A Survey of mSCTP for Transport Layer Mobility Management

Raman Kumar Goyal
Email: raman.uiet1985@gmail.com

Sakshi Kaushal
Department of Computer Science and Engineering, UIET, Panjab University, Chandigarh
Email: sakshi@pu.ac.in

Abstract—Wireless Communications have increased rapidly in recent years. Wireless LAN, Code Division Multiple Access (CDMA), Universal Mobile Tele-Communication System (UMTS), GPRS (General Packet Radio Service), Bluetooth, 3G (Third Generation networks), and GPS (Global Positioning System), etc., are common wireless technologies that are commercially available. Internet together with small hosts in the form of laptop, mobile phones, etc., has increased the demand of mobility support for moving nodes as when nodes move from one network to another the Internet connectivity should be maintained during handover. In this paper, various protocols for mobility management have been reviewed. Network layer protocols like Mobile IP and its variants and transport layer protocol mSCTP are compared. Various issues regarding mSCTP handover have been investigated like packet re-ordering, throughput, multi-path transfer, handover in case of fast moving nodes so that optimal handover can be achieved.

Index Terms—Mobility, SCTP, mSCTP, handover, DAR, MIP.

I. INTRODUCTION

As the most hosts that are providing Internet facility are mobile, so mobility management emerges as an open challenge. When a node moves from one network to another the communication with Internet is disconnected due to change of IP address [1]. Mobility can be classified into terminal, personal, session and service mobility [2]. Terminal mobility refers to maintaining the connection and maintaining the session when the network changes. Personal mobility refers to locate the users at several terminals. Session mobility is the ability of maintaining session when a user moves between different terminals. Service mobility can be defined as continuation of services even when terminals are moving [3]. Terminal mobility is of main consideration because it involves location and handover management. Handover management maintains communication with the Internet while changing the host’s point of attachment to the IP network. Handover between the networks give rise to challenges like handover latency, packet loss and security. Location management deals with updating the current location of the host so that any entity can communicate with the host at the new location, while being transparent to its peers.

Mobile IP (MIPv4) [4] was introduced by the Internet Engineering Task Force (IETF) to provide mobility support. With the introduction of IPv6, MIPv6 [5] was proposed to remove the deficiencies of MIPv4 but MIPv6 has also some shortcomings like handover latency [54], signaling cost and low bandwidth. Therefore variants of MIPv6 like Hierarchical MIPv6 (HMIPv6) [6] and Fast Handover MIPv6 (FMIPv6) [7] were introduced which required MN protocol stacks modifications. These protocols were introduced to reduce handover latencies mainly. All the above mention protocols provide mobility support at network layer. A transport layer protocol called Stream Control Transport Protocol (SCTP) [8] is proposed which gained interest for research in recent years due to its features like multi-homing and multi-streaming. A new protocol using SCTP with Dynamic Address Reconfiguration (DAR) [9] called mobile SCTP (mSCTP) [10] enables vertical handovers between heterogeneous networks.

This paper focuses on transport layer mobility management provided by mSCTP. Various issues of mobility management have been investigated and work done on these issues with mSCTP has been reviewed. In the next section, types of handovers are discussed in brief. In section 3 transport layer mobility management is discussed. mSCTP is compared with network layer mobility solutions in section 4. In section 5 enhancements in mSCTP for achieving smooth and seamless handover are discussed. Multi-path transfer in SCTP, performance of mSCTP in case of fast moving Mobile Nodes (MNs), various congestion control mechanisms and cross layer mobility solutions are discussed to achieve optimal handover by reducing packet losses and avoid spurious retransmissions. Section 6 concludes the paper.

II. TYPES OF HANDOVER

Handover in IP based networks can be defined as the process for redirecting IP packets to MN’s new location which are destined to MN’s old location. Different types of handovers that can occur in wireless networks [11-14]
based on various network parameters or properties as listed below:

A. **Horizontal and Vertical Handover**

Handover between same types of networks is known as horizontal handover while vertical handover means switching between different types of networks [11, 12]. Horizontal handovers occur between homogeneous cells of a wireless access systems. Vertical handovers occur between heterogeneous cells of wireless access systems having different features like bandwidth, data rate, etc. Vertical handovers are of two types: upward vertical handover and downward vertical handover.

B. **Upward–vertical and Downward-vertical Handover**

An upward-vertical handover occurs when a MN switches from a network having smaller coverage but having greater bandwidth to a network with larger coverage and lower bandwidth. While a downward vertical handover is a handover from larger coverage area and lower bandwidth to smaller coverage area and higher bandwidth [12, 15].

C. **Anticipated (Mandatory) and Unanticipated (Forced) Handover**

When a MN knows that it is in the range of a new access point and it lost the coverage from old access point, it will always perform the handover. This type of handover is known as anticipated handover. If the handover is performed due to radio link causes it is known as mandatory handover. The MN has to switch the access point in order to maintain its connection; therefore this type of handover is called mandatory handover [12, 14, 16]. Horizontal and upward vertical handover falls into the category of anticipated handover. But if the handover is done due to pre-process planning as in the case of congestion it is known as forced handover. In case of downward vertical handover a MN can choose to delay its handover process based on some requirements of applications running on MN. This type of handover is known as unanticipated handover.

D. **Hard and Soft Handover**

If the MN has only one active connection, i.e., it can be associated with only one access point that is known as hard handover. If the MN has more than one network interfaces it can associate with multiple access points in different networks. Handover in which MN can interact with multiple access points is known as soft handover [11-13].

E. **Mobile Controlled and Network Controlled Handover**

If the handover decision is taken by MN it is known as mobile controlled handover while if the handover decision is taken by the network it is known as network controlled handover [12].

### III. Transport Layer Mobility Management

The Stream Control Transmission Protocol (SCTP) [8] is being standardized by the IETF as a reliable transport protocol to transport SS7 signalling messages over IP networks [17]. The major differences between SCTP and TCP are illustrated in Table 1.

<table>
<thead>
<tr>
<th>S.No.</th>
<th>Features</th>
<th>TCP</th>
<th>SCTP</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Setup</td>
<td>Three way</td>
<td>Four way</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Handshake</td>
<td>Handshake</td>
</tr>
<tr>
<td>2.</td>
<td>Shutdown</td>
<td>Four way</td>
<td>Three way</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Handshake</td>
<td>Handshake</td>
</tr>
<tr>
<td>3.</td>
<td>Message Boundary</td>
<td>Stream Oriented</td>
<td>Message Oriented</td>
</tr>
<tr>
<td>4.</td>
<td>Half Open</td>
<td>Supported</td>
<td>Not Supported</td>
</tr>
<tr>
<td></td>
<td>Connection</td>
<td>and prone to</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>denial of</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>service</td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>SYN attacks</td>
<td>Present</td>
<td>Not present</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>due to</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>cookie</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>mechanism</td>
</tr>
<tr>
<td>6.</td>
<td>Multi-homing</td>
<td>Not Provided</td>
<td>Provided</td>
</tr>
<tr>
<td>7.</td>
<td>Multi-streaming</td>
<td>Not Provided</td>
<td>Provided</td>
</tr>
</tbody>
</table>

**A. Multi-homing and Transport Layer Handover Scheme**

SCTP has two important features which were not present in TCP called multi-streaming and multi-homing. Multi-streaming allows data from the upper layer application to be multiplexed into one channel called association in SCTP. Multi-homing allows an association between two endpoints span across multiple IP addresses or network interface cards as shown in Figure 1.

![Figure 1: Concept of SCTP multi-homing in two IP networks](image)

One of addresses is set as primary, while the other addresses are set as back-up addresses. mSCTP exploits this multi-homing feature, which enables an association among devices across multiple IP addresses for mobility support [20]. Mobility can be supported with Dynamic Address Resolution (DAR) [9] that has two additional chunks for address configuration namely Address Configuration Change Chunk (ASCONF) and Address Configuration Acknowledgement (ASCONF-ACK). ASCONF chunk communicates to the remote end point about one of the configuration changes that must be acknowledged. There are three parameters used in the ASCONF, which are as follows:
(1) Add IP Address: Adding an IP address to the association.
(2) Delete IP Address: Deleting an IP address from the association.
(3) Set Primary Address: Setting an IP address as a primary address where the data can be transmitted.

ASCONF-ACK is used by the receiver of the ASCONF chunk to acknowledge the reception. As shown in Figure 2, when a MN comes into the overlapping region, the MN gets router advertisement from New Access Router (NAR) and the new IP address is added into the association by ASCONF Add IP [18]. As the MN moves out of the range of Previous Access Router (PAR) and comes into the range of NAR the new IP address is set as the primary address by ASCONF Set Primary. When MN leaves the overlapping region and enters completely into the new network, deletion of previous IP address is performed. The complete message flow of mSCTP is shown in Figure 3. Firstly the Proxy Router Solicitation (PrRtSol) message is sent from MN to PAR. PAR responds with Proxy router advertisement (PrRtAdv) to ASCONF messages are sent from MN to Co Node (CN) for adding IP address of new network, seeing it as primary address and deleting the previous IP address respectively. All the ASCONF messages are responded with ASCONF-ACK message from CN to MN. The two chunks that are used for data transmission in SCTP are the DATA chunk and the SACK chunk. DATA chunk is used by the data sender that carries the user data and the SACK chunk is used by the data receiver to carry the acknowledgement of the receipt of the DATA chunk.

Heartbeat (HB) mechanism is equivalent to TCP’s keep-alive mechanism in which back-up (secondary) addresses are checked whether they are alive or not.

Figure 2: Working of mSCTP

Figure 3: Message Flow in mSCTP protocol

B. mSCTP for Advances in Information Technology
is a gateway based approach in which a gateway splits the connection between MN and CN and allows MN to change its connection with the gateway during handover. mSCTP [10] is a handover protocol that offers soft handovers. It does not include location management. Seamless IP diversity based Generalized Mobility Architecture (SIGMA) [26] is a complete handover solution similar to mSCTP that also includes location management using Domain Name Service. mSCTP is compared with network layer mobility approaches in Table 2.

C. Application Layer Approaches

The main application layer mobility approach is Session Initiation Protocol (SIP) [27] that offers personal mobility and terminal mobility can be extended by establishing a new connection during the start of a new session, when a MN has moved to a new network or in the middle of a session. Dynamic DNS (DDNS) is another application layer mobility management approach that can be used for location management where the mobile terminal act as a server and other nodes actively originate communication with the mobile terminal.

Mobility protocols in different layers have been compared by various researchers on different parameters. mSCTP performs better in terms of handover delay and packet transmission during a handover [28, 29].

### Table 2

<table>
<thead>
<tr>
<th>Feature</th>
<th>MIP</th>
<th>HMIPv6</th>
<th>FMIPv6</th>
<th>mSCTP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operating layer</td>
<td>Network</td>
<td>Network</td>
<td>Network</td>
<td>Transport</td>
</tr>
<tr>
<td>IP Connections</td>
<td>One</td>
<td>One</td>
<td>One</td>
<td>More than One</td>
</tr>
<tr>
<td>Packets Reordering</td>
<td>Very low</td>
<td>Very low</td>
<td>Minimu m</td>
<td>Required</td>
</tr>
<tr>
<td>Concurrent multi-path transfer</td>
<td>Not Possible</td>
<td>Not Possible</td>
<td>Not Possible</td>
<td>Possible</td>
</tr>
<tr>
<td>Type of Handover</td>
<td>Hard</td>
<td>Hard</td>
<td>Hard</td>
<td>Soft</td>
</tr>
<tr>
<td>Packet Loss</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>Least in Comparision</td>
</tr>
<tr>
<td>Handover Latency</td>
<td>Bad</td>
<td>Moderat e</td>
<td>Very Good</td>
<td>Very Good</td>
</tr>
<tr>
<td>Impact of speed of MN on Throughput</td>
<td>Bad</td>
<td>Bad</td>
<td>Very Bad</td>
<td>Least in Comparision</td>
</tr>
<tr>
<td>Location Management</td>
<td>Provided</td>
<td>Provide d</td>
<td>Provide d</td>
<td>Not Provided</td>
</tr>
</tbody>
</table>

V. ENHANCEMENTS IN MSCTP FOR SMOOTH AND SEAMLESS HANDOVER

As the terminals for accessing the networks are becoming smaller in size, demand of accessing the Internet on move is increasing. Downloading data and accessing various Internet applications while moving in a fast speed vehicle emerges as a great challenge due to Internet connectivity issues when terminal moves from one network to another. mSCTP provides soft handover for heterogeneous networks and performs better than the trivial MIP solutions in terms of handover latency, bandwidth, packet loss, etc. mSCTP is still evolving as it is in its research stage and its performance can further be optimized.
Various enhancements can be made in mSCTP for achieving the optimal handover. Data can be sent on all the paths for better bandwidth by modifying the multi-homing feature of SCTP in case of link failure. Handover triggering decisions can consider parameters other than signal strength for initiate the handover process in case of fast moving MNs. Congestion control mechanisms can be integrated with a mobility protocol so that smooth and seamless handover can be achieved. mSCTP can be collaborated with other layer mobility solutions to provide lower handover latency and higher throughput. The related work done on mSCTP on the above mentioned approaches are described below:

A. Multi-path Transfer of Data for Aggregating the Bandwidths of All the Paths

Many researchers have Multi-path feature of SCTP had been exploited multi-path feature of SCTP to provide multi-path data transmission. Multi-path data transmission is done in order to improve the bandwidth by sending the data on multiple paths. Load Sharing–SCTP (LS-SCTP) scheme [30] is the multipath based transport layer scheme that aggregates the bandwidth of all the paths and introduces a path congestion control by performing per-association, per-path data unit sequence numbering. But it requires the change in SCTP packet format. Independent Per-Path Congestion Control SCTP (IPCC-SCTP) scheme [31] was almost same as that of LS-SCTP scheme but it did not modify the packet format by keeping the state information of sender in the receiver. In Westwood SCTP scheme [32] the bandwidth for congestion control among the disjoint paths in the association has been estimated. SCTP- Concurrent Multipath Transfer (CMT) scheme [33] has proposed a multi-buffer sender structure and it focused on improving throughput by exploring retransmission policies for CMT and congestion was controlled per-path not per-association but it introduced Receiver Buffer (Rbuf) blocking problem (receiver buffer is filled out of order data). Rbuf problem was partially solved by CMT-PF (Potentially Failed) [34] which has marked the path that has experienced a failure as PF and stops communication on that path until a positive HB is acknowledged.

All the protocols discussed above are for improvement of throughput but do not handle mobility issues like path handover. Mobile Multi-path SCTP [35] allows a MN to perform two path handover when multipath transmission is adopted in wireless mobile networks. It deals with three concerns related to path handover that includes spurious retransmissions due to failed SACK, retransmissions of data lost before path handover and unnecessary congestion window reductions and the reordering problems due to path handover. Dynamic DNS approach has been used for location management of MN. Concurrent multi-path SCTP (empSCTP) [34] adds CMT and dynamic Multi-path Management support to SCTP for realizing SCTP handover. mSCTP-CMT [36] has been proposed to apply CMT to distribute data among two end to end paths of a mSCTP association during the handover process. Mobile CMT (mCMT) [37, 38] adds path-oriented multi-streaming to tackle Rbuf blocking problem with MIH (Media Independent Handover) [39] assisted handover to reduce handover loss and Association Resume (AR) to resume the mCMT association if it disconnects with the remote endpoint in the case of hard handover.

B. Handover for Fast Moving Mobile Nodes

Some research has also been done on mSCTP for very fast moving mobile nodes. Enhancement to mSCTP was made by adaptive primary switching scheme. This scheme had a better performance when MN is moving at various speeds between networks. Switching of primary path becomes important and this scheme switches between current primary path and the path of new network according to Round Trip Time (RTT) [40] and the velocity of movement of MN. A Smooth Handover Scheme based on mSCTP (SHSBM), better the performance of Seamless IP diversity based Generalized Mobility Architecture (SIGMA) [41] which performs poorly when the MN is moving at high speed by utilizing buffer and tunnel scheme. Fast Handover with Congestion control window size estimation (FHO-CE) [42] used cellular link as primary path when it is out of range of previous access router and new association is still not established. Comparison of these three techniques is done in Table 3. FHO-CE is implemented in case of MNs in Vehicular Ad-hoc Networks (VANETs) while the remaining two techniques are for general mobile nodes.
C. Congestion Control Mechanisms

Due to difference in the bandwidths of heterogeneous networks packets can come in out-of-order after vertical handovers which can further cause congestion. Also when the MN enters in the new network after handover it enters into a slow start phase, thus reduces data rate. SMART-FRX [14] technique has been proposed to improve performance of mSCTP during WLAN to cellular forced vertical handovers. It specifies a technique for when to switch on alternative method to reduce error losses and packet losses. Congestion control techniques [43] have been proposed by estimating the bandwidth available on the available path and then calculating congestion window. mSCTP with bicasting bicasts [44] the data packets to old IP and new IP addresses in the vertical handover region to reduce the packet loss. Enhancement of mSCTP has been made [45] to reduce handover latency and packet loss rate by eliminating ASCONF delete chunk and introducing handover SACK to inform handover. Congestion control parameter update and buffer transmission was combined to avoid reduction of data rate and packet re-ordering. Due to packet re-ordering, spurious fast retransmissions and unnecessary reduction of congestion window was done during mSCTP handover. To solve this, a new scheme [46, 47] has been introduced in which the CN transmits the outstanding data chunks just before the transmission of the newly generated data chunks over the new primary address.

D. Cross Layer Handovers

Network oriented solutions can provide handovers between heterogeneous Radio Access Networks (RANs) but each access network must be part of the same network operators infrastructure. By shifting the complexity out of the network and into the mobile devices, endpoint centric handover solutions have been developed. Endpoint Centric Handover (ECHO) [48] is a SCTP based handover solution for Voice over IP (VoIP). It has combined the ITU-T E-model for voice quality assessment to accurately estimate the Quality of Service (QoS) of candidate handover networks, thus facilitating a more intelligent handover decision. An effective handover management scheme for VoIP communication [1] makes handover decisions based on the number of retransmissions experienced by a data frame on layer 2. This scheme selects single-path transmission or multi-path transmission according to the number of frame retransmissions. mSIGMA [49] has been proposed to implement mobility in multicell networks with low delay and packet loss with efficient location management. Various mobility protocols are also combined with mSCTP such as:

(a) Integrating mSCTP with FMIPv6

To improve handover latency, features of mSCTP and FMIPv6 are integrated by some researchers. With DAR signal in mSCTP the MN can change its IP address dynamically [50]; therefore requirement of creating a tunnel does not exist. This reduces the handover latency significantly. Enhanced DAR (EDAR) is also implemented with FMIPv6 to reduce handover latency further. The main difference between DAR and EDAR is that in EDAR it sends both ADD-IP ASCONF and Set-Primary ASCONF messages simultaneously [51] as shown in Figure 5.

(b) Integrating with HMIPv6

Hierarchical Transport layer Mobility protocol (HTM) deals with local and global mobility. HTM provides low handover latency, greater throughput and reduces signaling cost by exploiting DAR feature of SCTP and introduces an Anchor Mobility Unit in order to achieve the smooth and complete handover in case of horizontal and vertical handovers [52].

![Figure 5: Message Flow in EDAR-FMIPv6](image)

(c) Integrating with SIP

A novel hybrid method for mobility called Hybrid–NEMO [53] has been proposed which provides a soft handover scheme at the transport layer by combining the
SIP and SCTP protocols to reduce handover latency and packet loss which are critical in providing QoS voice and multimedia applications.

VI. CONCLUSION

In this paper various network layer mobility protocols (MIP and its variants) and transport layer mobility protocol like mSCTP for mobility support have been compared. While mSCTP provides better handover latency and throughput, it requires reordering of packets as in comparison with all network layer protocols. Some enhancements to mSCTP protocol have also been reviewed that provide lesser handover latency and avoid packet loss and can improve bandwidth by concurrent multi-path transfer of data, so that optimal handover can be achieved. Further the transport layer handover techniques based on network parameters like bandwidth, velocity, handover latency, etc., needs to be proposed for smooth handovers in the case of fast moving MNs. Congestion control mechanisms with the proposed techniques can be implemented after the completion of handover to improve the overall network performance.

REFERENCES


© 2013 ACADEMY PUBLISHER


