Evaluation of End-to-End TCP Performance for Vertical Handover using Intermediate Switching Network

Aditya Yadav, Maushumi Barooah, Sandip Chakraborty*, Sukumar Nandi, Sanjay Ahuja

Department of Computer Science and Engineering, Indian Institute of Technology Guwahati
Assam India 781039,
Email:{y.aditya,maushumi,c.sandip,sukumar,s.ahuja}@iitg.ernet.in
*Corresponding author

Abstract: Most of the traffic in today’s network uses the Transmission Control Protocol (TCP) as the transport layer protocol for reliable end-to-end packet delivery. However, TCP considers packet loss to be the result of network congestion which makes it unsuitable for mobile wireless communication, where sporadic and temporary packet losses are usual due to fading, shadowing, hand-off and other radio effects. During the vertical hand-off between different wireless technologies, the problem of end-to-end connection and reliability management for TCP becomes more severe. Intermediate Switching Network (ISN) is an effective framework for vertical handover between the UMTS and the WLAN network. This paper evaluates the performance of TCP over ISN based framework. The proposed scheme uses a cross layer interaction between the data link and the transport layers to estimate TCP retransmit timeout and congestion window during handover. Simulation results establishes effectiveness of the proposed scheme.

Keywords: ISN; Vertical Handover; TCP; Congestion Window; RTT; MPLS-MP-BGP

1 Introduction

Transmission Control Protocol(TCP) is the most widely used end-to-end transmission layer protocol for today’s data network to provide reliable and service-oriented data delivery. At the time when TCP was designed, it was mainly made to perform on wired networks. So the packet loss was considered as the information of the congestion in the network, and upon detection of the packet loss, TCP reduces the sending rate to cope up with the network congestion. In wireless networks, the communication medium is error prone, and packet loss can occur because of several reasons like the channel fading, shadowing, hand-off and other radio effects. TCP faces several problems in the wireless networks [27], which can be summarized as follows,
1. Packet losses due to the radio effect is taken as congestion, and TCP unnecessarily drops its congestion window (CWnd). The problem becomes severe for the high bit error rate (BER) and error prone channels, where CWnd size tends to stay small for a long period of time.

2. TCP uses an exponential back-off mechanism for packet retransmission. With small CWnd size, the exponential back-off increases the retransmission timeout (RTO), resulting in long periods of silence or connection break-down.

3. TCP uses a set of retransmission timers at the transport layer, which are independent from the link layer timers. Uses of independent timers at the link layer and the transport layer may trigger unnecessary retransmissions.

The problems of the TCP become more severe during the vertical handover between two different wireless technologies. Vertical handover introduces additional problems over TCP as follows:

1. Different wireless technologies operate in different data rates. For example, IEEE 802.11 WLAN can operate up to 600 Mbps for IEEE 802.11n, whereas the UMTS operates theoretically up to 42 Mbps when high speed packet access (HSPA) is implemented in the network. However, in practical, users of an UMTS network can expect the transfer rate up to 7.2 Mbps with HSDPA handsets. Thus there is a significant gap between the data rate for an UMTS network and a WLAN network. After handover, TCP resumes its old values of CWnd and RTO, which are not suitable for the new network. So the performance degrades severely when handover occurs from a slow network to a fast network because of the low CWnd and high RTO of the old network. Similarly several packet drops and connection break-down occurs due to the spurious timeouts, when handover occurs from a fast network to a slow network. Spurious timeout is defined as a timeout which would not have happened if the sender waited long enough. It is a timeout resulting in retransmission due to a segment being delayed, but not lost, after RTO expires [23].

2. With a high-latency link technologies, such as the UMTS and its high buffering, it takes several seconds for the TCP CWnd to reach the new path capacity [12].

This paper evaluates the TCP performance for vertical handover between an UMTS and a WLAN networks, using the ISN framework. In [4], the performance of the TCP has been analyzed using ISN framework for GPRS-WLAN handover. An improved version of the TCP, called ISN-TCP, has been proposed in [4]. This paper extends the analysis for the UMTS-WLAN handover using ISN framework, and proposes an improved TCP variant, namely ISN-TCP-PLUS. In ISN-TCP-PLUS, TCP connection maintenance during handover is handled using a cross-layer interaction between the transport and the network layer, and estimation of updated TCP parameters after handover. By filtering out the duplicate ACKs (DACKs) and calculating the RTT of the new network on-the-fly, ISN-TCP-PLUS improves the performance of the TCP significantly during handover. The proposed scheme uses an approximation in CWnd calculation, similar to the approaches used
for the equation based TCP congestion control [13, 29]. The performance of the ISN-TCP-PLUS has been analyzed using simulation results.

The rest of the paper is organized as follows. The state-of-the-art works proposed till now in literatures are summarized in section 2. Section 3 provides the background of the ISN based vertical handover framework. The performance of the WP-TCP over ISN based framework during handover has been reported and analyzed in section 4. In section 5, a brief overview of ISN-TCP is provided with detail analysis of performance evaluation through simulation results for the WLAN to the UMTS handover and vice-versa, using the ISN based framework. The proposed improvement of the TCP variant for the UMTS to the WLAN handover and vice-versa, using the ISN based framework, is discussed in section 6. Section 7 reports the performance of the ISN-TCP-PLUS using simulation results. Finally, section 8 concludes the paper.

2 Related Works

The performance of the TCP protocol over the wired and wireless technologies has been widely studied in the literature [3, 18, 21, 24, 25, 26, 28, 31]. To handle the problems of TCP over wireless network, a TCP variant is introduced, called Wireless Profiled TCP (WP-TCP) [32]. WP-TCP uses large CWnd size based on the bandwidth-delay product (BDP). The delay for the BDP is assumed to be sufficiently large to handle channel fading and shadowing. However, increasing CWnd sizes may introduce problems during vertical handover [32]. First, during handover, increasing the maximum window size improves the efficiency in a high BER environment, but degrades the efficiency in a low BER environment. Second, depending on the duplicate ACK threshold, increasing the CWnd may also increase the chances of false fast-retransmits during the vertical handover.

Several works exist in the literature that deal with the TCP performance for the vertical handover [9, 10, 11, 16, 17, 19, 20, 22, 36]. However, all of these works are based on mobile IP (MIP) based handover schemes between two different wireless technologies. Most of these works use probing based mechanism after handover, based on the MIP framework, to cope up with the new channel characteristics. However, the MIP based handover faces the problem of triangular routing, and the layer-3 handover operates independently from the link layer handover. This introduces high handover delay in the network, and after the handover, the end-to-end delay increases because of the triangular routing. In [5], Barooah et al. have proposed a new framework for the vertical handover between an UMTS and a IEEE 802.11 WLAN networks. Their proposed framework is based on Intermediate Switching Network (ISN), where the ISN connects the UMTS network with the WLAN network. Multi-protocol Label Switching (MPLS) and Multi-protocol Border Gateway Protocol (MP-BGP) have been used for the packet forwarding. The layer-3 handover is executed using the mobility binding information. The ISN based handover scheme reduces the handover delay significantly compared to the MIP based handover scheme.

In [33], the authors have discussed about several TCP variants for a mobile wireless network. Out of them, Split-TCP [35] is a widely accepted variant for the TCP used in the mobile network. Split-TCP proposes to setup multiple proxies
along the path of the TCP connections, and the lost packets can be recovered from
the most recent last proxy. Thus after handover, only the link properties between
the Mobile Node (MN) to the first proxy needs to be updated. Split-TCP is not
suitable for the ISN based handover framework, because the path through the ISN
changes based on the mobility binding updates. Split TCP also needs to transfer
the state information in case of the handover, which increases the handover time.
It also violates the end-to-end TCP semantics.

Another variant of the TCP for vertical handover is proposed in [8, 30] that
freezes the TCP parameters during handover. This modification of TCP is termed
as the freeze-TCP. Freeze-TCP solves the problem of the CWnd dropping by
freezing the TCP parameters to avoid unnecessary triggering of the congestion
control actions during the handover. After the handover gets complete, TCP
resumes its old values of the congestion control parameters, and normal TCP
procedure continues. The problem with freeze-TCP is that, it requires considerable
amount of time to adopt to the new link characteristics after the handover gets
completed.

The existing TCP variants for the vertical handover, as mentioned earlier, are
not applicable for the ISN based framework, because of the reasons as follows,

1. Most of the existing schemes use the network probing mechanism to find out
the round-trip time (RTT) after the handover occurs [10, 16, 17, 20, 22].
For the MIP based handover, the network probing is mandatory because of
the triangular routing. However, the ISN framework uses the MPLS and the
mobility-binding for the layer-3 handover, and avoids the triangular routing.
Thus network probing is not required for the calculation of the RTT. A more
optimized scheme can be deduced for the RTT estimation, as reported in this
paper.

2. Some of the approaches for the vertical handover use the “make before break”
strategy [11] where the TCP first adopts the properties of a new connection,
and then the layer-3 handover takes place. This approach introduces extra
network delay which may not be tolerable for different applications. ISN
based framework is built upon the “break before make” strategy where the
layer-3 handover occurs after the layer-2 handover gets completed. After the
layer-2 handover, the mobility binding updates are used to find out the layer-
3 paths. The proposed ISN framework uses an “Alternate PDP Context” to
forward the data after the handover to the UMTS network in a slow data
rate, until the desired data rate is achieved. The analysis in [5] shows that the
time for setting up the “Desired PDP Context” is within the tolerable limit
for most of the applications. TCP can also adopt to the new environment
within this time duration.

3. Zhang et al. [36] uses a scheme where the link characteristics information
is piggybacked within the IP mobility packets. Similarly, the authors in [19]
solve the packet reordering problem using a TCP aware agent at the mobile
nodes. These schemes have their limitations on the MIP based handover
framework, and can not be adopted for the ISN based framework.

ISN-TCP [4] is the first solution that handles the problems of the packet
reordering, spurious timeouts, packet losses and the network under/over utilization
for the TCP connections during the WLAN-GPRS handover through the ISN framework. It introduces an additional layer at the MN and the corresponding node (CN), between the transport and the network layer. The cross layer information is used to trigger the TCP for handover related actions. During the handover from the GPRS to the WLAN, the MN first estimates the new RTT, and then updates the link parameters before sending any new packet. Similarly the CN updates the link parameters. During the handover from the WLAN to the GPRS, the MN at TCP freezes the connection to avoid any buffer over-flow. The connection is resumed as the link characteristics are updated.

ISN-TCP requires freezing to maintain the consistency in connection during the handover, which introduces extra delay in the network. Further, this TCP variant is designed for the GPRS to the WLAN handover or vice-versa. The freezing delay is tolerable for the slow GPRS connections. However, the delay is not tolerable for the UMTS network, which operates in a much higher data rate compared to the GPRS network. ISN-TCP can solve spurious timeouts and network under/over utilization for the GPRS network. However, the scheme fails for the UMTS network, where the link characteristics should be updated using a faster method, and the connection freezing is not acceptable at all.

3 Background: Intermediate Switching Network

![Figure 1](image)

Figure 1 Integrating of ISN with the UMTS and the WLAN Networks [5]

Figure 1 shows the integration of ISN with the UMTS and the WLAN networks. ISN is a mesh connection of the Label Edge Routers (LER) with the Mobility Support Functions (MSF) associated with them. The routers are termed as MSF-LERs. The MSF-LERs uses the Border Gateway Protocol (BGP) for
information update among them. Each of the MSF-LERs is a BGP speaker and a BGP Peer to every other MSF-LER in the ISN, thus forming a complete mesh. The MSF-LERs are connected with each other via Label Switched Path (LSP) tunnels. The integration points of ISN with the UMTS and the WLAN network are the Gateway GPRS-3G Support Node (GGSN/MSF-LER) and Access Point (AP/MSF-LER) respectively. MN has a dual stack with MSF operating between the IP and the SNDCP layers for the UMTS, and between the IP and the LLC layers for the WLAN. The handover procedure is initiated by the MN. For the sake of completeness, the handover procedure is briefly described here. The detailed handover procedure can be found in [5].

Handover between the UMTS and the WLAN networks is initiated by the MN depending on the presence or absence of a stable WLAN network. MN starts up in the WLAN and sends ICMP registration message containing the pair \([MN \text{ IP Address}, MAC \text{ Address}]\) to AP/MSF-LER. AP/MSF-LER generates a unique Mobility Label (ML) and sends it back to the MN. AP/MSF-LER generates a Mobility Binding \([ML, MN \text{ IP Address}, AP \text{ IP Address}]\) and distributes it to all other MSF-LERs for storage, using BGP update message. MSF-LER from ISN forwards IP datagrams destined for the MN IP Address to AP/MSF-LER using MPLS LSP tunnels. AP/MSF-LER uses ML to identify the MAC Address for the MN and forwards packets to the MN using Layer 2 path. Registration here also contains the setting up of an “Alternate PDP Context” in the UMTS network (using dummy ICMP messages). After handover, the “Alternate PDP Context” is used initially to transmit data with a lower data rate.

3.1 Handover from the WLAN to the UMTS

When the MN moves out of the WLAN network, it registers with the ISN via GGSN/MSF-LER using an “Alternate PDP Context” containing the pair \([ML, MN \text{ IP Address}]\). In the UMTS network the Layer 2 path is provided by the PDP Context. The GGSN/MSF-LER then creates the mobility binding \([ML, MN \text{ IP Address}, GGSN \text{ IP Address}]\) and sends it to other MSF-LERs using BGP update message. MSF-LER uses LSP tunnels to redirect traffic to the GGSN/MSF-LER. The GGSN/MSF-LER uses ML to get the “Alternate PDP Context” and uses it to forward the IP Datagrams to the MN. Simultaneously, MSF at GGSN also generates a “Desired PDP Context” (with desired data rates). Once “Desired PDP Context” is established, it is used to forward IP Datagrams.

3.2 Handover from the UMTS to the WLAN

When the MN moves into a WLAN network, the registration process using the same ML is carried out again with AP/MSF-LER. The mobility binding sent out by AP/MSF-LER prompts to switch traffic from ISN to AP/MSF-LER which forwards packets to the MN using the WLAN MAC Address.
Table 1  Simulation Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Application Traffic</td>
<td>FTP/GENERIC</td>
</tr>
<tr>
<td>Packet Size</td>
<td>512 Bytes</td>
</tr>
<tr>
<td>Number of Packets</td>
<td>40000</td>
</tr>
<tr>
<td>Mobile Node Speed</td>
<td>1.2m/s</td>
</tr>
<tr>
<td>PDP Context Max Idle Time</td>
<td>45s</td>
</tr>
<tr>
<td>Registration Lifetime</td>
<td>20s</td>
</tr>
<tr>
<td>Max Number of Registration Retries</td>
<td>3</td>
</tr>
<tr>
<td>Registration Retry Interval</td>
<td>3s</td>
</tr>
</tbody>
</table>

4 Performance Evaluation of TCP for ISN Based Framework

The TCP performance is evaluated over ISN based framework using Qualnet 5.0.1 [34] network simulator. The simulation scenario is based on ISN architecture shown in Figure 1. Two scenarios have been used for simulation,

1. In first case, initially the MN is in the UMTS network and moves into coverage area of AP1 of the WLAN network as shown in Figure 1. It finally moves inside the coverage area of AP2 of the WLAN network.

2. In second case, initially the mobile node is in the WLAN network and moves out of the WLAN network. It finally stops in the UMTS network.

The TCP version used for simulation is TCP new-Reno with wireless profiled modifications (WP-TCP) as proposed in [32]. WP-TCP is implemented with all mandatory requirements (RFC 1112 [7], RFC 2581 [2] and selective acknowledgement [6]) and some important optional requirements such as large initial window (RFC 3390 [1]) and time-stamp option (RFC 1323 [15]) for RTT measurement. The maximum CWnd size is taken as 31 packets and DACK threshold is taken as 5 packets. The MN acts as the TCP receiver and the CN acts as the TCP sender. The simulation parameters are shown in Table 1.

To evaluate the performance of TCP during handover, the CWnd and duplicate packets sent during handover have been measured.

4.1 UMTS to the WLAN Handover

The handover from the UMTS to the WLAN network starts at 474.8 seconds. The simulation graphs are shown in Figure 2 and Figure 3. During the handover significant amount of packets are dropped from the UMTS interface. It can be seen from Figure 2 that, after the handover procedure gets completed, duplicate packets are transmitted, and the packets are delivered out of order. As a result, CWnd drops to its initial value, as shown in Figure 3. It has been observed from the simulation traces that 24 Kbytes of data have been retransmitted due to the handover. From extensive simulation results it has been also observed that amount of data retransmitted increases with the increase of packet size and the number of simultaneous TCP connections. It can also be observed from Figure 3, that CWnd
4.2 WLAN to the UMTS Handover

The handover procedure from the WLAN network to the UMTS network is initiated at 1150 sec. During the handover, multiple RTO occur due to large difference in the RTT values in the UMTS network and the WLAN network. In Figure 4, data receiving stops at 1150.67 secs and data receiving again starts at 1156.14 sec. This break in the transmission is because of the multiple RTOs that occur at the sender side. At every retransmit timer firing event, TCP sender reduces the congestion window to one segment, retransmits that segment, and doubles up the retransmit timer. The drop in congestion window is shown in Figure 5. In the present simulation scenario, three consecutive RTOs occur before the first packet is successfully delivered after the handover. It has been observed from the simulation traces that even after doubling up the retransmit timer, the updated RTO value is not sufficient for the UMTS network. Hence a mechanism is required for the TCP RTT estimation after the handover, that overcomes this drawback.
The observations from simulation analysis of WP-TCP over ISN based vertical handover framework is summarized below.

1. During the UMTS to the WLAN handover, CWnd drops to its initial value because of two reasons,
   
   (a) The RTO at the UMTS network is less than handover delay. Thus during handover, the retransmission timer expires, and TCP resets CWnd value to its initial value.
   
   (b) Significant number of packets are dropped from the UMTS interfaces during handover. These TCP packets are retransmitted through the WLAN interface after the handover gets completed. This is because, the TCP sender (here the CN) is not aware of the impending handover. The packet loss during the handover generates DACKs. Thus the DACK threshold (which is normally 5 packets for WP-TCP) gets expired, and TCP resets its CWnd.

2. Premature RTOs occur during the handover from the WLAN to the UMTS. This is because RTOs are calculated based on the RTT values. The RTT for the WLAN network is much lower compared to the UMTS network. Because of these premature RTOs, CWnd drops to one segment, and TCP goes to the slow-start phase.

5 ISN-TCP and Its Performance for Handover between the UMTS and the WLAN

To solve the problem of duplicate packet delivery and premature RTOs, ISN-TCP [4] is proposed for GPRS-WLAN handover using ISN based framework. In this section, the performance of ISN-TCP is evaluated for the handover between an UMTS and a WLAN networks. A brief description of the ISN-TCP is reproduced from [4] for shake of completeness. In ISN-TCP, a sub-layer has been introduced between the transport and the network layer, at the two end nodes of a communication path, with minimal dependencies on the intermediate nodes. No additional functionality is required in the network architecture. Upon notification of handover, the intermediate layer executes two processes at the MN, called \( PR_{MN} \) as receiver process and \( PS_{MN} \) as sending process. Two similar processes are executed at CN, called \( PR_{CN} \) and \( PS_{CN} \). \( PR_{MN} \), \( PS_{MN} \), \( PR_{CN} \) and \( PS_{CN} \) interact with the corresponding TCP process at the MN (\( TCP_{MN} \)) and the CN (\( TCP_{CN} \)).

5.1 Handover from the UMTS to the WLAN

The detailed handover procedure from an UMTS network to a WLAN network is as follows.

1. When the MN becomes aware of an impending handover with the receipt of a Layer 2 trigger, the \( PR_{MN} \) process saves the sequence number of the last packet (\( SeqL_{MN} \)) on the old interface and then sends a Handover Notification packet to the \( PR_{CN} \).
2. The $PR_{CN}$ process also saves the sequence number of the last packet ($SeqL_{CN}$) on the old network and handover begins.

3. The MN sends the path establishment notice after handover to the $PR_{MN}$ process. The MN estimates the RTT and sends a path establishment notification to the $PR_{CN}$. MN also updates the parameters at $TCP_{MN}$. This notification includes the estimated value of the bandwidth and the RTT by the MN.

4. $TCP_{CN}$ estimates the RTT value and updates parameters. Both the MN and the CN save the sequence number of the first packet ($SeqF_{MN}$ and $SeqF_{CN}$) on the new interface.

5. After this, the MN forwards packets with the sequence number $SeqL_{MN}$ to $SeqF_{MN}$ to the CN. Similarly the CN forwards packets with the sequence number $SeqL_{CN}$ to $SeqF_{CN}$ to the MN.

While handing over from an UMTS network to a WLAN network, the problem faced are mainly of the packet loss, packet reordering and the network under utilization. Since the RTT of the WLAN is much smaller than that of the UMTS network, the packets arrive out-of-order at the receiver generating DACKs to falsely trigger TCPs fast retransmission of the in-flight packets. ISN-TCP solves this problem by closing the old link immediately after Handover and falsely create a situation of packet loss to avoid the problem of packet reordering. $TCP_{CN}$ is also notified to avoid taking any congestion avoidance actions. The CN and the MN know exactly what packets are to be resent as the sequence number of the last packet arrived on the old interface before handover and the first packet on the new interface after handover are stored. After the new path is established, the leftover packets are sent from either side. So packets do not come out-of-order in ISN-TCP.

5.2 Handover from the WLAN to the UMTS

The detailed procedure is as follows,

1. When the MN sends a Layer-2 trigger to the process $PR_{MN}$ indicating a possible handover, it sends a handover notification packet including type of
handover and Zero Window Advertisements (ZWA) to the $PS_{CN}$. A sender that receives a ZWA must stop sending further packets until it receives a positive window.

2. The $PR_{MN}$ process at the MN freezes its TCP process.

3. The $PS_{CN}$ process notifies about the impending handover to its $TCP_{CN}$ process and freezes it.

4. After the Handover is complete, the MN sends the path establishment notification to the $PR_{MN}$ and estimates RTT and CWnd. These values are updated at $PR_{MN}$.

5. A notification of new path establishment by the MN is sent to the $PR_{CN}$ process and estimation of RTT takes place by CN followed by the updation of parameters.

6. This update notification is sent to the $PS_{MN}$, and $PS_{CN}$ sends unfreeze notification to $TCP_{CN}$ process and TCP becomes active once again with new parameters.

Here, while handing over from the WLAN to the UMTS, the problem faced is of spurious timeouts. The drastic increase in RTT results in spurious timeouts and as a result TCP enters the slow start phase by reducing the CWnd to 1 and injecting too many packets into the slow network. In ISN-TCP both the CN and the MN measure RTT and bandwidth after handover and the estimation of RTT at new interface takes place. For TCP to adapt readily to the new link characteristics, ISN-TCP estimates RTT and available bandwidth.

5.3 Analysis of ISN-TCP for the UMTS-WLAN Handover

The performance of ISN-TCP for the UMTS-WLAN handover is evaluated using the similar simulation scenario as described earlier in section 4.

5.3.1 Handover from the UMTS to the WLAN:

The simulation results for ISN-TCP is shown in Figure 6 and Figure 7. From the simulation analysis for the UMTS to the WLAN handover, it has been observed that ISN-TCP can reduce number of CWnd drops during handover. However, it can not eliminate CWnd drops fully. From the simulation analysis, it has been observed that amount of retransmissions increases for ISN-TCP. In the present simulation scenario, there are 35KB of retransmissions per TCP connection. The amount of retransmission increases because ISN-TCP retransmits packets that were not acknowledged through the old-interface. Because of these increased amount of retransmissions that generates DACKs for the sender, the CWnd is dropped. Amount of retransmissions was not a problem for GPRS network, as shown in [4]. GPRS network is much slower compared to the UMTS network, and thus the amount of retransmissions is very less in case of GPRS to the WLAN handover. It has been observed from extensive simulations that the amount of retransmission increases with the increase of TCP packet size and number of active TCP connections during handover.
5.3.2 Handover from the WLAN to the UMTS:

The performance of ISN-TCP for the WLAN to the UMTS handover is shown in Figure 8 and Figure 9. As ISN-TCP freezes the TCP connection, it doesn’t cause successive time outs as in the case of WP-TCP. However estimating the parameters after handover and freezing of connection affects the TCP performance as shown in Figure 8. In Figure 8, data receiving stops at mobile node at 1150.67 secs and data receiving again starts at 1155.3 sec. This break in transmission is because of freezing the transmission during handover and estimating the parameters of network. The freezing time is acceptable for GPRS network, but it affects TCP performance in the UMTS network.

The performance of ISN-TCP for the UMTS-WLAN vertical handover can be summarized as follows;

1. To solve the problem of packet loss and packet reordering at the time of the UMTS to the WLAN handover, ISN-TCP saves the sequence number of the last packet acknowledged through the UMTS interface before handover, and sends all the packets after that sequence number through the WLAN interface, once the handover gets complete. Though this can solve packet loss and packet reordering, it generates lots of duplicate transmissions. This is because, the packets which were sent through the UMTS interface before handover, but not acknowledged, are retransmitted. This large amount of retransmission causes DACKs, and the expiration of DACK threshold. This causes TCP to reset its CWnd.

2. Spurious RTOs during the WLAN to the UMTS handover are avoided in ISN-TCP by freezing the TCP parameters and estimating the new value after handover. Though the freezing time is tolerable for GPRS network, but it affects TCP performance in UMTS network.
6 ISN-TCP-PLUS: Improved TCP for ISN based Handover Framework

In the proposed TCP modification, a Handover layer is added between the transport and the network layer, which implements the solution to improve the TCP performance. This Handover layer is added only at the end devices.

It can be noted that the handover affects mostly the TCP connection from the CN to the MN as the CN is not aware of the impending handover. The connections from the MN to the CN suffer less because the handover is initiated by the MN. That is why the modifications in TCP is reported with respect to the connection from the CN to the MN. The reverse way connection can be handled in a simpler way.

6.1 Handover from the UMTS to the WLAN

In this case, DACKs are generated because of out-of-order packet delivery. TCP perceives three DACKs as congestion, and reduces the CWnd. Due to congestion control mechanism, TCP also makes unnecessary retransmissions. While transferring data from the CN to the MN, CN can receive the DACKs from the MN, which will trigger congestion control in CN. In ISN-TCP [4], the UMTS interface is immediately closed after the handover and all the outstanding packets are retransmitted on the WLAN interface. This leads to unnecessary retransmission, which can be avoided. Simulation results in Section 4 and Section 5 show that there are significant number of retransmissions per TCP connection in the case of WP-TCP and ISN-TCP.

ISN-TCP-PLUS filters out unnecessary DACKs so that the congestion control actions are not triggered at CN during handover. The sequence of actions that are triggered by the handover layer of ISN-TCP-PLUS at the MN and the CN are as follows;

1. When the MN starts to handover from the UMTS to the WLAN network, there is a cross layer interaction between MAC layer and the handover layer at the MN, which sends Handover Notification message to the handover layer at the MN. In turn, the handover layer at the MN sends a handover notification to the CN.

2. On receiving the handover notification message from the MN, the CN saves the sequence number of the packet last acknowledged in lastAcked variable, and TCP transmission continues.

3. When the handover completes at the MN, handover completion notification message from the MAC layer is sent to the handover layer at the MN, which again sends a handover completion notification to the CN.

4. On receiving the handover completion notification, the handover layer at the CN saves the highest sequence number sent so far in HigestSeqSent variable. Now the handover layer at the CN monitors the ACKs received. If the sequence number of the ACK received at handover layer lies between lastAcked and HigestSeqSent, then it is a DACK, and the ACK frame is dropped.
5. When the sequence numbers of ACKs become greater than \( HigestSeqSent \), then handover layer stops the filtering of the ACKs and the normal TCP operation resumes.

As the DACKs are the wrong indicators for the network congestion in the case of the handover from the UMTS to the WLAN, the handover layer filters out the DACKs between \( lastAcked \) and \( HigestSeqSent \). TCP doesn’t invoke congestion control as the DACKs are filtered out by the handover layer. If packets on the UMTS interface are lost permanently, then TCP will eventually time out and retransmit those packets. Filtering period at the handover layer could maximum go up to the RTO in the UMTS network (\( RTO_{UMTS} \)). TCP does not immediately change to the updated RTT values of the WLAN networks on handover, because this may lead to the spurious timeouts. As DACKs are filtered, TCP perceives the same RTT as in the UMTS network. So filtering the ACKs does not cause spurious timeout at the TCP sender. If the ACKs are piggybacked, then instead of filtering the ACKs, the ACK bit of the data packet can be reset. In this method, no change to existing TCP implementations is required, and only a handover layer below the TCP is required to be implemented at the end devices.

6.2 Handover from the WLAN to the UMTS

During handover from the WLAN to the UMTS network, the problem is spurious RTOs. The RTO at the WLAN network (\( RTO_{WLAN} \)) is lesser than \( RTO_{UMTS} \). The ACKs for the packets transmitted by the TCP sender in the WLAN network, just before the handover, arrive through the UMTS network. RTO perceived by the TCP sender in the WLAN network is significantly lesser than the current RTT for ACKs through the UMTS network. Therefore, TCP triggers RTOs, drops CWnd to 1, retransmits the packet, and exponentially backs off the retransmission timer. This leads into a coarse packet transmission in the period after the handover and reduces TCP performance.

Moreover, even after exponential back-off of the RTO, multiple retransmit timeout occur before the TCP estimates the RTT of the UMTS network (\( RTT_{UMTS} \)). In ISN-TCP, TCP connection is freeze during the handover and is unfreeze only after the measurement of the required parameters after the handover. This affects the TCP performance as shown in section 5. Performance degradation by freezing the TCP connection would be more in the case of “make before break” handover [11]. The RTT calculation method proposed in this section can be used to estimate the RTT of probable networks before the handover. It can also be used to estimate the end-to-end delay parameter of the QoS, while scanning for available networks. In ISN-TCP, RTT calculation can only be performed by freezing the connection and after the handover has taken place. WP-TCP also estimates the RTT by usual mechanism after triggering many RTOs. An improved version of the RTT estimation mechanism is proposed in this paper that significantly improves the TCP performance.

ISN-TCP-PLUS uses this RTT estimation to update the TCP parameters during handover. The detailed actions performed by ISN-TCP-PLUS for the WLAN to the UMTS handover are as follows;

1. The MAC layer at the MN sends a handover notification to the handover layer at the MN about the impending handover.
2. The handover layer at the MN, on receiving the handover notification, measures the the RTT between the MN and the WLAN AP ($RTT_{MN-AP}$), by sending ICMP ECHO message to the current WLAN AP. Here either multiple ICMP ECHO requests can be sent and average can be taken, to get accurate $RTT_{MN-AP}$, or the handover layer can continuously monitor $RTT_{MN-AP}$ and apply TCP like smoothing of the estimated RTT.

3. The handover layer also sends an ICMP ECHO request to the GGSN through the UMTS interface, and measures the RTT between the MN and the GGSN ($RTT_{MN-GGSN}$).

4. After measuring $RTT_{MN-AP}$ and $RTT_{MN-GGSN}$, the handover layer at the MN sends a update parameter message to the CN, with these values.

5. The CN, on receiving handover update parameter from the MN, does a cross layer interaction with the TCP, and saves the current estimated RTT (smoothed RTT) in variable $EstRTT_{WLAN}$. Then the handover layer estimates the $SampleRTT_{UMTS}$ by equation (1),

$$SampleRTT_{UMTS} = EstRTT_{WLAN} - RTT_{MN-AP} + RTT_{MN-GGSN}$$  \hspace{1cm} (1)

6. This gives the Sample RTT of the UMTS network. Now this can be used to calculate $EstimatedRTT_{UMTS}$ using a similar procedure as [14]. Initially the value of $Deviation_{UMTS}$ is set as half of the $SampleRTT_{UMTS}$. Then the difference between sample RTT and estimated RTT ($Difference_{UMTS}$) for the UMTS network can be calculated by equation (2),

$$Difference_{UMTS} = SampleRTT_{UMTS} - EstimatedRTT_{UMTS}$$  \hspace{1cm} (2)

Initially $EstimatedRTT_{UMTS} = SampleRTT_{UMTS}$. $EstimatedRTT_{UMTS}$ can be calculated using equation (3) as follows;

$$EstimatedRTT_{UMTS} = EstimatedRTT_{UMTS} + \delta \times Difference_{UMTS}$$  \hspace{1cm} (3)

Here $\delta$ is a smoothing fraction between 0 and 1. From equation (2) and equation (3), the deviation for the UMTS network ($Deviation_{UMTS}$) can be calculated as follows;

$$Deviation_{UMTS} = Deviation_{UMTS} + \delta \times (Difference_{UMTS} - Deviation_{UMTS})$$  \hspace{1cm} (4)

Based on the estimated RTT of the UMTS network and the RTT deviation, TCP computes the timeout value as a function of $EstimatedRTT_{UMTS}$ and $Deviation_{UMTS}$, as given in equation (5),

$$TimeOut_{UMTS} = \mu \times EstimatedRTT_{UMTS} + \phi \times Deviation_{UMTS}$$  \hspace{1cm} (5)

The value of $\mu$ and $\phi$ is calculated based on experience. $\mu$ and $\phi$ depends on the link quality and the link speed difference between the WLAN and the UMTS network. For the current analysis, $\mu$ is set to 1, and $\phi$ is set to 5.
7. By cross layer interaction between the handover layer and the TCP, \( TimeOut_{UMTS} \), is updated in the TCP. Current retransmit timer, that was set to \( TimeOut_{WLAN} \), is also updated according to \( TimeOut_{UMTS} \) by equation (6) and equation (7).

\[
diffT = TimeOut_{WLAN} - CurrRetransmitTimer \tag{6}
\]

\[
CurrRetransmitTimer = TimeOut_{UMTS} - diffT \tag{7}
\]

8. After updation of the above mentioned values, handover layer becomes passive and the normal TCP processing begins.

6.3 Estimation of congestion window during handover

During handover from the WLAN to the UMTS network, CWnd in the WLAN network remains high compared to the UMTS network. Just after the handover, this difference in CWnd may trigger unnecessary congestion control actions in the UMTS network. As TCP reduces CWnd to its minimum value, and then increases it linearly, it takes some time to TCP to adjust its CWnd to appropriate value. Similarly in the case of handover from the UMTS to the WLAN network, value of the CWnd in the UMTS network is less than value of CWnd in the WLAN network and TCP increases CWnd linearly. Thus an estimation of appropriate CWnd for the new network just after the handover would increase the performance of TCP. The appropriate sending rate \( T \) of TCP is adjusted after handover using equation (8) as follows,

\[
T = \frac{s}{R \sqrt{\frac{3p}{8}} + t_{RTO}(3\sqrt{\frac{3p}{8}})p(1 + 32p^2)} \tag{8}
\]

In equation (8), \( T \) is in bytes/sec, \( s \) is the packet size, \( R \) is the RTT in new network, \( p \) is the steady state loss event rate, \( t_{RTO} \) is the TCP RTO value in the new network. Equation (9) is used to calculate the CWnd after the handover, as follows

\[
CongestionWindow = T \times RTT \tag{9}
\]

To estimate the sending rate \( T \), as given in equation (8), \( R \) can be calculated by equation (3), and \( t_{RTO} \) can also be calculated by equation (5). \( p \) is the average loss event probability as defined in [29]. \( p \) can be measured continuously by the handover layer at the sender side, and equation (8) and equation (9) can be used to estimate the congestion window, and is updated by the handover layer into the TCP control block.

7 Simulation and Analysis

The proposed ISN-TCP-PLUS has been implemented and simulated using Qualnet-5.0.1 [34] network simulation framework. The simulation scenario is taken
similar to Figure 1, and the network parameters are taken as described earlier in section 4. The proposed scheme is compared with ISN-TCP and WP-TCP. It can be noted that ISN-TCP already incorporates freezing concept for the WLAN to the UMTS handover, similar to freeze TCP [8, 30] and rate adaptation techniques for the UMTS to the WLAN handover [16, 17, 22]. So the proposed scheme is not compared explicitly with these techniques, and compared with ISN-TCP and WP-TCP.

7.1 Packets Sent and CWnd Variation During Handover

This subsection provides the performance of ISN-TCP-PLUS with respect to data packets sent and CWnd variations during handover. The simulation results for both the UMTS to the WLAN handover and vice-versa are explained next.

7.1.1 UMTS to the WLAN Handover

The simulation results for the UMTS to the WLAN handover is shown in Figure 10 and Figure 11. The handover is started at 474.5 seconds and is completed at 474.8 seconds. From Figure 10, both higher sequence number packets from the WLAN interface and lower sequence number packets from the UMTS interface are forwarded. This is because the UMTS interface is
not closed immediately after handover. However, as DACKs are filtered out based on handover notifications, TCP does not invoke congestion control actions. In this way unnecessary retransmissions and CWnd variations are avoided, as shown in Figure 11. Effectively, even after handover, data rate is guided by the UMTS network for some duration. It has been observed that amount of packet retransmissions is considerably lower compared to WP-TCP and ISN-TCP.

### 7.1.2 WLAN to the UMTS Handover

The simulation results for the WLAN to the UMTS handover is shown in Figure 12 and Figure 13. In ISN-TCP-PLUS, RTT value calculated using the procedure mentioned in section 6, is updated in TCP control block. This prevents the spurious RTOs from occurring and improves the performance. As shown in the Figure 12, data receiving by mobile node stops at 1150.67 secs and again continues from 1153.80 secs. This break in transmission is due to underlying handover “break before make” scheme. It can be seen from Figure 13, there is no CWnd drops after handover.

### 7.2 Amount of Retransmission for the UMTS to the WLAN Handover

The problem for the UMTS to the WLAN handover is amount of unnecessary retransmissions and expiration of DACK threshold. The improvement of ISN-TCP-PLUS with respect to amount of retransmissions is shown in Figure 14. Amount of retransmission is considerably less compared to WP-TCP and ISN-TCP. In case of ISN-TCP-PLUS, the only packets which are lost during handover is retransmitted through the WLAN interface. Amount of retransmission is very high for ISN-TCP because ISN-TCP explicitly retransmits all previously transmitted packets through the UMTS interface during handover, and the UMTS interface is immediately closed after handover.
7.3 **Spurious RTOs during the WLAN to the UMTS Handover**

The problem for the WLAN to the UMTS handover is the spurious RTOs. Figure 15 presents the percentage of two consecutive and three consecutive spurious RTOs as a function of the UMTS network one way channel delay. Maximum CWnd is taken as 10 packets which is equal to bandwidth delay product (BDP). It can be observed that the lowest delay that can result two or three consecutive spurious RTOs is considerably higher for the ISN-TCP-PLUS compared to the WP-TCP and the ISN-TCP. The spurious RTOs occur when the RTO is less than the maximum time between the arrival of the last ACK from the WLAN network and the first ACK from the UMTS network. The RTO estimation procedure as proposed for ISN-TCP-PLUS estimates the new RTT and RTO for the UMTS network immediately after the handover, and TCP parameters are updated accordingly. Thus ISN-TCP-PLUS avoids spurious timeouts. It can be seen from the figure that two consecutive spurious RTOs occur for ISN-TCP-PLUS when the UMTS one way delay is more than 1.3 seconds and three consecutive spurious RTOs occur when the one-way UMTS delay is more than 1.5 seconds. The estimation procedure fails for high UMTS delay because while estimating initial $\text{Deviation}_{\text{UMTS}}$ is taken as zero, whereas actual $\text{Deviation}_{\text{UMTS}}$ is very high. Therefore, the convergence requires considerable amount of rounds, before which the RTOs occur. However, this much of high one way delay for the UMTS network never occurs in practice.

7.4 **TCP Efficiency**

TCP efficiency is defined as the ratio of the number of packets transmitted successfully to the actual number of packets transmitted. In Figure 16, the efficiency for TCP flow is shown with respect to mean handover interval. Mean handover interval is defined as the average time between two consecutive handover. It can be seen from the figure that the efficiency of ISN-TCP-PLUS is significantly higher compared to the ISN-TCP and the WP-TCP, when mean handover interval is very high. The figure shows that the proposed ISN-TCP-PLUS is very adaptive with frequent handover.
This paper provides an improved TCP variant for the vertical handover between the UMTS and the WLAN networks using an ISN based framework. The shortcomings of the WP-TCP and the ISN-TCP are analyzed using the simulation results. It has been observed that the main problem during the UMTS to the WLAN handover is the unnecessary retransmissions, and during the WLAN to the UMTS handover the spurious RTOs. The proposed TCP variant, namely the ISN-TCP-PLUS, uses an estimation mechanism for measurement of the new RTT and the RTOs after the handover, and avoids spurious RTOs during the WLAN to the UMTS handover. The unnecessary retransmissions during the UMTS to the WLAN handover is avoided through the filtering of unnecessary DACKs. The sending rate for the TCP is estimated after the handover is over, and the CWnd is updated accordingly so that TCP can rapidly adopt to the new link characteristics after the handover gets completed. Simulation results confirm that the proposed TCP variant performs considerably better than other schemes, like the WP-TCP and the ISN-TCP.

References


