An Enhancement of mSCTP Handover with an Adaptive Primary Path Switching Scheme

Jinsuk Baek, Paul S. Fisher
Department of Computer Science
Winston-Salem State University
Winston-Salem, NC, 27110, United States
{baekj, fisherp}@wssu.edu

Abstract—We propose a primary path switching scheme to provide a seamless handover for dual-homed mobile terminals. This scheme is suggested as an enhancement to the mSCTP protocol. With our scheme, a mobile terminal performs primary path switching before it becomes unavailable due to its primary path drop. The improvement of the scheme comes from considering the temporal velocity of the mobile terminal with relative RTT variances of all available paths when it performs a handover process in the overlapped area between two different networks. Our simulation results show the proposed scheme provides a better overall performance than other schemes and the anticipatory switching is more important for faster moving terminals.

Keywords—mSCTP; Primary path switching; RTT-based; Speed-based;

I. INTRODUCTION

As a variety of networks (e.g. LAN, WLAN, and 3GPP) are currently deployed, and the next generation network is converging to an all-IP-based, unified network, many mobile terminals are configuring multi-homed environments by simply installing two or more network interfaces. In this environment, the mobile stream control transmission protocol (mSCTP) [1] has been considered as a proper transport layer protocol. The mSCTP provides connection-oriented reliable transmission over the IP core network via selective ACK (SACK), flow control, congestion control and avoidance, as well as failure detection and recovery.

Among its many functionality, its multi-homing feature can maximize the utilization of the multi-homed environment and increase network availability. The multi-homing feature enables a mobile terminal (MT) to use more than one IP address in order to support more than one communication path, namely a primary path together with several alternative paths in a single SCTP session. The primary path is used to transport the data packets, and the MT will change its primary path to an alternative path when its current primary path has failed. A link failure can easily occur, especially when an MT performs a vertical handover between heterogeneous, access networks.

In order to provide a seamless handover, the primary path switching scheme should be judiciously designed, otherwise it can have a significant, negative effect on the performance of the mSCTP. According to the specification of the mSCTP [1], primary path switching is not conducted until the current primary path has explicitly failed. However, one of the most challenging issues is how to determine when an MT has to change its primary path to the alternative path while it is under the dual-homing mode. In this condition, the MT is in an overlapping state between different networks and both paths are available. This issue has led to numerous proposals [2]–[7] to determine the appropriate network conditions for primary path switching. Unfortunately, the existing schemes have not sufficiently considered the switching conditions.

We propose a new primary path switching scheme, which considers the network conditions including round trip times (RTTs) of all available paths and velocity of the MT. In other words, an MT takes into account the historical and relative difference among the evaluated round trip times (RTTs) of the multiple paths against its estimated movement speed when it is located in the overlapped area between two different networks. Based upon these measurable characteristics, an MT adaptively determines when it needs to change its primary path. This case-adaptive feature of the proposed scheme guarantees network availability for the MT and provides advantages over the existing schemes: 1) it improves the performance in terms of transmission delay by reducing the probability of data packet loss at the MT; 2), it increases the overall throughput of the MT since the reduced number of retransmission requests from the MT allows the correspondent terminal (CT) to maintain its current congestion window size.

The rest of this paper is organized as follows. Section 2 describes related work. In Section 3, we propose a new primary path switching scheme taking into account network conditions and movement speed of the MT. Section 4 includes performance evaluations and discussion of the effect of the proposed scheme. Finally, we conclude this paper in Section 5.

II. RELATED WORK

This section describes the existing primary path switching schemes that have been proposed as an extension to the legacy mSCTP. These schemes essentially differ in the strategies they use for deciding which conditions should be satisfied for performing the path switching. The RTT-based adaptive scheme [2] considered both the absolute RTT gap and the relative RTT gap for both paths. Even though the current RTT of the alternative path is smaller than that of current primary path, the MT refrain from switching to the alternative path as long as the relative gap ratio is less than a fixed constant time. This conservative approach allowed the MT to maintain its current congestion window size. Therefore, the scheme could
make a promise of consistent throughput to the MT. Unfortunately, the scheme considered only the current RTTs for both paths at a given moment without consideration of historical RTTs. Also, the time window for evaluating the RTTs was not defined. As the current RTT cannot well reflect the comprehensive network condition for two paths, more consideration is needed for a more accurate determination.

Other proposals [3]–[4] considered bandwidth for both paths as a decision metric. Based upon the narrowest end-to-end bandwidth and required minimum bandwidth by application, the scheme classified four different cases and suggested different switching conditions in terms of bandwidth gaps for two paths. However, the scheme did not consider the end-to-end delay for both paths, which significantly affects the mSCTP retransmission strategy.

The RTT has been considered in [6] for some cases. Under this scheme, the link bandwidth has been considered as the most important decision metric for primary path switching while a RTT has been considered later. That is, link bandwidths of both paths have been measured and relative RTT is compared whether or not to switch the primary path. However, all mentioned bandwidth based schemes are essentially depending on the available minimum bandwidth of the local paths, which will not reflect on the actual mSCTP throughput for the end-to-end path.

III. PROPOSED SCHEME

In order to determine primary path switching criteria for each MT, we take into account two metrics including the historical relative difference of the RTT against the estimated movement speed of the MT. Both metrics are calculated only when the MT is under the dual-homing mode as it is located in the overlapping state among different networks as shown in Figure 1. In order to simplify our description, we assume each MT has only one alternative path, since the proposed scheme can be applied in a straightforward manner to the multi-alternative paths environment.

\begin{equation}
D(i) = RTT_{<i} - RTT_{<i+1}, \quad (1 \leq i \leq n)
\end{equation}

where $RTT_{<i}$ and $RTT_{<i+1}$ is RTT for the primary path and the alternative path at the $i$-th measurement step, respectively.

Note that the $D(i)$ is considered only when the $RTT_{<i}$ is larger than $RTT_{<i+1}$, that is when $D(i)$ is positive. In addition, a weighted average of the measured $D(i)$ values can be expressed as $D_i$, such that

\begin{equation}
D_i = \sum_{j=1}^{i} a_{ij} \cdot D(j),
\end{equation}

where $a_{ij}(j)$ is the weight for the $j$-th measurement step.

As we have more interest in the recent deviation of the difference, our weighted average puts more weight on $D(i)$ than on $D(i-1)$. Here, we define $D_T$ as configured threshold time ($D_T > 0$) that is used to determine the relative network condition of the MTs and $\alpha$ as network condition coefficient ($\alpha \geq 1$). The $D_T$ is initially set to $\alpha \cdot D_f$. At a given measurement step $i$, based upon the size of the value $D$ from Equation (2), each MT can be broadly divided into three different categories including:

1. An MT with a large RTT gap such that $D \geq \alpha \cdot D_f$;
2. MT with small RTT gap such that $D_f \leq D < \alpha \cdot D_f$; and
3. MT with extra-small RTT gap such that $D < D_f$.

Typically, the fast MTs show a large RTT gap while the slow MTs show a small RTT gap. Also, the slow MTs with a ping pong movement show an extra-small RTT gap. The proposed scheme requires the MTs with the large RTT gap to immediately switch its primary path to the alternative path, while the MTs with an extra small RTT gap stay on the primary path until the associated network condition incurs a significant change. On the other hand, the MTs with a small RTT gap switches its primary path only when $RTT_{<i} \geq \beta \cdot RTT_{<i+1}$, where $\beta$ is a fixed constant ($\beta \geq 1$), and it is called the switching coefficient.

The two coefficients, $\alpha$ and $\beta$, are devised for two purposes. First, the estimated RTTs are not favorable to the current primary path because these RTTs have been estimated based upon different data volumes. As the transmission delay is directly affected by the data size, the RTTs for the current primary path can be relatively overestimated compared to the RTTs for the alternative path which may have been estimated based upon small data size. Second, they provide an acceptable compromise between the RTTs and congestion window (cwnd) size. If we simply allow an MT to switch its packets. It can also check the connectivity of the alternative path with interaction of periodical HEARTBEAT and HEARTBEAT-ACK packets. These packet exchanges allow the MTs to measure their RTTs for both paths in a timely manner.

In order to determine the primary path switching criteria, the proposed scheme first considers the difference between the RTT for both paths. This can be expressed as Equation (1). Note that MTs start to measure the RTTs for both paths when the alternative path is activated at time $T_E$. On the other hand, they terminate measuring when one of the paths is deactivated at time $T_F$. Once we assume the MT has measured the RTTs $n$ times during $(T_E - T_F)$ time period, the size of measurement interval $\Delta_T$ is equal to $(T_E - T_F)/n$ and the absolute difference of the RTT between two paths at the $i$-th measurement step can be expressed as $D(i)$ such that

\begin{equation}
D(i) = RTT_{<i} - RTT_{<i+1}, \quad (1 \leq i \leq n)
\end{equation}

where $RTT_{<i}$ and $RTT_{<i+1}$ is RTT for the primary path and the alternative path at the $i$-th measurement step, respectively.

Note that the $D(i)$ is considered only when the $RTT_{<i}$ is larger than $RTT_{<i+1}$, that is when $D(i)$ is positive. In addition, a weighted average of the measured $D(i)$ values can be expressed as $D_i$, such that

\begin{equation}
D_i = \sum_{j=1}^{i} a_{ij} \cdot D(j),
\end{equation}

where $a_{ij}(j)$ is the weight for the $j$-th measurement step.

As we have more interest in the recent deviation of the difference, our weighted average puts more weight on $D(i)$ than on $D(i-1)$. Here, we define $D_T$ as configured threshold time ($D_T > 0$) that is used to determine the relative network condition of the MTs and $\alpha$ as network condition coefficient ($\alpha \geq 1$). The $D_T$ is initially set to $\alpha \cdot D_f$. At a given measurement step $i$, based upon the size of the value $D$ from Equation (2), each MT can be broadly divided into three different categories including:

1. An MT with a large RTT gap such that $D \geq \alpha \cdot D_f$;
2. MT with small RTT gap such that $D_f \leq D < \alpha \cdot D_f$; and
3. MT with extra-small RTT gap such that $D < D_f$.

Typically, the fast MTs show a large RTT gap while the slow MTs show a small RTT gap. Also, the slow MTs with a ping pong movement show an extra-small RTT gap. The proposed scheme requires the MTs with the large RTT gap to immediately switch its primary path to the alternative path, while the MTs with an extra small RTT gap stay on the primary path until the associated network condition incurs a significant change. On the other hand, the MTs with a small RTT gap switches its primary path only when $RTT_{<i} \geq \beta \cdot RTT_{<i+1}$, where $\beta$ is a fixed constant ($\beta \geq 1$), and it is called the switching coefficient.

The two coefficients, $\alpha$ and $\beta$, are devised for two purposes. First, the estimated RTTs are not favorable to the current primary path because these RTTs have been estimated based upon different data volumes. As the transmission delay is directly affected by the data size, the RTTs for the current primary path can be relatively overestimated compared to the RTTs for the alternative path which may have been estimated based upon small data size. Second, they provide an acceptable compromise between the RTTs and congestion window (cwnd) size. If we simply allow an MT to switch its
primary path to the alternative path when the alternative path shows a shorter RTT at a current measurement step, the MT will experience performance degradation because this switching requires significant overhead, since the switched MT is required to re-start its congestion control at the slow start phase.

### B. Speed-Based Scheme

The above proposed RTT-based scheme shows partial success in terms of throughput. However, the sole RTT measure alone does not effectively determine the primary path, switching criteria because it does not adequately reflect the velocity of the MTs. Even though an MT has a much shorter RTT in its alternative path than that of its current primary path, and our weighted average value $D$ is larger than $\alpha T$, we cannot obviously conclude that the velocity of the MT is fast because the RTT is likely to fluctuate from chunk to chunk, depending upon the level of congestion in the routers, and upon the varying load on the MT’s. One example of this would be the fast moving MT experiences an even larger RTT for the alternative path that will be used in its new network when the new access router (AR) is suffering from congestion. According to our Linux test-bed experiments, we found the velocity of MT’s should be considered as an important factor in determining the primary path switching especially for the fast moving MT’s. The usual primary path switching procedure performs the following sequence of actions:

1. The MT detects a link-up of a new path and configures a new IP address at the network layer.
2. The MT sends “Add-IP” message to the correspondent terminal (CT).
3. The MT sends “Primary-Switching” message to the CT if some conditions have been satisfied.
4. The MT sends “Delete-IP” message to the CT.
5. The MT detects link-down of its previous primary path.

Note that the alternative path is used for a CT-to-MT data packet after event 3, while the primary path is still used for MT-to-CT data packet before processing event 4. That is, for the receiving data packets, the MT immediately uses its alternative path as a primary path as soon as it explicitly requests primary path switching. On the other hand, for sending data packets, the MT does not use the alternative path until it explicitly deletes its current primary path.

However, fast moving MT’s can experience a link-down event in its primary path even before sending a Primary-Switching message to the CT. That is, the events 1, 2, and 5 can sequentially occur with extremely short intervals. Therefore, the two SCTP hosts cannot communicate until the MT sends the Primary-Switching message and the CT updates its data forwarding table in its kernel. Of course, both CT-to-MT and MT-to-CT data packets are lost during this period. To eliminate this critical period, the fast moving MT’s have to compress three messages (“Add-IP”, “Primary Change”, and “Delete-IP”) into a single message and send it to the CT. Therefore, this forces us to define a new type of message. Therefore, the proposed scheme uses a different approach to minimize this period by requiring them to sequentially send Add-IP, Primary Change, and Delete-IP messages with the system allowed, shortest intervals.

<table>
<thead>
<tr>
<th>Join SCTP session</th>
<th>Begin loop</th>
<th>Switch(event)</th>
</tr>
</thead>
<tbody>
<tr>
<td>event: $(E(S) \geq \gamma S_T)$</td>
<td>If $[(D \geq \alpha D_T) \land (RTT_{CP} \geq \gamma RTT_{CP}')]$ Sends “Primary-Switching”</td>
<td>Else Stay on the current primary path</td>
</tr>
<tr>
<td>event: $(S_T \leq E(S) \leq \gamma S_T)$</td>
<td>If $[(D \geq \alpha D_T) \land (RTT_{CP} \geq \beta RTT_{CP}')]$ Sends “Primary-Switching”</td>
<td>Else Stay on the current primary path</td>
</tr>
<tr>
<td>event: $(E(S) &lt; S_T)$</td>
<td>If $[(D \geq \alpha D_T) \land (RTT_{CP} \geq \beta RTT_{CP}')]$ Sends “Primary-Switching”</td>
<td>Else Stay on the current primary path</td>
</tr>
</tbody>
</table>

**Figure 2. Primary path switching algorithm**

We estimate the moving speed of the MT’s based upon the elapsed moving time and the moving distance from the previous network to the current network. We define time $T(k)$ as a time when the MT sends the ADD-IP message to the CT in its previous network $k$. As a MT records these times whenever it moves to the different networks, it keeps a series of $T(k)$ such that $1 \leq k \leq K-1$ by assuming it has visited $K$ networks. Therefore, the elapsed moving time $\Delta T(k)$ of the MT between two successive networks $k-1$ and $k$ can be obtained by

$$\Delta T(k) = T(k) - T(k-1)$$

The estimated speed of the MT at a given network domain can now be obtained by dividing the moving distance that the MT has traversed in the current network by the elapsed moving time $\Delta T$. Also, the $\Delta T$ tends to be frequently changed from network to network, the proposed scheme only considers a limited number of recent historical speeds rather than a full history. Therefore, the expected speed $E(S)_{h+1}$ of the MT in its next network $k+1$ considers the recent $h$ histories, and this will obey the following equation.

$$E(S)_{h+1} = \sum_{j=1}^{k} \omega_k(j) \cdot \frac{d_j}{\Delta T(j)}$$

where $d_j$ is the average moving distance of MTs in network $j$ and $\omega_k(j)$ is weight value at network $j$. As the MT tends to keep its current speed, the proposed scheme maximizes the size of $j$ and puts much more weight, say 0.875, on the most recent speed of the MT. In order to differentiate among MT’s in terms of velocity, we also define $S_T$ as a preconfigured threshold time ($S_T > 0$) that is used to determine the relative speed of the MT’s and $\gamma$ as a fast speed coefficient ($\gamma \geq 1$). Each MT can now be classified into three different categories including 1) fast moving MT’s with large $S$ value such that $E(S) \geq \gamma S_T$; 2) slow moving MT’s with a small $E(S)$ value such that $S_T \leq E(S) < \gamma S_T$; and 3) slow moving MT’s having a ping pong movement pattern with extra-small $E(S)$ values such

---

$\sum$
that $E(S) < S_T$. Based upon these observations, each MT determines whether it performs primary path switching or not. The more detailed, overall actions are described in Figure 2.

If the mSCTP is applied to the single-homed MT, which has only one IP address at one time, the three events, including 1) detecting link-up of its new path; 2) sending an Add-IP message to the CT; 3) detecting link-down of its old path, can sequentially occur with non-negligible intervals because it relies on the time that a single NIC is activated and deactivated. This time introduces some delay. Moreover, unlike the dual-homed MT situation, the CT and the MT can now only communicate through the new IP address. Therefore, the CT will not transmit any more data packets until it receives a Primary-Switching message from the MT, which introduces another delay. In order to reduce this delay, it is better to send the Primary-Switching message as soon as the MT detects link-down over its old path. In case of extremely fast moving MT’s, link down of the current path can be detected even before the MT sends an Add-IP message to the CT. Such MT’s are temporarily unreachable from the CT. In this case, some mobility prediction support will be needed with assistance from the network layer protocols. However, this prediction algorithm is beyond the scope of this paper, although we have developed such an algorithm.

IV. PERFORMANCE EVALUATION

In order to clarify the performance improvement of the proposed scheme, we consider a small test topology as shown in Figure 3.

We consider the mSCTP MT has two different NICs, which use IEEE 802.11b as a wireless link while the mSCTP CT has one NIC, which uses Ethernet as a wired link. We also consider these two mSCTP hosts are interconnected through a 100Mbps IP core network, each having an average of 10 network hops.

Under this topology, we have developed a system to emulate the variations of RTTs between the two hosts. With our emulator, the effective bandwidth for the two paths has been changed between 0.5Mbps and 1.5Mbps and two propagation delays for the MT at a given time also have been proportionally changed based upon the MT’s physical distances to its ARs. According to the statistical random multiplexing of the packet switched routers, the queuing delay has been generated as a Poisson random number having an average 50ms over time. Finally, we do not consider the processing delay because this is relatively small and constant compared to other delays.

Even though we set the data packet size to five times larger than the control packet size, we set the network condition coefficient $\alpha$ to 3 because we consider the segmentation and reassembly overhead when the data size is larger than the path maximum transmission unit (PMTU). For simplicity, we consider the MT’s maintaining their current moving speed under dual-homing status. Their moving directions are also considered to be steady except for the MT’s having a ping pong movement pattern. We define slow moving MT’s to be when their moving speed is slower than 10km/h. As we set the speed coefficient $\gamma$ to 4, the MT’s, which are faster than 40km/h in speed, are considered as fast moving MT’s. The experiment parameters are summarized in Table I. For MT’s with different moving speeds, we require the MT to estimate RTTs of both paths every 1 second.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum network condition coefficient: $\alpha$</td>
<td>3</td>
</tr>
<tr>
<td>Minimum switching coefficient: $\beta$</td>
<td>2</td>
</tr>
<tr>
<td>Minimum speed coefficient: $\gamma$</td>
<td>4</td>
</tr>
<tr>
<td>Minimum frequency coefficient: $\delta$</td>
<td>5</td>
</tr>
<tr>
<td>Historical RTT threshold: $D_T$ (ms)</td>
<td>140</td>
</tr>
<tr>
<td>Historical Speed threshold: $S_T$ (km/h)</td>
<td>10</td>
</tr>
<tr>
<td>Data packet size (Kbyte)</td>
<td>1</td>
</tr>
<tr>
<td>Control packet size (byte)</td>
<td>200</td>
</tr>
<tr>
<td>RTT measurement interval: $\Delta_T$ (sec)</td>
<td>1</td>
</tr>
<tr>
<td>Standard moving distance: $SD$ (meter)</td>
<td>1000</td>
</tr>
<tr>
<td>Standard moving distance in overlapping area: $DOA$ (meter)</td>
<td>400</td>
</tr>
<tr>
<td>Wireless link speed: $LS$ (m/sec)</td>
<td>$10^4$</td>
</tr>
</tbody>
</table>

Figure 3. Network model

Figure 4. RTTs variations of MT with 10km/h moving speed

Figure 5. RTTs variations of MT with 80km/h moving speed
Due to the difficulty in obtaining the actual moving distance, we use an average moving distance of MT’s that has been pre-calculated in each network. Even though this distance is different from the actual distance, it still works well for the proposed scheme because our interest is the relative moving speed among the MT’s. Figure 4 and 5 shows the evaluated RTTs for the MT’s with a 10km/h and 80km/h moving speed, respectively. At first, we can observe the slow speed MT and fast speed MT spent about 145 second and 30 second in the overlapped area respectively. We now compare the proposed scheme to the RTT-only scheme [2], which has most recently been proposed but considered only RTT difference at a given time. Under the scheme, the MT switches its primary path as long as RTT < \( iR \) condition is satisfied. We do not show the performance of the MT’s having irregular movement patterns. Due to the conservative approach, both schemes will not allow them to switch their primary path as long as their movement patterns are not significantly changed.

For the slow moving MTs, as we set the minimum switching coefficient \( \beta \) to 2, the RTT-based scheme will allow the MT to switch its primary path at 100 seconds because 288 ≥ 2·102. On the other hand, the proposed scheme delays this path switching to 125 seconds when the historical RTT condition is satisfied for recent \( \delta = 5 \) consecutive measurement steps. This delayed switching affects the throughput since it allows both CT and MT to continuously increase their cwnd rather than restarting at the slow-start phase. In our experiment, we put the different weight \( w_R(j) \) at the measurement step \( j \) by obeying the following rule where the weight of most recent RTT \( w_R(i) \) is equal to 0.875. In the case of the fast moving MT, our experiment shows more meaningful results. Note that the dashed line shows the expected RTT of the new primary path when the MT switches its primary path at 1 second with the proposed scheme since RTT < \( P \) ≥ RTT < \( P \). After 19 seconds, the current primary path is not available as the MT is out of range of the AR of the previous network. In Figure 5, under the RTT-only scheme, note that the RTT < \( P \) ≥ \( \delta \)RTT < \( P \) condition is not satisfied until the MT detects link-down of its current primary path. Therefore, the MT will not change its primary path until its new path is activated and its new IP address is updated to the CT. The legacy-SCTP also suffers from the same problem. Until then the MT is in an unreachable status causing performance degradation.

Figure 6 shows the throughput of the proposed scheme compared to the other two schemes when the fast moving MTs are on the overlapped area. We assume the CT sends 128 data packets per second and set the wireless link error probability to 0.01. We set the new NIC activation delay to 2 seconds. Owing to our fast primary path switching, the proposed scheme will not suffer from this delay. Also, the average RTT of the proposed scheme in overlapped area is shorter than the other scheme. As a result, it shows better throughput than the RTT-only scheme. The proposed scheme considers the moving speed of the MT in order to cope with the trade-off relationship between the RTT and the cwnd. We give different conditions in terms of the RTT against the MT’s different levels of moving speed. These conditions more severely impact the MT’s with a slow moving speed as shown in Figure 2. The proposed scheme indicates the RTT benefit when the MT actually switches its primary path. If this expected benefit is not large enough, it requires the MT to stay on the current primary path. This benefit can be compensated for by preventing its congestion control from entering the slow-start phase. On the other hand, the cwnd benefit is relatively small for the fast moving MT’s because they can easily experience packet losses due to the delayed switching. Therefore, more generous RTT conditions are given to those MTs. In addition, the proposed scheme allows non-fast moving MT’s to continuously increase their cwnd in their current primary paths because the immediate primary path switching of the fast moving MTs can alleviate the congestion of their current primary path.

V. CONCLUSION

We have proposed a new scheme for switching the primary path and an alternative path within the SCTP protocol for MTs that are moving at various speeds between networks. This scheme provides a better overall performance than other existing schemes as shown by the simulation of performance. This is due to the fact that the proposed scheme utilizes switches between the present primary path and the alternative path according to the RTT and the velocity of movement of the MT.

ACKNOWLEDGEMENT

This research was supported by Ministry of Culture, Sports and Tourism and Korea Culture Content Agency in the Culture Technology Research & Development Program, the Korean government and was supported by the Brain Korea 21 program, Ministry of Education, Science and Technology, the Korean government.

REFERENCES