

Mobility Protocols for Handoff Management in Heterogeneous Networks

F. Siddiqui¹ and S. Zeadally²

¹ Wayne State University, Detroit, MI 48201, USA

² University of the District of Columbia, Washington, DC 20008 USA

Abstract. Future generation networks are expected to be a combination of several types of access technologies that vary in their characteristics. Efficient handoff management techniques are required to enable end-users to seamlessly access these networks as they roam across different geographic locations. We describe recent protocols (application, transport, and network) such as Mobile IP, Session Initiation Protocol (SIP), and Stream Control Transmission Protocol (SCTP) that have been deployed to handle handoffs. We present an empirical performance evaluation of the three protocols using performance metrics such as handoff delay. We found that Mobile IP yields the highest handoff delay out of all the three mobility protocols. SIP and SCTP yield (33 %) and (55 %) lower handoff delays compared to Mobile IP.

Keywords: Handoff, Mobility, Heterogeneous, Protocols, Networks.

1 Introduction

The demand for ubiquitous information access has led to the convergence of several types of networks including Ethernet Local Area Network (LAN), General Packet Radio Service (GPRS), Global System for Mobile Communication (GSM), Wireless Local Area Network (WLAN), Bluetooth, etc. In such heterogeneous environments mobility management is the basis for providing continuous network connectivity to mobile users roaming between these access networks. There are two major components of mobility management: Location management and Handoff management. Location management enables the network to discover the current attachment point of the mobile user. Handoff management enables the mobile node to maintain the network connection as it continues to move and change its access points or base stations to the network.

Several protocols have been proposed [9] [10] [11] to address the issue of mobility management in heterogeneous networks. These approaches operate at different levels of the network protocol stack.

- **Network Layer:** Mobile IP [1] was proposed by the Internet Engineering Task Force (IETF) to handle mobility management at the network layer. It handles mobility by redirecting packets from a mobile node's home network to the mobile node's current location. Deployment of Mobile IP requires network servers including a home agent and a foreign agent that are used to bind the home address

of a Mobile Node (MN) to the care-of address at the visited network and provide packet forwarding when the MN is moving between IP subnets.

- **Application Layer:** The Session Initiation Protocol (SIP) [2] is an application layer protocol that keeps mobility support independent of the underlying access technologies. In the SIP approach, when an MN moves during an active session into a different network, it first receives a new network address, and then sends a new session invitation to the correspondent node. Subsequent data packets from the CN are forwarded to the MN using the new address. The MN also needs to register its new IP address with a SIP server called a Registrar to enable other nodes on the network to reach it by querying the Registrar server.
- **Transport Layer:** A third approach for mobility management has been proposed at the transport layer in the form of the Stream Control Transmission Protocol (SCTP) [3]. The SCTP-based approach uses multihoming for implementing mobility management. The multihoming feature allows a SCTP to maintain multiple IP addresses. Among those addresses, one address is used as the primary address for current transmission and reception. Other addresses (secondary) can be used for retransmissions. The multihoming feature of SCTP provides a basis for mobility support since it allows a mobile node (MN) to add a new IP address, while holding an old IP address already assigned to it.

In this paper we present a comparison of SIP, Mobile IP and SCTP for supporting handoff management in heterogeneous networks. We present an empirical evaluation of handoff latency achieved in the case of each protocol when a mobile user roams across different types of networks. We also identify issues in setting up a testbed to conduct handoff delay tests. The rest of this paper is organized as follows. In section 2 we give an overview of the three mobility management protocols: SIP, Mobile IP, and SCTP. Section 3 discusses the experimental procedures and testbed setup used for conducting our performance evaluation tests. In section 4 we present a performance analysis of handoffs conducted across different network types. Finally, in section 5 we make some concluding remarks.

2 Mobility Management Protocols

2.1 SIP-based Terminal Mobility

SIP is an application-layer control protocol that can establish, modify and terminate multimedia sessions [2]. SIP defines several logical entities, namely user agents, redirect servers, proxy servers and registrars. SIP inherently supports personal mobility and can be extended to support service and terminal mobility [8]. Terminal mobility allows a device to move between IP sub-nets, while continuing to be reachable for incoming requests and maintaining sessions across subnet changes. Mobility of hosts in heterogeneous networks is managed by using the terminal mobility support of SIP.

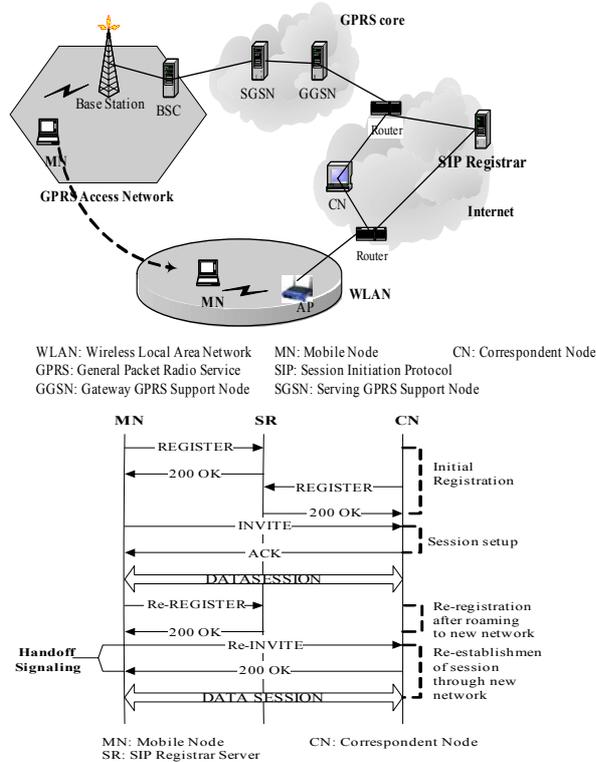


Fig. 1. SIP-based Mobility Management

Terminal mobility requires SIP to establish a connection either during the start of a new session, when the terminal or MN has already moved to a different location, or in the middle of a session. The former situation is referred to as *pre-call mobility*, latter as *mid-call* or *in-session mobility*. For pre-call mobility, the MN re-registers its new IP address with the Registrar server by sending a REGISTER message, while for mid-call mobility the terminal needs to notify the correspondent Node (CN) or the host communicating with the MN by sending a re-INVITE message about the terminal's new IP address and updated session parameters. The CN starts sending data to the new location as soon as it receives the re-INVITE message. The MN also needs to register with the redirect server in the home network for future calls. Figure 1 shows the messages exchanged for setting up a session between a mobile node and a correspondent node and continuing it after changing the access network.

2.2 Mobile-IP-based Mobility

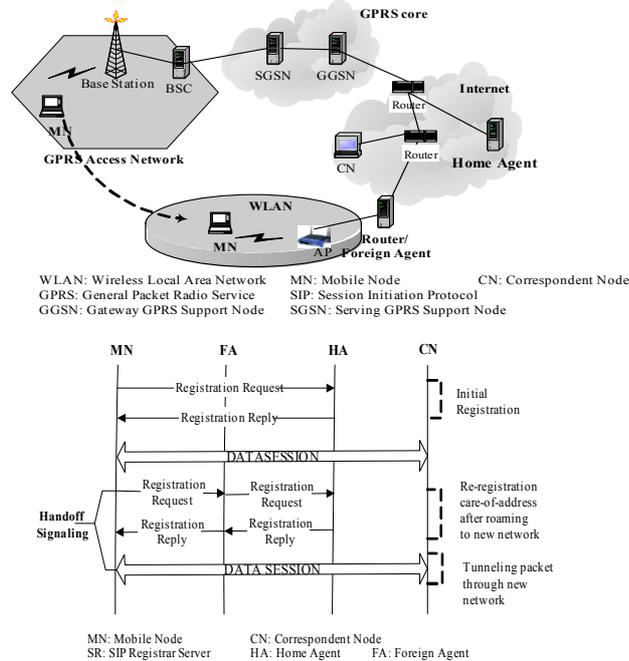


Fig. 2. Mobile-IP-based Mobility Management

Mobile IP is a mobility management protocol proposed to solve the problem of node mobility by redirecting packets to the mobile node's current location. The Mobile IP architecture is shown in figure 2. Its main components include a Home Agent (HA) and a Foreign Agent (FA). HA is a router on a mobile node's home network, which encapsulates datagrams for delivery to the mobile node when it is away from home, and maintains current location information for the mobile node. FA is a router on a mobile node's visited network (foreign network) that provides routing services to the mobile node when registered. The FA decapsulates and delivers datagrams, tunneled by the mobile node's HA to the mobile node. When a mobile node moves out of its home network it must obtain another IP. So, in Mobile IP, a mobile host uses two IP addresses: a fixed home address (a permanent IP address assigned to the host's network) and a *care-of-address* - a temporary address from the new network (i.e. foreign network) that changes at each new point of attachment. When the mobile node moves, it has to first discover its new care-of-address. The care-of-address can be obtained by periodic advertising from the FA through broadcasting. The mobile node then registers its care-of-address with its home agent by sending a Registration Request to its home agent via the foreign agent. The HA then sends a Registration Reply either granting or denying the request. If the registration process is successful, any packets destined for the MN are intercepted by the HA, which encapsulates the packets and tunnels them to the FA where decapsulation takes place and the packets are then forwarded to the appropriate MN.

2.3 SCTP based Mobility

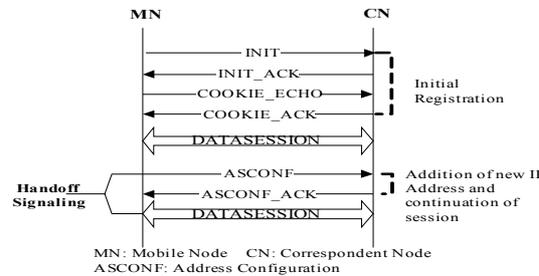


Fig. 3. SCTP-based Mobility Management

The Stream Control Transmission Protocol (SCTP) [3] is a reliable connection-oriented transport protocol that operates over a potentially unreliable connectionless packet service, such as IP. Before peer SCTP users can send data to each other, a connection must be established between two endpoints. This connection is called an association in SCTP context. A cookie mechanism is employed during the initialization of an association to provide protection against security attacks. Figure 3 shows a sample SCTP message flow. An essential property of SCTP is its support for multihomed nodes, i.e. nodes that can be reached under several IP addresses. If a client is multi-homed, it informs the server about all its IP addresses with the INIT chunk's address parameters. An extension to the SCTP called mSCTP (Mobile SCTP) also allows dynamic addition and deletion of IP addresses from an association, even if these addresses were not present during association startup. This feature of SCTP is used to support mobility of hosts across different networks.

3 Performance Evaluation of Mobility Protocols

3.1 Experimental testbed

We conducted experimental measurements to determine the handoff delay experienced while roaming across different networks. The handoff tests were conducted for each of the mobility protocols: SIP, Mobile IP, and SCTP.

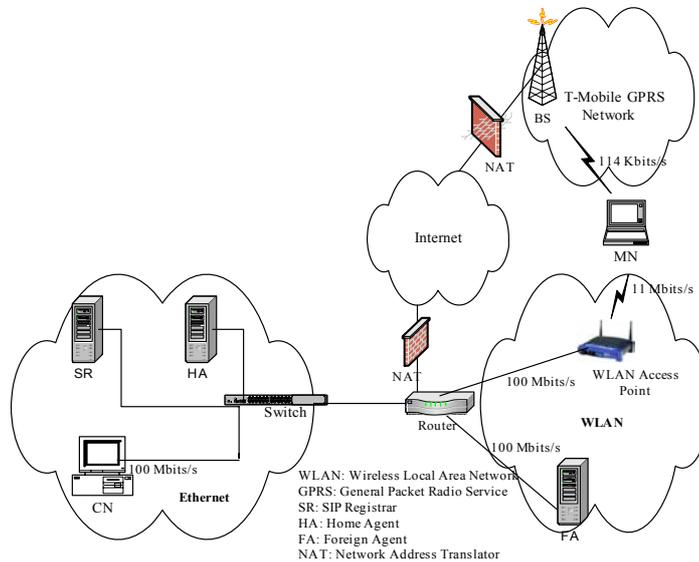


Fig. 4. Experimental Testbed used for SIP, Mobile IP, and SCTP Handoff Performance Measurements

Figure 4 shows the experimental testbed used for conducting the handoff measurements. The setup consists of a DELL laptop (client machine) equipped with three network interface cards (NICs): a built-in Natsemi Ethernet NIC (100 Mbps), a built-in Orinoco WLAN NIC (11 Mbps) and an external PCMCIA GPRS Sierra Wireless aircard 750 (144 Kbps). The Ethernet interface (eth0) of the client machine is connected to a 100 Mbits/sec switch that connects to the external IP network (Internet). The WLAN interface (eth1) of the client machine is associated with a WLAN access point, which is in turn connected to the router for Internet access. The GPRS interface (ppp0) is associated with a T-Mobile GPRS base station, which connects to the Internet via the GPRS core network. In order to use the GPRS network, we purchased a GPRS data plan subscription from the T-Mobile service provider [5]. Other components of the testbed include a SIP Registrar server, a Home Agent, and a Foreign Agent.

The client (Mobile Node) and the server (Correspondent Node) machines were loaded with Redhat 9.0 Linux operating system and used a kernel version of 2.4.20-8. For SCTP-based mobility tests, a user-level implementation of SCTP called Sctplib-1.3.1 [6] (developmental version) was used. For Mobile-IP-based tests, a Mobile IP user-level implementation called Dynamics [7] from Helsinki University of Technology was used. SIP-based mobility was tested by implementing a simple SIP user-agent client [8], a SIP user agent server and a SIP Registrar server using the SIP methods (INVITE, ACK, BYE, REGISTER, and CANCEL) described in RFC 3261 [2].

3.2 Measurement Procedures and Performance Metrics

We measured the handoff delay experienced when roaming across three types of networks: Ethernet, WLAN and GPRS by implementing mobility protocols at the application (SIP), network (Mobile IP) and Transport (SCTP) layers.

In the case of SIP, we measured the handoff delay experienced by a mobile node in six different cases:

- GPRS to WLAN
- WLAN to GPRS
- Ethernet to WLAN
- WLAN to Ethernet
- Ethernet to GPRS
- GPRS to Ethernet

In the case of SCTP and Mobile IP, we measured the handoff delay in two different cases:

- Ethernet to WLAN
- WLAN to Ethernet

The performance metrics that we measured are as follows:

- **Total Handoff Delay:** The total handoff delay is the time difference between the last data packet received at the old network interface and the first data packet received on the new network interface. The total handoff delay includes the handoff time as well as the time taken for the first data packet to arrive from the mobile node to the correspondent node.
- **Handoff Signaling Time:** The handoff signaling time is a measure of the time required to exchange signaling messages to execute a handoff. The number of signaling messages exchanged is different for each mobility management protocol.
- **Packet Transmission Delay after handoff:** The packet transmission delay after the handoff is a measure of the transmission time of a packet from the mobile node to the correspondent node after the mobile node has moved to a new network.

3.2.1 SCTP and Mobile IP Issues for NAT Traversal

It was not possible to measure the handoff delay (for SCTP and Mobile IP) while moving from the GPRS network to the other networks (Ethernet and WLAN) and vice versa because the GPRS operator assigns a dynamic, private IP address to the mobile node. A dynamic IP address is one that is not manually specified and is not a permanent address. It is a temporary address that is dynamically configured using the Dynamic Host Configuration Protocol (DHCP). A private IP address is one that can

be used by any machine and is therefore re-usable. However, private IP addresses are not routable over the public Internet. They are used in private networks due to the shortage of public, routable IP addresses. The range of IP addresses reserved for private use includes 10.0.0.0 – 10.255.255.255, 172.16.0.0 – 172.31.255.255, 192.168.0.0 – 192.168.255.255. Also, each Internet provider network employs a Network Address Translator (NAT) for providing Internet access to the internal nodes with private IPs and also for security purposes.

The problem with Dynamics implementation of Mobile IP is that it is not "NAT traversal" capable. When a mobile node moves to the GPRS network, it acquires a care-of-address (CoA), which is a private address. Then the mobile node sends a Registration Request to the Home Agent (HA) to register its new CoA. However, at the NAT gateway, the private IP address of this packet (source IP address in the IP header) is replaced by the public IP address of the NAT gateway. When the Registration Request arrives at the HA, the HA detects that that the source address of the packet (which is the public address) is different from the CoA inside the Registration Request message (present in the Mobile IP header). Therefore the HA drops the request. Thus, in the case of the Dynamics, it is necessary to have a public, static IP address for the mobile node. Hence, handoffs involving the GPRS network could not be tested due to the assignment of a private IP.

In the case of SCTP, when the mobile node is located in the GPRS network and the correspondent node is located on a different network, all packets from the mobile node have to pass through the NAT. SCTP has certain issues related to NATs. If Network Address Port Translation is used with a multihomed SCTP endpoint, then any port translation must be applied on a per-association basis such that an SCTP endpoint continues to receive the same port number for all messages within a given association. The NAT needs to understand this requirement to allow mobility support using SCTP. Since existing NATs are not designed to support SCTP, a NAT assigns a different port number when the SCTP association changes its primary address. The SCTP server does not accept the change in the port number and breaks the association. Thus SCTP cannot be experimented with a GPRS network employing a NAT that is not configured to support SCTP.

4 Experimental Results and Discussion

In this section we present an analysis of the handoff performance obtained for the three mobility management protocols. Figure 5 shows the total handoff delay obtained while roaming from Ethernet to WLAN and vice versa using SIP, Mobile IP and SCTP. It is worthwhile mentioning that SIP, Mobile IP and SCTP operate at the application, network, and transport layers respectively.

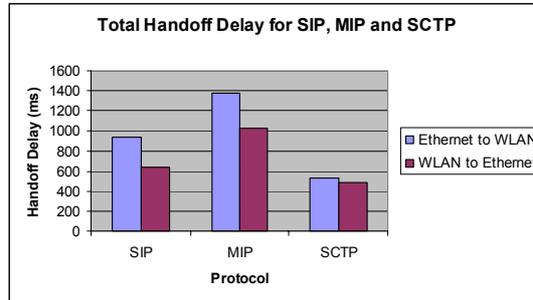
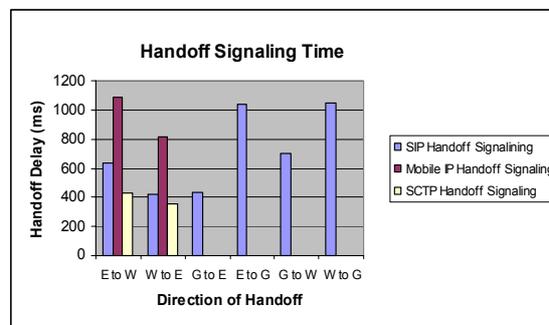


Fig. 5. Total Handoff Delay for SIP, Mobile IP and Sctp



E: Ethernet W: WLAN G: GPRS

Fig. 6. Handoff Signaling Time for SIP, Mobile IP, and Sctp

It can be observed from figure 5 that the total handoff delay in either direction (Ethernet to WLAN and vice versa) is the lowest in the case of Sctp followed by SIP and is the highest in the case of Mobile IP. The total handoff delay is lowest for Sctp (31% lower compared to SIP and 55% lower compared to Mobile IP for WLAN to Ethernet handoff). The reason for the low handoff delay in the case of Sctp is because the Sctp client immediately adds the IP address of a newly discovered network to its list of available networks and also relays this information to the Sctp server. When a handoff is initiated due to the unavailability of the current network, the client sends an ASCONF_DELETEIP message to the server (for removing the old IP address) and starts using the interface with the new IP address for data transmission. Thus, the handoff process with Sctp involves very few signaling messages thereby resulting in a low total handoff time. Table 1 lists the signaling messages exchanged for implementing handoffs using SIP, Mobile IP and Sctp.

Table 1. Components of Handoff Signaling: SIP, Mobile IP, and SCTP

Protocol	Handoff Messages
SIP	Re-Register ACK
Mobile IP	Registration Request, Registration Reply
SCTP	ASCONF_DELETE IP

In the case of SIP, when a handoff is initiated, the SIP client sends a Re-INVITE message to the SIP server using the new interface. After the SIP server acknowledges the Re-INVITE, the communication between the client and the server is continued. Thus, handoff delay in the case of SIP is the two-way delay involved in sending the Re-INVITE message and receiving an acknowledgement. We determine the handoff delay at the correspondent node as the time difference between the last data packet received at the old network interface and the first data packet received at the new network interface. Thus, the handoff delay also includes the transmission time of the first packet following the handoff signaling. In the case of Mobile IP, the handoff involves a higher number of signaling messages compared to SIP and SCTP. Mobile IP requires the mobile node needs to send a Registration Request to the Foreign Agent that forwards the request to the Home Agent. The Registration Reply is sent by the Home Agent to the Foreign Agent which then gets forwarded to the mobile node. Due to the high signaling overhead involved in the case of handoffs based on Mobile IP, the signaling time is also higher.

Figure 6 shows the handoff signaling time in the case of the SIP protocol when the mobile node moves across various networks. It can be observed that the signaling time is the highest when the mobile node makes a handoff to a GPRS network. The signaling time is comparatively lower when the mobile moves to the WLAN and is the lowest in the case of transition to an Ethernet network. We note that the low signaling delay associated with transition to an Ethernet network is probably because of Ethernet's lowest transmission latency. To confirm this explanation, we performed a simple test using Netperf [4] to determine the available bandwidth and the latency offered by each of these networks. As shown in table 2, the latency incurred on the GPRS network is comparatively higher as compared to Ethernet and WLAN. This accounts for the high handoff signaling delay when the mobile node moves to the GPRS network. We also observe (from figure 6) that there is a 41 % reduction in the handoff signaling time in the case of SIP when compared to Mobile IP (for handoff to a WLAN) and a 60 % decrease in the handoff signaling time in the case of SCTP as compared to Mobile IP.

Table 2. Network Characteristics determined by running a Netperf test

Network Type	Link Speed	Actual Measured Bandwidth	Average Latency (one-way)
GPRS	114 Kbps	28.9 Kbps	891 ms
WLAN	11 Mbps	5.51 Mbps	61 ms
Ethernet	100 Mbps	88.8 Mbps	36 ms

Figure 7 shows the transmission delay incurred by packets arriving at the correspondent node after the handoff. We observe that in the case of Mobile IP, we

obtained highest packet transmission delay. As observed from figure 7, there is a 47 % decrease in the packet transmission delay in the case of SIP as compared to Mobile IP (in the case of handoff to a WLAN) and a 54 % decrease in the packet transmission delay with SCTP as compared to Mobile IP (in the case of handoff to a WLAN). This is because, after handoff, packets from the Mobile Node to the Correspondent Node have to be routed through the Home Agent and the Foreign Agent before they can reach the Correspondent Node. This introduces additional delay in the transmission time. The packet transmission delay for SCTP and SIP is almost the same. In both these cases, the packets following handoff are sent directly from the Mobile Node to the Correspondent Node. This results in a lower packet transmission delay for SIP and SCTP as compared to Mobile IP.

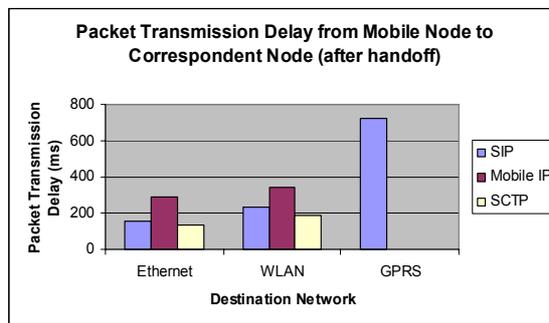


Fig. 7. Packet Transmission Delay after Handoff

5 Conclusions and Future Work

In this paper we have compared the handoff performance of three types of mobility management protocols: SIP, Mobile IP and SCTP. We found that SCTP performs well both, in terms of handoff delay, as well as the packet transmission time after a handoff. The SIP protocol incurred a higher handoff delay compared to SCTP but the packet transmission time for packets after a handoff was almost comparable for the two protocols. Mobile IP showed higher handoff delay as well as longer packet transmission time following handoff to a new network. However, Mobile IP keeps the change in the IP address completely transparent to the other end-system. In the case of SIP and SCTP, the change in the destination IP address has to be conveyed to the node at the other end. SCTP-based mobility is however completely transparent to the application, whereas in the case of SIP, applications need to be aware of mobility.

We also discussed the issues that exist in connection with deploying Mobile IP and SCTP over networks using private IP addresses and deploying NATs. Mobile IP and SCTP are not capable of operating in networks with NAT mechanisms. Since almost all network operators use NATs in their networks, it is crucial to extend these protocols to enable them to operate across heterogeneous domains. One method that can be used to enable this feature is to use UDP encapsulation in each of these protocols. Since most NATs are already designed to provide support for UDP packets, encapsulating SCTP packets inside UDP can make SCTP operate across NATs

belonging to different network domains. However, this would introduce additional encapsulation-decapsulation delays. In our future work, we aim at overcoming the drawbacks of these protocols and design a solution for handoff management that is not specific to a single layer of the network protocol stack, but employs a cross-layer design for achieving seamless handoffs across heterogeneous networks. We plan to implement a mobility middleware that performs handoffs using the information from various layers (such as link quality information from layer 2, QoS information from layer 4, etc) of the protocol stack and is deployable across existing network configurations.

References

1. C. Perkins, "Mobile networking through Mobile IP", IEEE Internet Computing, Volume 2, Issue 1, January - February 1998 Pages: 58 – 69.
2. Rosenberg et al., "Session Initiation Protocol", RFC 3261.
3. Stewart et al., "Stream Control Transmission Protocol", RFC 2960.
4. Netperf, <http://www.netperf.org>
5. T-Mobile, <http://www.tmobile.com>
6. SCTP Implementation, <http://www.sctp.de>
7. Mobile IP, Dynamics Implementation, <http://www.cs.hut.fi/Research/Dynamics>
8. Schulzrinne H., "Application Layer Mobility with SIP", ACM SIGMOBILE Mobile Computing and Communications, Volume 4, Issue 3, July 2000, Pages: 47 – 57.
9. Eddy, W., "At What Layer Does Mobility Belong", IEEE Communications Magazine, Volume 42, Issue 10, October 2004 Pages: 155 – 159.
10. Chiussi, F.M.; Khotimsky, D.A.; Krishnan, S., "Mobility management in third-generation all-IP networks", IEEE Communications Magazine, Volume 40, Issue 9, September 2002 Pages:124 – 135.
11. Banerjee, N.; Das, S.K.; Acharya, A., "SIP-Based Mobility Architecture for Next Generation Wireless Networks", in proceedings of the third IEEE International Conference on Pervasive Computing and Communications, 8-12 March 2005 Pages:181 – 190.