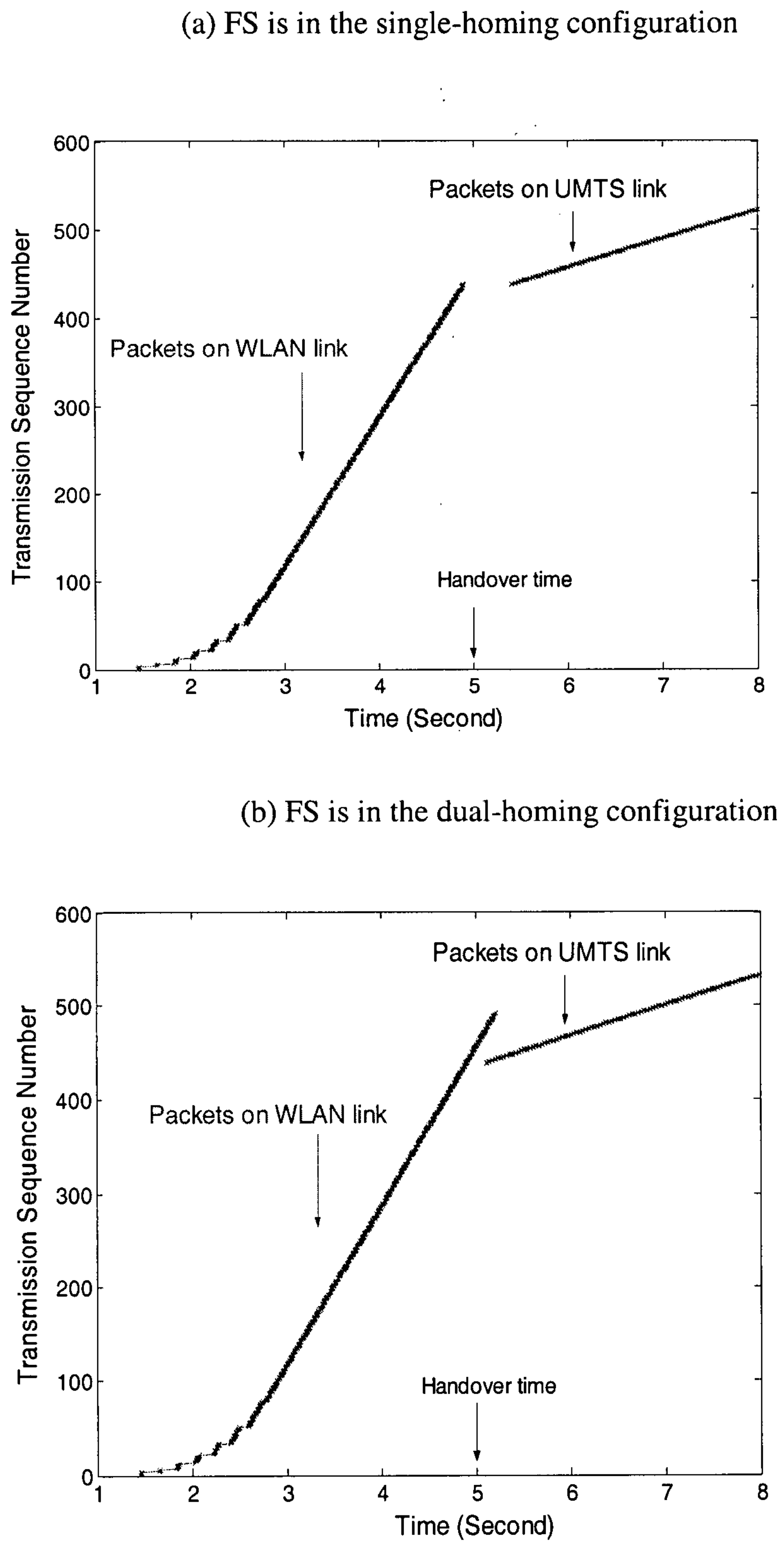


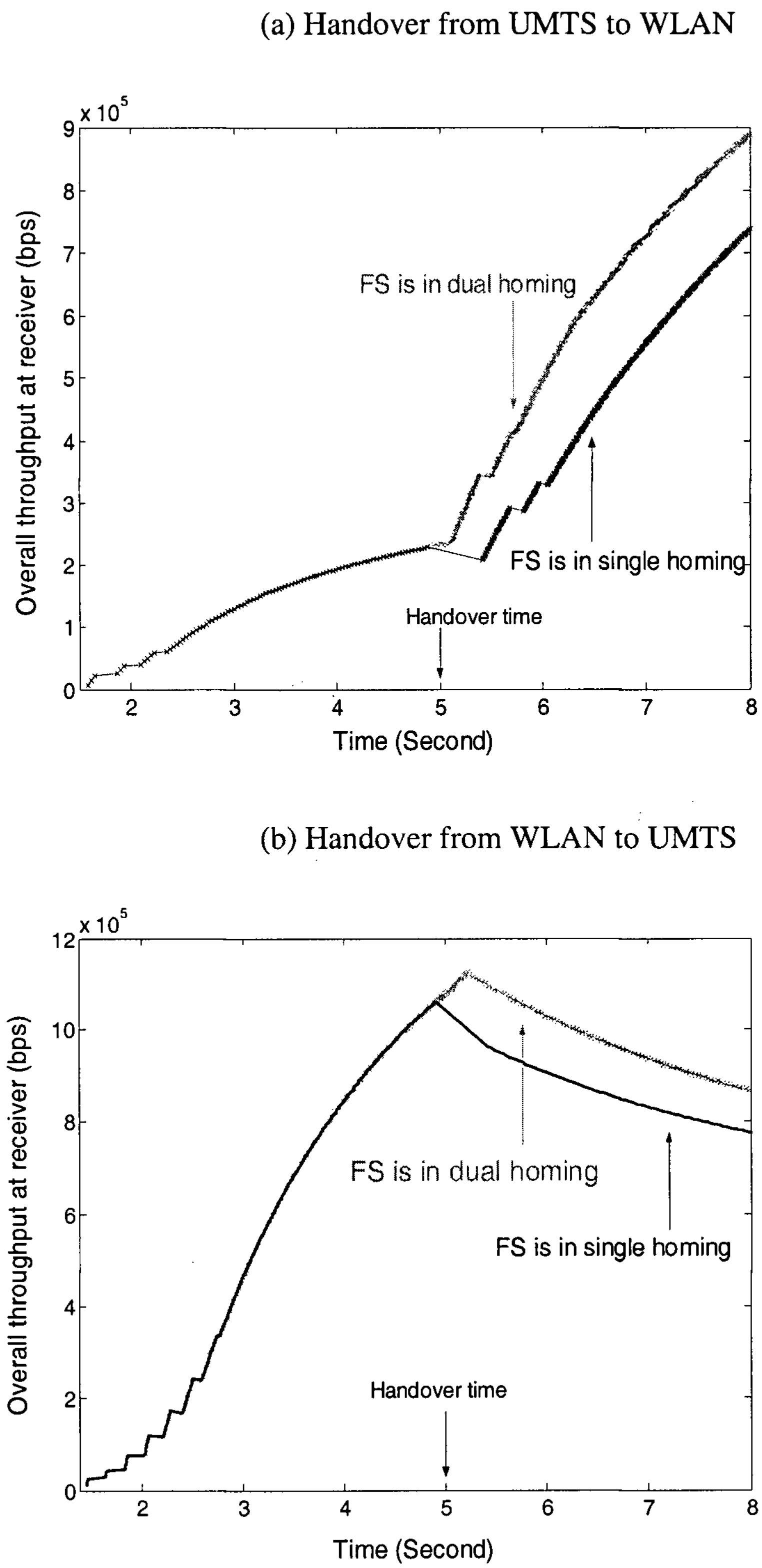
Fig. 5-5 Delay performance of the proposed VHO scheme (from UMTS to WLAN)



(a) FS is in the single-homing configuration

(b) FS is in the dual-homing configuration

Fig. 5-6 Delay performance of the proposed VHO scheme (from WLAN to UMTS)



(a) Handover from UMTS to WLAN

(b) Handover from WLAN to UMTS

Fig. 5-7 Throughput performance of the proposed vertical handover scheme

proposed scheme. The configuration used in this simulation is a dual-homing configuration as shown in Fig 5-4. The threshold Path.Max.Retrans is set to 5 as specified in the SCTP standard. The other parameters are set as the following: RTO on WLAN link is 1 s, MTU is 1500 bytes, data chunk size is 1468 bytes, and the network propagation delay is 120 ms.

Figs 5-8 to 5-11 show the received TSNs on both WLAN and UMTS links, and the throughput comparison with and without the proposed scheme. In each of the figure, (a) shows the received TSNs without the SMART-FRX scheme; (b) shows the received TSNs with the scheme; and (c) compares the throughput with and without the proposed scheme. The throughput is defined as the total number of bits received at the receiver divided by the total time elapsed.

As shown by (a) and (b) in all these figures, at the beginning, the data traffic is on the primary WLAN link. We drop a block of chunks from TSN 214 to 260 on the WLAN link at time 3.9 s, so that the error counter on this link can be set to 1. After about 1 s, i.e., at about 4.95 s, the TO of TSN 214 is detected. Then, on the WLAN link, the cwnd size becomes 1, the *T3-rtx* increases to two RTO, i.e., 2 s, and the link enters into a failure detection period. Immediately after the error counter becomes to non-zero, TSN 214 to 260 are retransmitted on the alternative UMTS link. At the same time, on the WLAN link, with the new *T3-rtx* and the new cwnd value, TSN 261 is sent to test if the WLAN link has recovered. If we keep on dropping the chunks from TSN 261 to 266, we can make the error counter *n* to be increased to 2 up to 6. According to our observations on the TSN, with and without SMART-FRX, with the error counter increasing, data

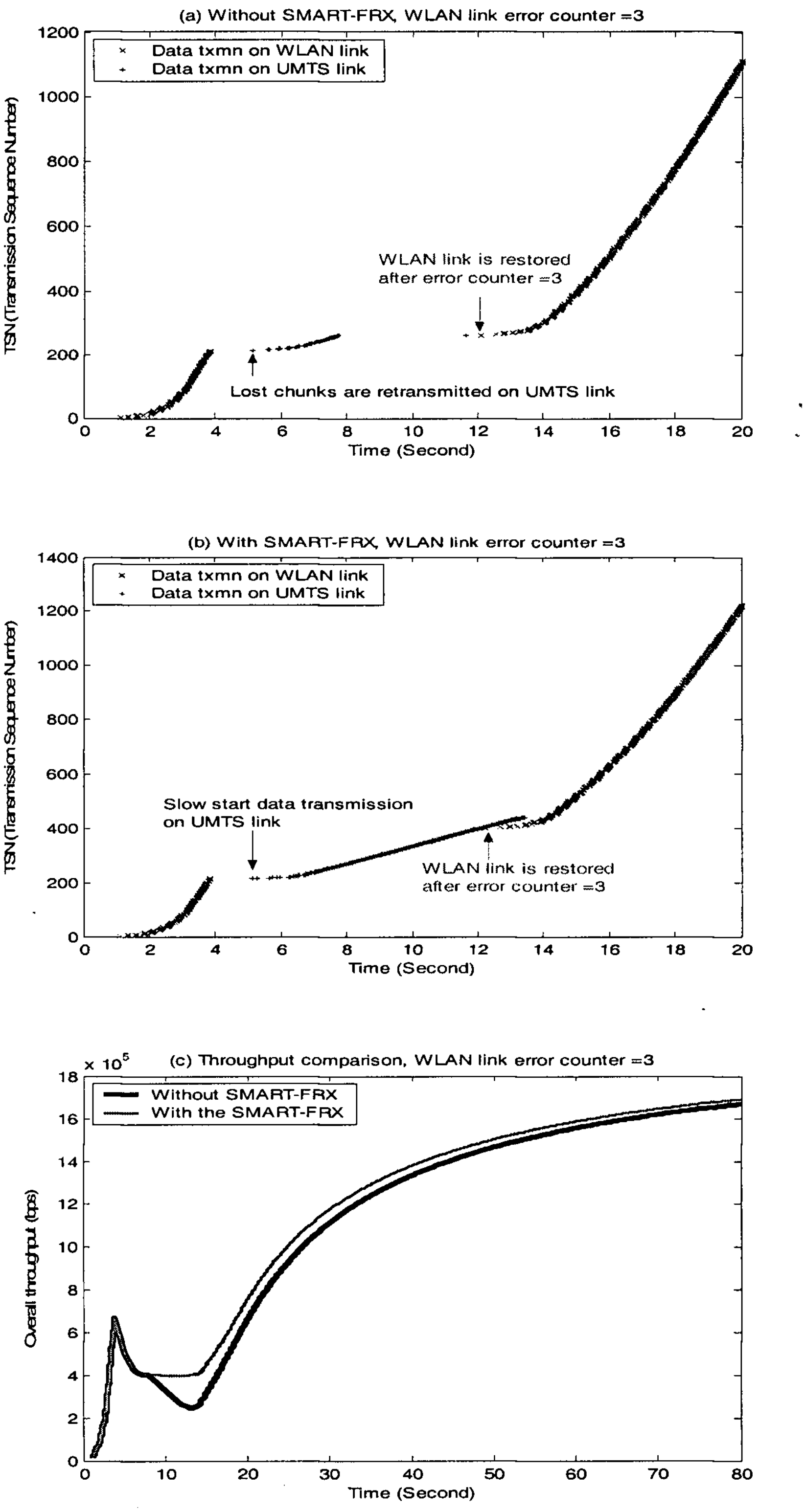


Fig. 5-8 Comparison with and without SMART-FRX ($n = 3$)

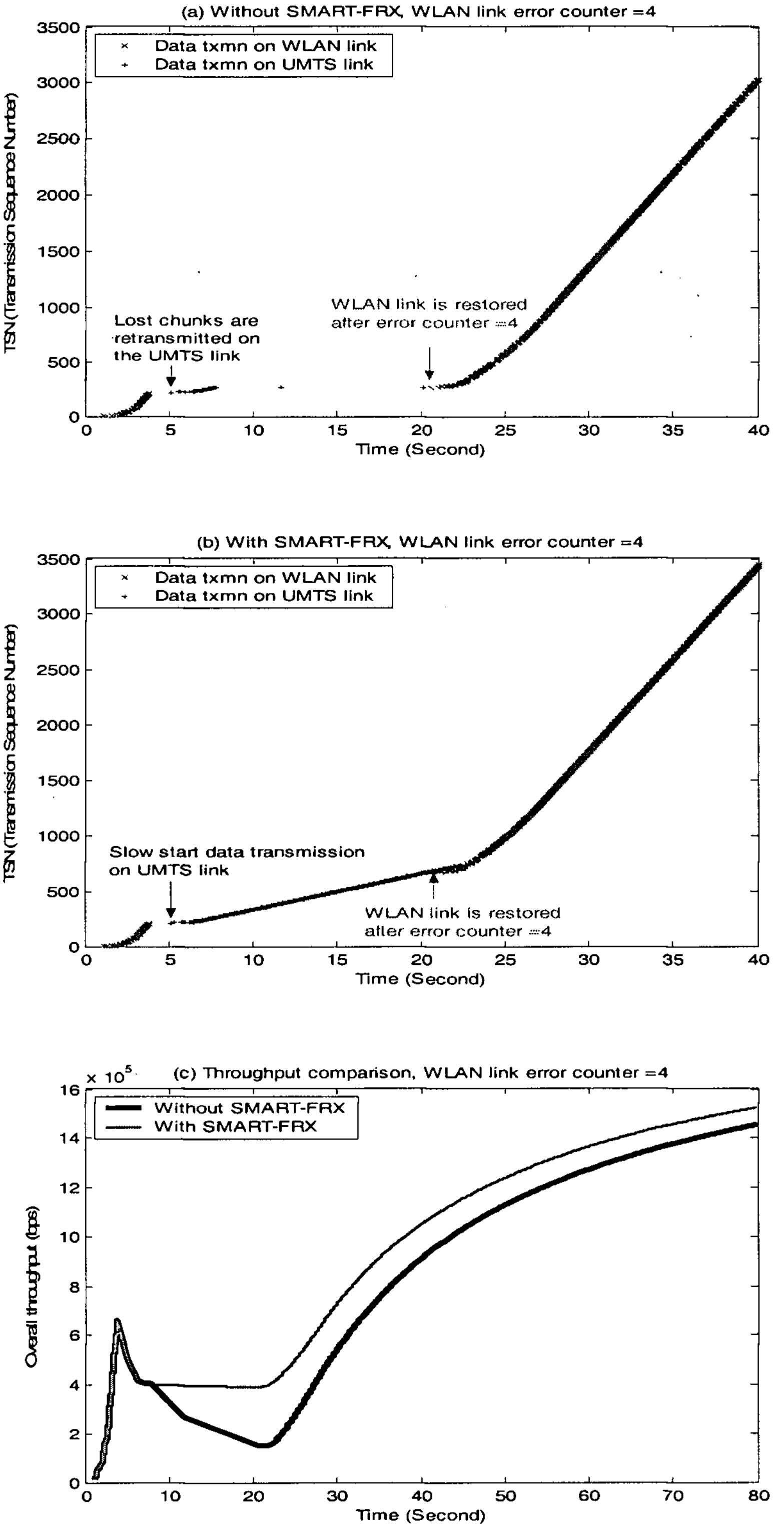


Fig. 5-9 Comparison with and without SMART-FRX ($n = 4$)

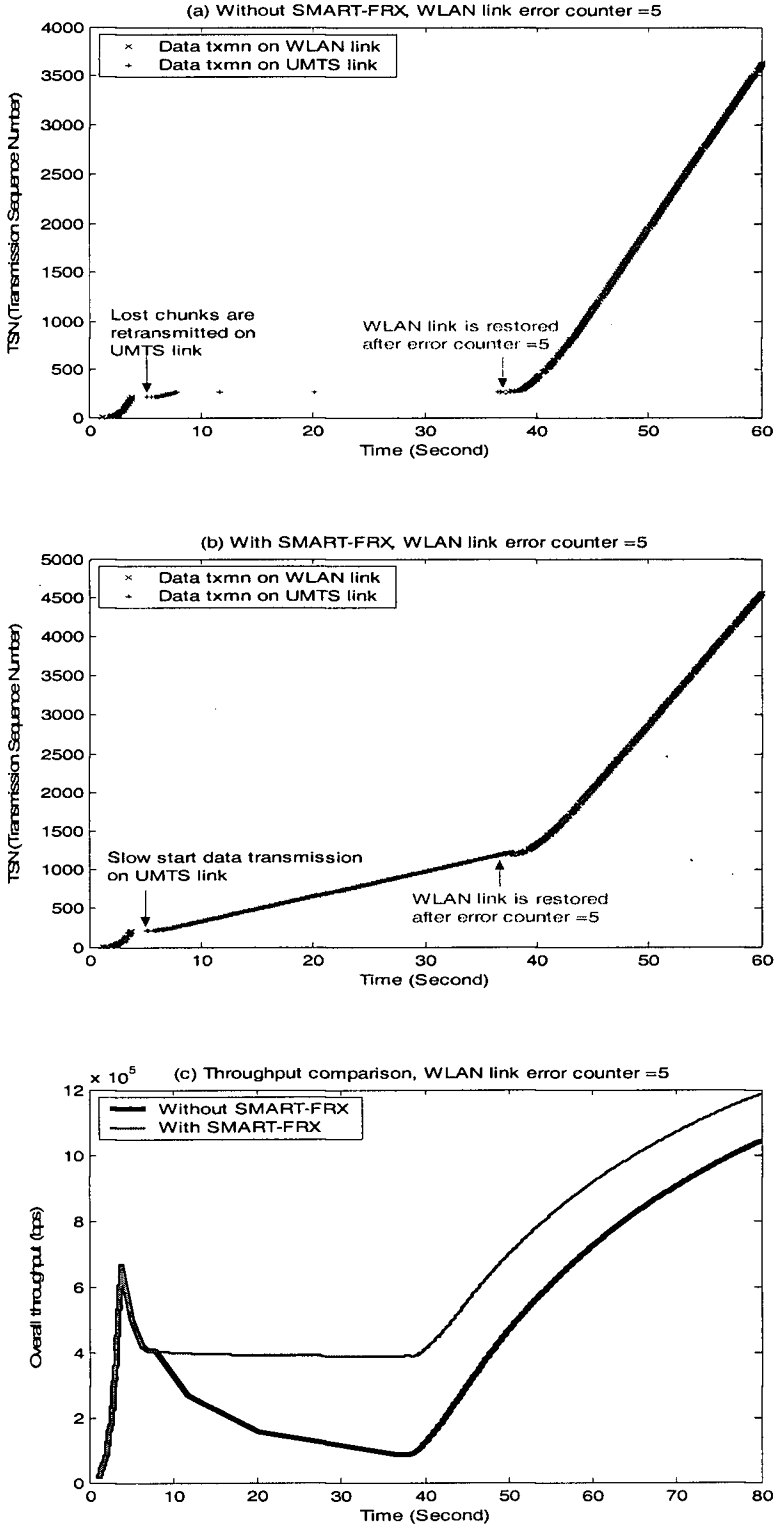
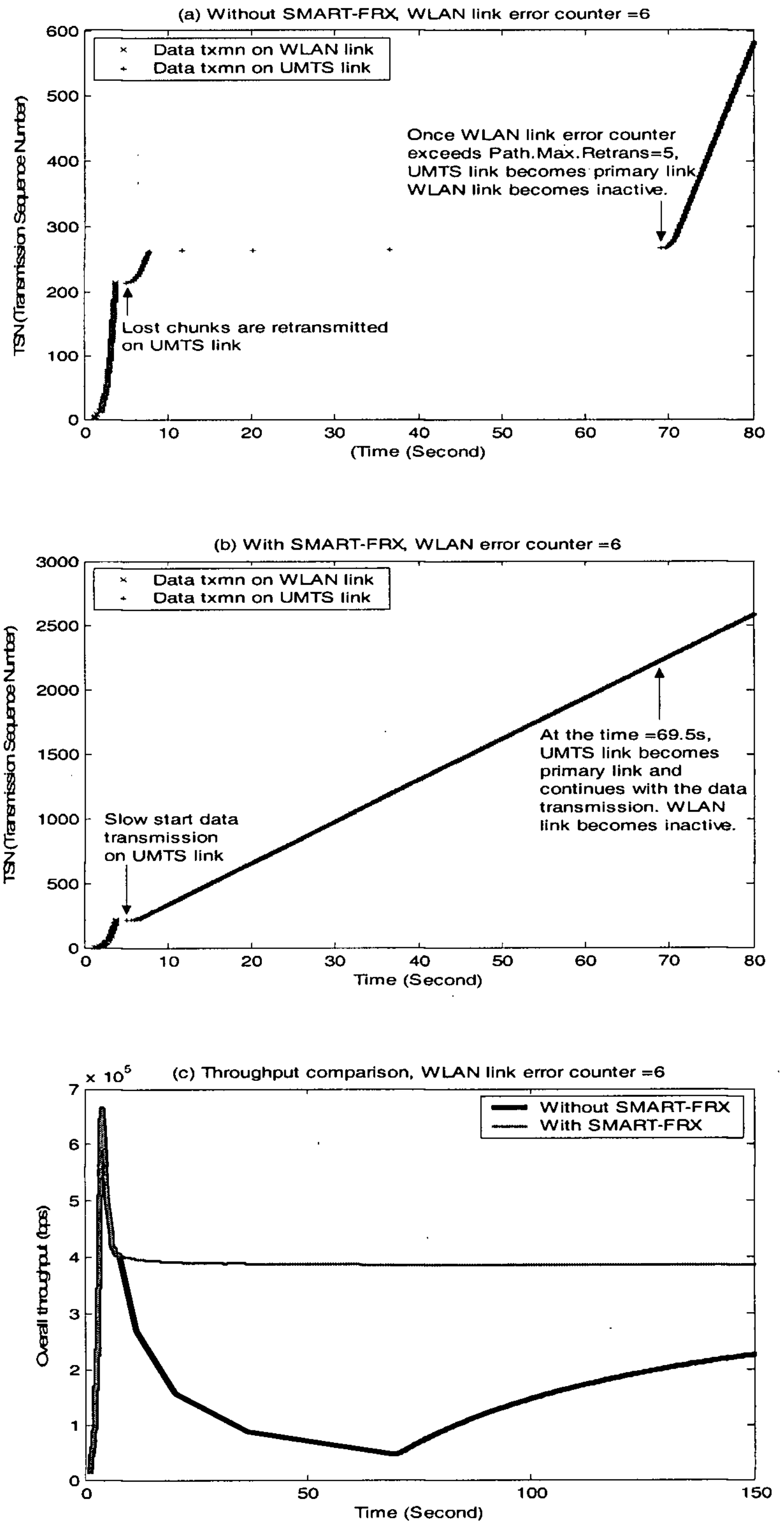


Fig. 5-10 Comparison with and without SMART-FRX (*n* = 5)



(a) Without SMART-FRX, WLAN link error counter =6

Once WLAN link error counter exceeds Path.Max.Retrans=5, UMTS link becomes primary link, WLAN link becomes inactive.

Lost chunks are retransmitted on UMTS link

(b) With SMART-FRX, WLAN error counter =6

At the time =69.5s, UMTS link becomes primary link and continues with the data transmission. WLAN link becomes inactive.

Slow start data transmission on UMTS link

(c) Throughput comparison, WLAN link error counter =6

Fig. 5-11 Comparison with and without SMART-FRX ($n = 6$)

transmissions on the WLAN link become disrupted until either of the following happens,

1) A SACK over WLAN link is received at the sender before the WLAN error counter *n* exceeds the Path.Max,Retrans. In this case, the error counter *n* resets to 0, the WLAN link enters into slow start and the traffic falls back to the WLAN link. Figs. 5-8 to 5-10 show the SCTP data transmission situation for this case.

2) The WLAN error counter *n* exceeds the Path.Max.Retrans. In this case, data transmissions are switched to the new primary link, the UMTS link. Fig. 5-11 shows SCTP data transmission situation for this case.

As shown in Figs. 5-8 (c) to 5-11 (c), SMART-FRX does make difference on the throughput performance for $n = 3$ to 6:

- Without SMART-RFX, due to the backoffs on the WLAN link, the secondary UMTS link is always waiting for data from the WLAN link for retransmission. The stop-and-wait on the WLAN link affects the utilization of the secondary link.

- When the SMART-FRX is applied, the sender is able to multicast the buffered data on both the WLAN and UMTS links. Therefore, the UMTS link can enter into a slow start data transmissions process instead of a stop-and-wait process following with the WLAN link. Contrary to a sudden drop of the overall throughput, the overall throughput can be increased significantly because the UMTS link becomes much more aggressive. Figs. 5-8 (c) to 5-11 (c) show the throughput quickly converges to the UMTS data rate 384 Kbps with SMART-FRX.

Figs. 5-8 to 5-11 also show that the larger the error counter n becomes, which means the longer time the MC staying outside of WLAN coverage, the more obvious the effectiveness of the SMART-FRX scheme. In Fig. 5-11, for both cases with and without SMART-FRX, when the WLAN link error counter n exceeds the Path.Max.Retrans at the time of 69.86475 s, the WLAN link is switched to be from "Active" state to "Inactive" state. A forced changeover makes the UMTS become the primary connection and data are only queued on the new primary link for transmission.

From our simulations, we also notice that the SMART-FRX has no effect for n=1 and 2. The throughput performance with and without SMART-FRX are similar. This is because the transmission rate of the WLAN link is much higher than that of the UMTS link. Therefore, when the first block of data TSN 214 to 260 are retransmitted on the UMTS link, the data transmission needs more time than that over the WLAN link. When the error counter is small, the WLAN link is restored before the UMTS link finishes the data transmission for TSN 214 to 260. Therefore, the throughput with and without the SMART-FRX scheme are the same.

5.2.3 Comparison with Analytical Results

In order to validate the proposed analytical model, we focus our performance study on the UMTS link. We firstly compare the throughput performance predicted by the developed analytical model with that of simulation results when there is no EL on the UMTS link. Figs. 5-12 to 5-13 show this comparison for the cases with and without the SMART-FRX scheme. We count the beginning of the WLAN possibly unreachable period as the beginning of the time-axis. Then, the UMTS link starts to retransmit data at the time of 1 s (the RTO of the WLAN link is set to be 1 s). Y-axis is the throughput defined as the total number of chunks SACKed at the sender being divided by the total time elapsed.

According to the analysis in Chapter 4, the WLAN link takes a total of 63 s to detect a link failure, i.e., the WLAN link error counter *n* increases to 6. As shown in Fig. 5-12, we see that, without the SMART-FRX scheme, both simulation results and analytical model predictions show the throughput dropping at the time that the WLAN link error counter *n* becomes 2. With SMART-FRX, Fig. 5-13 shows the data transmissions on the UMTS link with the WLAN error counter *n* increases the Path.Max.Retrans. We see that with the error counter increasing, instead of dropping during the link failure detection period, the throughput value from both the analytical model and simulations converge to the maximum data rate of the UMTS link (384 Kbps or 32 packets per second). We can see that for both cases of with and without the SMART-RFX scheme, the analysis results match the simulation results fairly well in all the tests.

We have compared the analytical and the simulation results without EL on the UMTS link. We now consider the affects of EL on the accuracy of the analytical predictions. The following validation test is with the SMART-FRX implemented. As shown in Fig. 5-4, we employ a random loss error model to simulate EL on the UMTS link. Note that the results of both the analytical model and the simulation are based on normal operating conditions with a negligible probability of both WLAN and UMTS links falling into TO retransmission at the same time. In each figure of Figs. 5-14 to 5-19, we show the throughput comparison by the analysis and by the simulation over different

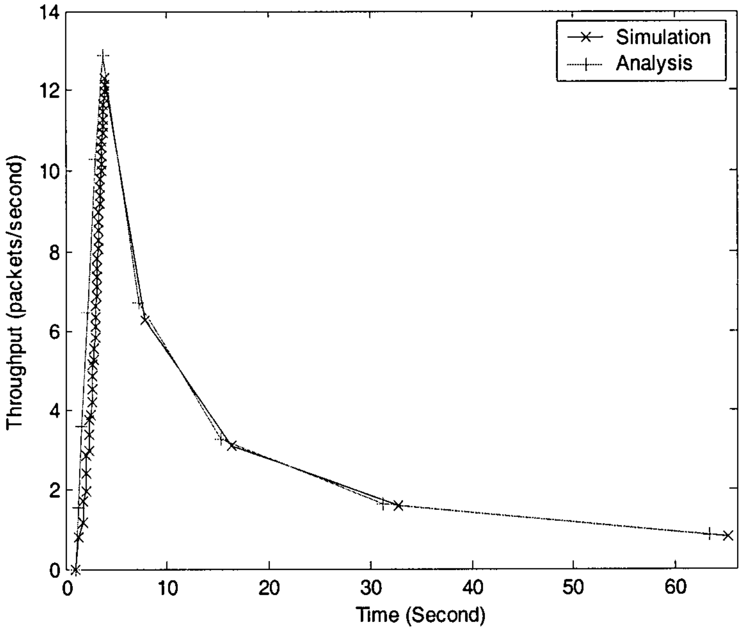


Fig. 5-12 Comparison of analysis and simulation results (without SMART-FRX)

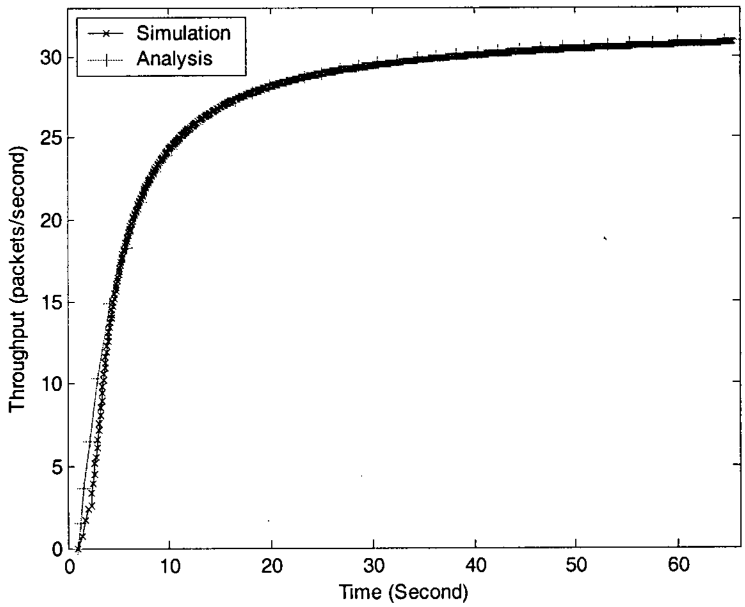


Fig. 5-13 Comparison of the analysis and simulation results (with SMART-FRX)

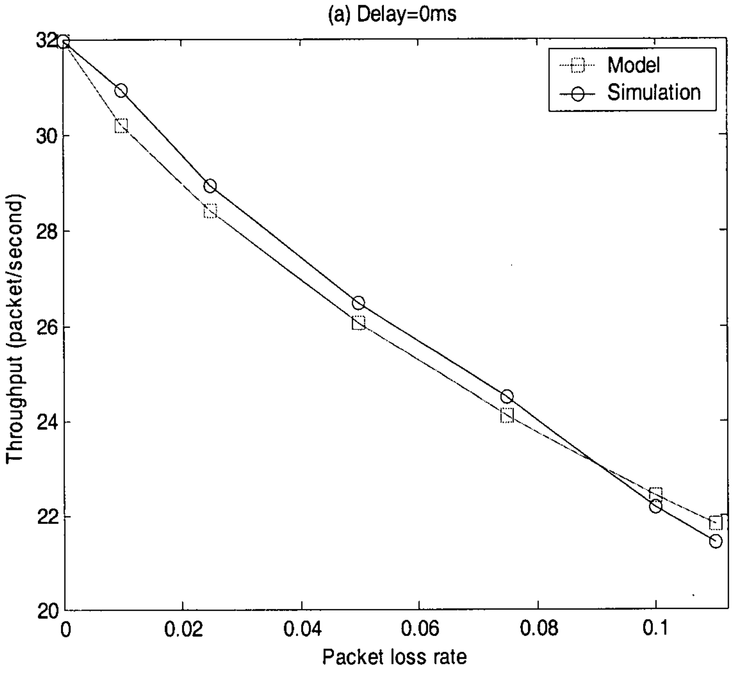


Fig. 5-14 Throughput vs. packet loss rate of UMTS link (Delay = 0 ms)

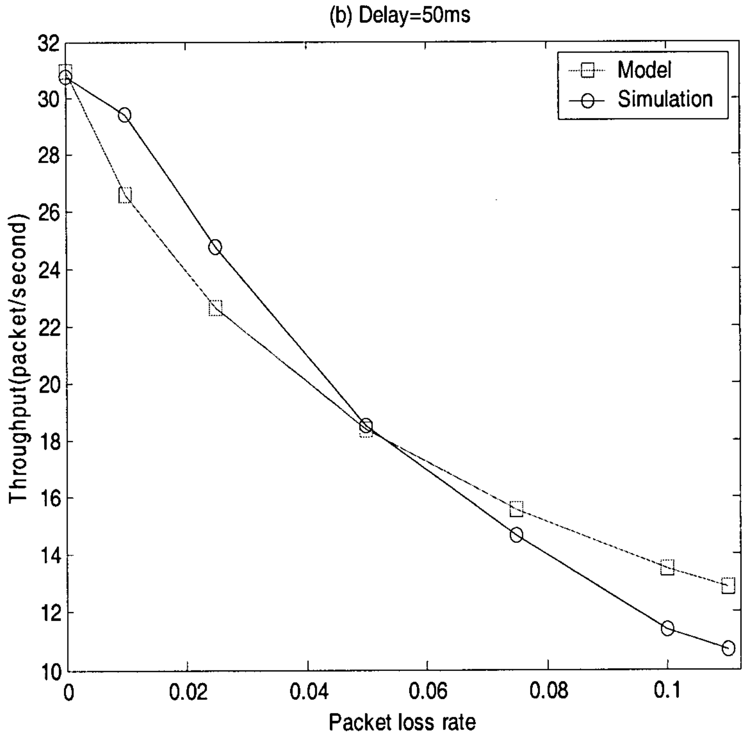


Fig. 5-15 Throughput vs. packet loss rate of UMTS link (Delay = 50 ms)

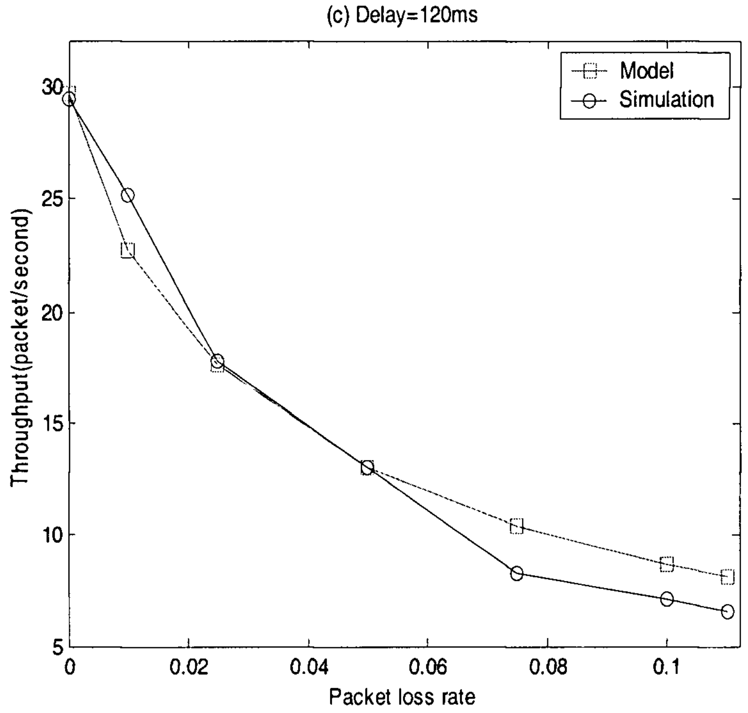


Fig. 5-16 Throughput vs. packet loss rate of UMTS link (Delay = 120 ms)

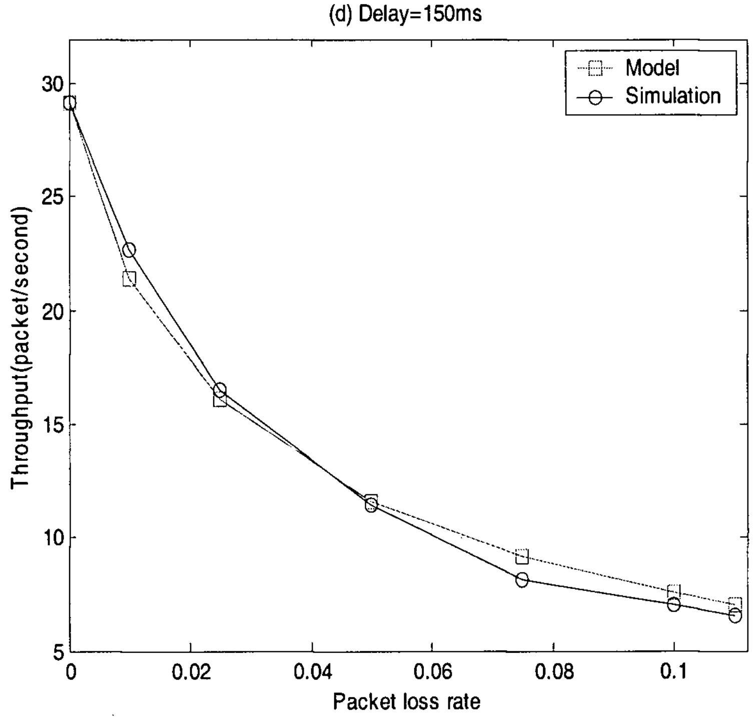


Fig. 5-17 Throughput vs. packet loss rate of UMTS link (Delay = 150 ms)

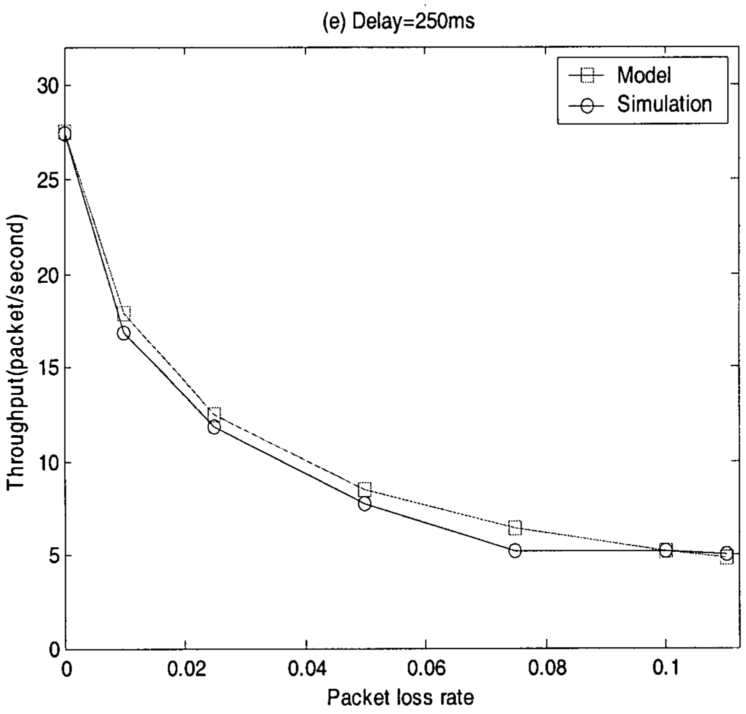


Fig. 5-18 Throughput vs. packet loss rate of UMTS link (Delay = 250 ms)

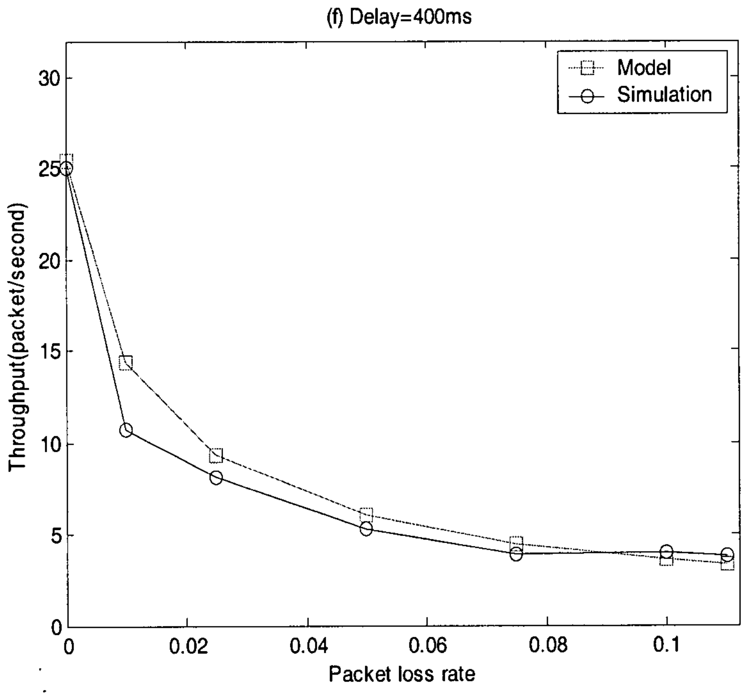


Fig. 5-19 Throughput vs. packet loss rate of UMTS link (Delay = 400 ms)

packet loss rates from 0 to 0.11. From Figs. 5-14 to 5-19, different delays from 0 ms to 400 ms are applied in the tests. On all these plots, we see that the predictions given by the analytical model are quite closely matched to the simulation results, and there is in general a good agreement between the theoretical model and simulations. By comparing each figure with different delay settings, we see that the model predictions are more accurate as loss rates and network delays increase. The average error, which is calculated by equation (5.1), is about 4.98% for 10 simulation runs. This is not a significant level of error when an analytical model is used to predict the performance of transport layer protocols. In fact, with the loss rate and network delay increasing, the RTT would increase dramatically as the throughput approaches the capacity of the link. The analytical model takes into account of this factor and therefore can provide a fairly accurate prediction.

Chapter 6 Conclusions

In this thesis, a new method to support the UMTS/WLAN VHO using MSCTP has been proposed. We have studied different scenarios employing the single-homing and the dual-homing FS to support VHO. Our simulation results showed that the delay and the throughput performance could be improved significantly using the dual homing configuration. In the dual homing configuration, the duplicate transmissions of buffered data over both old and new paths can help the receiver and sender to adapt easily and quickly to a sudden change of link characteristics during and after a vertical handover. SCTP the dual-homing configuration readily copes with the sudden change of the link bandwidth.

Under the current SCTP implementation, the SCTP congestion control causes significant throughput deteriorations when applied to wireless link with error losses. It misinterprets packet losses as indications of network congestions. In order to improve the low throughput performance problem during the WLAN to UMTS forced handover, we have proposed the SMART-FRX scheme to forced the UMTS link to engage in the slow start once the sender experiences the first timeout, and fast retransmissions over the UMTS link to the UMTS destination IP address. Because of the data multicasting on both UMTS and WLAN links during the WLAN possibly unreachable period, the throughput can be increased significantly. We have presented simulation results to validate this claim.

Moreover, we have developed a new analytical model to study the performance of SCTP during a WLAN to UMTS forced VHO. By comparing the numerical results for the analytical model with simulation results, we demonstrate that our model is able to accurately predict SCTP throughput. The analytical model provides a useful tool to estimate of the SCTP throughput.

We summarize our main contributions in this thesis as the following:

- We have proposed a SCTP-based handover management scheme to support UMTS/WLAN integration and given an implementation of mobile SCTP (MSCTP) using ns-2.
- In order to develop the UMTS/WLAN bi-directional VHO procedures, we have proposed to use SCTP bundling and unbundling message technology to simplify the handover procedures.
- We have compared two possible configurations that MSCTP may employ to support VHO: the single-homing and the dual-homing configurations. Our simulation results show that using MSCTP in the dual-homing configuration gives a better delay and throughput performance than that of the single-homing configuration.
- In order to improve the overall handover throughput performance, we have proposed to use a more effective queue management and retransmission scheme that we call SMART-FRX to assist WLAN to UMTS forced VHO.
- We have implemented the SMART-FRX scheme in the simulation model and the simulation results show the effectiveness of the proposed scheme.
- We have proposed a new analytical model to study the SCTP throughput performance on both WLAN and UMTS links during the WLAN possibly unreachable period. The

accuracy of the model has been verified against simulation results over wireless links with random packet losses.

In summary, SCTP is a new transport protocol with many superior features to TCP. Due to its attractive features, SCTP has been receiving more and more attention from the research community and industry. SCTP can be deployed in existing IP networks easily, because SCTP open source implementations are readily available. We believe that SCTP is not only a promising solution to support mobility in the next generation IP network, but also a good candidate for UMTS/WLAN integration. Our proposed methods are not limited to support mobility between UMTS and WLAN, but they can be applied to other inter-system handover processes as well.

Not withstanding the above, some problems need further investigations:

- **Both WLAN and UMTS Links in TO Retransmissions**: Considering the possibility that both WLAN and UMTS links fall into TO retransmissions during a VHO, is there any room to further reduce the data transmission delay and increase the throughput performance?
- **More Realistic Traffic to Validate the Analytical Model**: In future work, it is hoped to extend the applicability of the developed analytical model to the traffic in real MSCTP supporting UMTS/WLAN VHO networks, so that the accuracy of the analytical model predication can be further verified.

Bibliography

[1] ETSI, "3GPP General Packet Radio Service (GPRS) Service description (Stage 2), TS 23.060, Version 3.12.0 Release 1999," www.3gpp.org, ETSI 2002.

[2] "IEEE Standard for Wireless LAN MAC and PHY Specifications," PDF: ISBN 0-7381-1812-5, Jan. 2000.

[3] R. Becher, *et al.*, "Broadband wireless access and future communication networks," in *Proc. of the IEEE*, vol. 89, no. 1, pp. 58-75, Jan. 2001.

[4] C. Perkins, "IP mobility support," RFC2002, Oct. 1996.

[5] M. Moh, G. Berquin, and Y. Chen, "Mobile IP telephony: mobility support of SIP," in *Proc. of IEEE Computer Commun. and Networks*, Oct. 1999.

[6] V. Madisetti and A. Argyrious, "Transport layer QoS management for wireless multimedia services," www.soft-networks.com/vkm-vomo.pdf, Sept. 2002.

[7] J. H. Saltzer, D. P. Reed, and D. D. Clark, "End-to-end arguments in system design," *ACM Trans. on Comp. Sys.*, vol. 2, no. 4, pp. 277-288, Nov. 1984.

[8] A. Jungmaier, M. Schopp, and M. Tuxen, "Performance Evaluation of the SCTP," in *Proc. of IEEE ICATM 2000 Conf. on High Performance Switching and Routing*, Jun. 2000.

[9] R. Stewart and C. Metz, "SCTP: new transport protocol for TCP/IP," in *IEEE Internet Computing*, vol. 5, no. 6, pp. 64-69, Nov. 2001.

[10] L. One and J. Yoakum, "An introduction to the Stream Control Transmission Protocol (SCTP)," RFC 3286, May 2002.

[11] R. Stewart and Q. Xie, "Stream Control Transmission Protocol," IETF RFC 2960, Oct. 2001.

[12] M. Riegel and M. Tuexen, "Mobile SCTP," *draft-riegel-tuexen-mobile-sctp-03.txt*, Aug. 2003, work in progress.

[13] S. J. Koh, M. J. Lee, M. Riegel, L. Ma, and M. Tuexen, "Mobile SCTP for Transport Layer Mobility," *draft-sjkoh-sctp-mobility-03.txt*, Feb. 2004, work in progress.

[14] L. Ma, F. Yu, V.C.M. Leung, and T. Randhawa, "A New Method to Support UMTS/WLAN Vertical Handover Using SCTP," in *Proc. of IEEE VTC'03 Fall*, Oct. 2003.

[15] UC Berkeley, LBL, USB/ISI, and Xerox Parc., "Network Simulator ns-2," documentation and software, Version 2.1b8, http://www.isi.edu/nsnam/ns, 2001.

[16] A. Caro and J. Iyengar, "ns-2 SCTP Module," ver. 3.4, http://pel.cis.udel.edu, Jul. 2003.

[17] J. Padhye, *et al.*, "Modeling TCP throughput: A simple model and its empirical validation," in *ACM SIGCOMM'98*, pp. 303-314, Vancouver, Canada, Sept. 1998.

[18] N. Cardwell, S. Savage, and T. Anderson, "Modeling TCP latency," in *Proc. Of IEEE INFORCOM'00*, vol. 3, pp. 1742-1751, Tel Aviv, Israel, Mar. 2000.

[19] R. Brennan and T. Ravier, "TCP analytic models applied to SCTP: An experimental evaluation," in *Proc. of Information Technology and Telecommunications* (*IT&T*)*'02*, Oct. 2002.

[20] ETSI, "Requirements and Architectures for Inter-working between HIPERLAN/3 and 3rd Generation Cellular Systems," Tech. rep. ETSI TR 101 957, Aug. 2001.

[21] V. K. Varma, *et al.*, "Mobility management in integrated UMTS/WLAN networks," in *Proc. of IEEE ICC'03*, May 2003.

[22] M. Buddhikot, *et al.*, "Integration of 802.11 and third-generation wireless data networks," in *Proc. of IEEE INFOCOM'03*, Apr. 2003.

[23] S. Tsao and C. Lin, "Design and evaluation of UMTS-WLAN interworking strategies," in *Proc. of VTC'02 Fall*, Sept. 2002.

[24] F. Teraoka, *et al.*, "LIN6: A solution to mobility and multi-homing in IPv6," *draft-teraoka-ipng-lin6-02.txt*, Jun. 2003, work in progress.

[25] P. Nikander and J. Arkko, "End-host mobility and multi-homing with host identity protocol," *draft-nikander-hip-mm-01.txt*, Dec. 2003, work in progress.

[26] "Defining strategies to protect against TCP SYN denial of service attacks," Cisco Systems, tech. memo, http://www.cisco.com/warp/public/707/4.html, Aug. 2001.

[27] A.C. Snoeren and H. Balakrishnan, "An end-to-end approach to host mobility," in *Proc. of the 6th ACM/IEEE Intl. Conf. on Mobile Computing and Networking (MobiCom)*, Boston, MA, Aug. 2000.

[28] A.C. Snoeren, H. Balakrishnan, and M. Frans Kaashoek, "Reconsidering Internet mobility," in *Proc. of the 8th Workshop on Hot Topics in Operation Systems (HotOS-VIII)*, May 2001.

[29] R. Stewart and Q. Xie. *Stream Control Transmission Protocol (SCTP): A Reference Guide.* Addison Wesley, New York, NY, 2001.

[30] R. Stewart, *et al.*, "SCTP partial reliability extension," *draft-stewart-tsvwg-prsctp-04.txt*, May. 2003, work in progress.

[31] R. Stewart, *et al.*, "Stream Control Transmission Protocol (SCTP) dynamic address reconfiguration," *draft-ietf-tsvwg-addip-sctp-07.txt*, Feb. 2003, work in progress.

[32] W. Xing, H. Karl, and A. Wolisz, "M-SCTP: design and prototypical implementation of an End-to-End mobility concept," in *Proc. of the 5th Intl Workshop: The Internet Challenge: Technology and Applications*, Berlin, Germany, Oct. 2002.

[33] L. Coene, "Multi-homing issues in the SCTP," *draft-coene-sctp-multihome-04.txt*, Jun. 2003, work in progress.

[34] S. Biaz and N. Vaidya, "Discriminating congestion losses from wireless losses using inter-arrival times at the receiver," in *Proc. of IEEE Symposium ASSET'99*, Mar. 1999.

[35] N. Katsuhiro, *et al.*, "New analytical model for the TCP throughput in wireless environment," in *Proc. of VTC'01 Spring,* May 2001.

[36] M. Allman, *et al.*, "TCP Congestion Control," IETF RFC2581, Apr. 1999.

[37] A. Caro, *et al.*, "Retransmission policies with transport layer multihoming," in *Proc. of ICON 2003*, Sept. 2003.

[38] S. Fu, M. Atiquzzaman, and W. Ivancic, "Evaluation of delay spike on SCTP, TCP Reno, and Eifel in a wireless mobile environment," in *Proc. of Intl. Conf. on Computer Communications and Networks*, pp. 575-578, Miami, FL, Oct. 2002.

[39] R. Stewart, *et al.*, "Stream Control Transmission Protocol (SCTP) Implementer's Guide," *draft-ietf-tsvwg-sctpimpguide-10.txt*, Nov. 2003, work in progress.

[40] Daimler Chryler Research, "Extension to the ns Network Simulator," http://www.informatik.uni-mannheim.de/informatik/pi4/projects/MobileIP/ns-extension/.

Appendix A. Pseudo Code for the Analytical Model

```
%=============================================================
% Set initialize value.
% parameters for "chunk_size", "sack_size", "bandwidth", "ntw_delay" and
% "loss_rate" and "predefined_round_number" need to be set to start the algorithm
%=============================================================

cwnd = [2];
qq = [1];
current_element =1;

p = loss_rate;
q =1-p;
delay = ntw_delay;

pkt_cwnd = cwnd*q;
cum_delay = ntw_delay;
cum_rtt = [(cwnd*chunk_size+(cwnd/2)*sack_size)]/bandwidth;

round = predefined_round_number;

%-------------------------
% Do for every round
%-------------------------
for r = 1:(round-1)

   m = current_element;
   n = 2*current_element

   %------------------------------------------------------------------
   % next round, there are n elements in the cwnd and qq  arrays
   % generate next round qq, cwnd arrays
   %--------------------------------------------------------------------

   %-------------------------------------
   % create 2*current_clement array
   %-------------------------------------
   array = zeros(1,n);

   %-----------------------
   % for next round qq
   %-----------------------
   array(1:2:n) = qq;
   array(2:2:n) = qq;
```

```
qq = array;
for i = 1:m
   qq(2*i) = qq(2*i)*q^cwnd(i);
   qq(2*i-1) = qq(2*i-1)*(1-q^cwnd(i));
end

%-------------------------
% for next round cwnd
%--------------------------
array(1:2:n) = (cwnd*0.5);
array(2:2:n) = cwnd+1;
cwnd = array;
% allow slow start until cwnd(n) = 45
if(r =2); cwnd(n) = 4; end;
elseif(r = 3); cwnd(n) = 8; end;
elseif(r = 4); cwnd(n) =16; end;
elseif(r = 5); cwnd(n) =32; end;
elseif(r = 6); cwnd(n) = 45; end;

%----------------------------------
% for new round rtt and packet
%----------------------------------
array(1:2:n) = cum_rtt;
array(2:2:n) = cum_rtt;
cum_rtt = array;

array(1:2:n) = pkt_cwnd;
array(2:2:n) = pkt_cwnd;
pkt_cwnd = matrix;

for i = 1:n
  cum_rtt(i) = cum_rtt(i)+ (cwnd(i)*chunk_size+(cwnd(i)/2)*sack_size)/bandwidth;
  pkt_cwnd(i) = pkt_cwnd(i)+cwnd(i)*q;
end

pkt = 0;
c_rtt = 0;
cum_delay = cum_delay+ntw_delay;
for i = 1:n
    pkt = pkt+pkt_cwnd(i)*qq(i);
    c_rtt = c_rtt+cum_rtt(i)*qq(i)+cum_delay;
end

%---------------------------------------------------
%  to implement the combine element method
%--------------------------------------------------
```

```
    temp_cwnd = sort(cwnd);

    counter = 1;
    new_cwnd(1) = temp_cwnd(1);
    for i = 2:n
      if(temp_cwnd(i) ~= new_cwnd(counter))
        counter = counter+1;
        new_cwnd(counter) = temp_cwnd(i);
      end

   end

    %------------------------------------------------------------------------------------------
--------
    % generate new_qq arrary, add the transition probability with the same cwnd size
together
    %------------------------------------------------------------------------------------------
--------
    for i = 1:counter
     for j = 1:n
        if(cwnd(j) = new_cwnd(i))
          new_qq(i) = new_qq(i)+qq(j);
        end
     end

    current_element = counter;
    cwnd = new_cwnd
    qq = new_qq
  end;

  %--------------------------------
  % to calculate the throughput
  %--------------------------------
  thr = pkt/c_rtt;
```

Glossary of Acronyms

3G	3rd Generation
AAA	Authentication, Authorization and Accounting
AC	Address Check
ACK	Acknowledgement
ACR	Address Check Reply
AP	Access Point
API	Application Programming Interface
ASCONF	Address Configuration chunk
ASCONF_ACK	Address Configuration Acknowledgement chunk
Association.Max.Retrans	Maximum Association Retransmissions
BS	Base Station
CH	Corresponding Host
CL	Congestion Loss
cmTSN	Cumulative Transmission Sequence Number
cwnd	Congestion Window
DAR	Dynamic Address Reconfiguration
DHCP	Dynamic Host Configuration Protocol
DNS	Domain Name System
ECN	Explicit Congestion Notification
EL	Wireless Channel Error Loss
ETSI	European Telecommunications Standards Institute

FA	Foreign Agent
FD	Four Duplicate SACKs
FIFO	First In First Out
FS	Fixed Server
FTP	File Transfer Protocol
GGSN	GPRS Gateway Node
GI	Generalized Identity
GPRS	General Packet Radio Service
HA	Home Agent
HHO	Horizontal Handover
HI	Host Identifier
HIL	Host Identity Layer
HIP	Host Identity Protocol
HIP AC	HIP Address Check
HIP ACR	HIP Address Check Reply
HIP REA	HIP Readdress packet
HL	Handover Loss
HTTP	Hypertext Transfer Protocol
IETF	Internet Engineering Task Force
IOTA	Inter-working Access gateway
IP	Internet Protocol
IWU	Inter-working Unit
LIN6	Location Independent Network for IPv6

LIN6 ID	LIN6 Node Identifier
LIN6 GI	LIN6 Generalized ID
LINA	Location Independent Network Architecture
MA	Mapping Agent
MAC	Medium Access Control
MC	Mobile Client
MH	Mobile Host
MIP	Mobile IP
MIPBS	Mobile IP Base Station agent
MIPMH	Mobile IP Mobile Host agent
MSCTP	Mobile Stream Control Transmission Protocol
M-TCP	Migrate Transmission Control Protocol
MTU	Message Transmission Unit
NOAH	Non-Ad-Hoc Routing Agent
NS	Network Simulator
Path.Max.Retrans	Maximum Path Retransmissions
PMTU	Path Maximum Transmission Unit
PR	Partial Reliable data transfer
REA	HIP Readdress
RR	Round Robin scheduling
RSS	Received Signal Strength
RTO	Retransmission Timeout
RTO_{min}	Minimum Retransmission Timeout

RTT	Round Trip Time
rwnd	Receiver's Advertised Window
QoS	Quality of Service
SACK	Selective Acknowledgement
SCTP	Stream Control Transmission Protocol
SGSN	GPRS Support Node
SIP	Session Initiation Protocol
SMART-FRX	Sending-buffer Multicast with Fast Retransmission
ssthresh	Slow Start Threshold
T3-rtx	Timer 3 for Retransmission timeout
TCP	Transmission Control Protocol
TELNET	TCP/IP Terminal Emulation Protocol
TO	Timeout
TSN	Transmission Sequence Number
UDP	User Datagram Protocol
UMTS	Universal Mobile Telecommunication System
VHO	Vertical Handover
WLANs	Wireless Local Area Networks