PAPER Special Section on Networking Technologies for Mobile Internet Systems

Transport Layer Mobility Support Utilizing Link Signal Strength Information

Moonjeong CHANG[†], **Meejeong LEE**[†a)], *and* **Seokjoo KOH**[††], *Nonmembers*

SUMMARY Recently, mobile SCTP (mSCTP) has been proposed as a transport layer approach for supporting mobility. mSCTP is based on the 'multi-homing' feature of Stream Control Transmission Protocol (SCTP), and utilizes the functions to dynamically add or delete IP addresses of end points to or from the existing connection in order to support mobility. In this paper, we propose a transport layer mobility supporting scheme, which is based on mSCTP and utilizes the link layer signal strength information in order to determine when to add or delete end-point IP addresses of mobile node and how to change data delivery paths when handover happens. Exploiting the fact that the transport layer is aware of the mobility in the proposed scheme, we also propose error and congestion control enhancement to cope with handover efficiently. The simulation results show that the performance of proposed scheme is competitive compared to the traditional network layer mobility supporting approach. Especially, when the moving speed of mobile node is fast or new path acquisition takes long, it shows better performance than the traditional network layer approach.

key words: *transport layer, mobility, mSCTP, SCTP*

1. Introduction

For the next-generation Internet, one of the essential requirements of the users is being connected to the network while roaming. A number of mobility management techniques have been proposed in the literature, and these existing techniques can be classified into network layer techniques and transport layer techniques. Mobile IP is the proposed standard of IETF for supporting mobility based on IP [1]. Various protocols to enhance the performance of Mobile IP have also been proposed [1]–[6]. These protocols, including Mobile IP, take a common stance in the sense that they all manage mobility at the network layer. If mobility is managed at the network layer, transport connections may remain transparent to the user movement. Mobility management at the network layer requires special entities such as HA (Home Agent) and FA (Foreign Agent) to be deployed in the network, and this involves overhead and inefficiency such as tunneling and/or triangle routing as well as security related complexities [1]–[6].

Most of the existing transport layer techniques proposed for mobility cannot deal with mobility on their own since they depend on the network layer mobility management techniques for the IP address management required by handovers. The main purpose is, therefore, merely to minimize the degradation of transport performance caused by handovers. For this purpose some of the transport layer techniques require the mobility management related network entities to implement special mechanisms concealing and/or minimizing the losses occurring during handovers so that spurious timeout or rate reduction do not occur at the transport layer of sending side [7], [8]. For this type of transport layer mechanisms, handover latency may greatly increase due to the increment of amount of processing at the mobility management related network entities. The other types of transport layer techniques inform the sending transport layer of the handover so that the transmission stops temporarily and resumes after the handover is completed [9], [10]. For these mechanisms, resumed transmission on the new path may cause congestion since the transmission window adjusted to the old path is continuously applied to the new path. In addition, they do not provide any mechanism for the fast recovery of losses occurred during handovers. Among the existing transport layer techniques, Migrate TCP is the only one that may handle mobility on its own at the transport layer. However, it is known that Migrate TCP suffers from disconnection when handover latency becomes large, and its transport throughput is lower than pure TCP over Mobile IP [11], [12].

Some of the newly emerging transport protocols, such as Stream Control Transmission Protocol (SCTP) or Datagram Congestion Control Protocol (DCCP) [13], [14], suggest the possibility of independent management of mobility by the transport layer. Especially with the SCTP, an extension named mobile SCTP (mSCTP), which facilitates mobility has been drafted in [15]. mSCTP is targeted for the Client-Server services, in which the mobile client initiates an SCTP session with the fixed server. For supporting the peer-to-peer services, the mSCTP must be used along with an additional location management scheme such as Session Initiation Protocol (SIP) [16] or Dynamic DNS (DDNS) [17] in order to find out the current location of the peer to initiate the association. Once the association is established mSCTP takes care of mobility on its own without any intervention from the network.

SCTP is a new IETF standard track general purpose transport protocol for the Internet [13]. Similar to TCP, SCTP provides a connection oriented reliable service, and a connection between two SCTP endpoints is called as an *association*. One of the major features that SCTP provides is multi-homing. Multi-homing allows an endpoint of an SCTP association to be mapped to multiple IP-addresses.

Manuscript received January 16, 2004.
Manuscript revised May 27, 2004.

[†] The authors are with the Department of Computer Science, Ewha Womans University, 11-1, Seoul, Korea.

[††] The author is with the Department of Computer Science, Kyungpook National University, 702-701, Daegu, Korea.

a) E-mail: lmj@ewha.ac.kr

Among those addresses, one is chosen for the 'primary path' and is used as the destination for normal transmission, and the other addresses could be used for retransmissions only. Multi-homing feature of SCTP provides a basis for mobility support since it allows a mobile node (MN) to add a new IP address, while holding the old IP address already assigned to itself. On top of SCTP multi-homing feature, mSCTP proposed to utilize ADDIP, DELETEIP, and Set-Primary functions which enable dynamically adding or deleting IP addresses to and from the list of end points, and to replace the primary path for an on-going association [18].

The current specification of mSCTP [15] is, though, at a very primitive stage, and some essential issues, such as when and by which criteria the primary path to be changed or the addition and deletion of the IP addresses mapped to the SCTP association should occur in order to deal with handover seamlessly, are yet left for future elaboration. Without these issues being defined, though, the current mSCTP cannot practically handle mobility. For example, without appropriate mechanism to determine when to change and how to select the primary path, a serious oscillation problem, which may degrade the performance to the minimum, could occur during handover.

In this paper, we identify the loosely defined or missing aspects of the current mSCTP and propose a transport layer mobility supporting scheme which addresses all of those aspects. The proposed scheme addresses when to add or delete IP addresses of the mobile node and how to select and when to change the primary path for handovers. Furthermore, Error and congestion control mechanisms to reduce the handover latency and loss recovery time are also proposed. Unlike the existing mobility management techniques, the proposed scheme has the following features: 1) complete end-to-end mobility management is provided by the transport layer without requiring any support from the underlying network layer, 2) no handover related extra processing is imposed at the network entities, 3) no handover related congestion is incurred, and 4) fast error recovery mechanism is provided for the losses occurred during handovers.

The rest of this paper is organized in the following way. Section 2 gives a detailed explanation on the operation of proposed scheme. Simulations and its numerical results are presented in Sect. 3. Finally, Sect. 4 concludes the paper.

2. Transport Layer Mobility Support

In this section, we will describe the operation of the proposed scheme especially focusing on the following two aspects: 1) utilizing the layer 2 signal strength information for end-to-end IP address management to cope with the mobility, and 2) reducing the transport performance degradation caused by handovers. 1) is mainly an operation of MN, whereas 2) is implemented at the CN's side by an error and congestion control mechanism. The operation of MN is described in Sect. 2.1, and Sect. 2.2 explains the operation of CN.

2.1 Operation of MN

When MN moves into a new subnet, layer 2 (L2) handover and new IP address acquisition should happen. The proposed scheme assumes that L2 handover and the acquisition of a new IP address is achieved in the same way as they are done with Mobile IP. That is, for IPv4, it is assumed that the new IP address is obtained by DHCP (Dynamic Host Configuration Protocol) or CCoA (Co-located Care-Of Address) is deployed; for IPv6, the new IP address is assumed to be obtained by Stateless Address Auto Configuration [19].

When handover happens, mSCTP at MN should trigger ADDIP for the new IP address and DELETEIP for the old one. In the proposed scheme, mSCTP at MN triggers ADDIP as soon as the signal strength of the new access router exceeds the minimum signal strength that enables communications (hereinafter, it is called L2-TH). Once an IP address is added, DELETEIP for that address is not triggered until the signal strength from the corresponding access router becomes lower than the L2-TH. With these policies, an SCTP association of the proposed scheme maintains the MN's IP addresses corresponding to all of the accessible subnets. Furthermore, an accessible IP address is added to the SCTP association as early as possible. The main purpose of these principles regarding adding or deleting end point IP addresses is to maximize the chance that an end point IP address is ready when it is needed for handover.

Minimum signal strength that enables communication is the signal strength measured at the boundary of transmission range, and is determined by radio propagation model. For Two-Ray Ground Reflection model, which is the radio propagation model assumed in our simulation experiment, the minimum signal strength that enables communication (i.e., L2-TH) is computed by the following formula [20]:

$$\frac{P_t G_t G_r h_t^2 h_r^2}{d^4 L}, \tag{1}$$

where P_t, G_t, G_r, h_t, h_r, d and L denote transmit power, transmit antenna gains, receiver antenna gains, transmit antenna height, receive antenna height, diameter of transmission range, and system loss, respectively.

When handover happens, the primary path also needs to be changed. The current mSCTP does not specifically mention about how to change the primary path for handovers. In SCTP, sender is in charge of changing the primary path and it changes the primary path if the primary path experiences repetitive losses over a certain threshold. If it is adopted in mSCTP, therefore, CN should experience multiple data packet losses for each handover before it finally determines to change the primary path and it will lead to significantly long handover latency.

In order to prevent this, the proposed scheme makes MN, which is the receiver, be in charge of the primary path change, and trigger Set-Primary toward CN when handover happens. Set-Primary from MN notifies CN to change the primary path. In order to determine when to trigger Set-

Primary, MN uses L2 radio signal strength information. If the radio signal strength of the primary path becomes lower than a certain threshold (hereinafter it is called Primary-TH), primary path is replaced. The value of Primary-TH should be higher than L2-TH at the minimum in order for MN to trigger Set-Primary before DELETEIP of the primary path. Furthermore, it is desirable for Set-Primary to arrive at CN before MN completely moves out of the transmission range of the old access point. In order to satisfy this condition, the signal strength corresponding to Primary-TH should be at least the signal strength at the boundary of transmission range with diameter $(d - d')$, where d is the transmission range of the access router and d' is the distance that MN can move during the time for which Set-Primary is delivered from MN to CN. Therefore, based on the formula in (1), the minimum signal strength for Primary-TH that can satisfy the condition is obtained as follows:

$$\frac{1}{\left(\frac{d-d'}{d}\right)^4} \times L2 - TH \tag{2}$$

Note that the value determined by (2) depends on the moving speed of MN and the delay from MN to CN. Actually, the Primary-TH value computed by (2) is an optimal one since increasing Primary-TH value any further would only increase the chance of unnecessary primary path changes. The proposed scheme also let MN select a new primary path utilizing the L2 radio signal strength information of the wireless subnet, and inform it to CN. Among the accessible subnets, the one providing strongest radio signal is selected as the new primary path in order to minimize the possible oscillation.

2.1.1 Address Management Module (AMM)

The functions of utilizing L2 signal strength information to trigger mSCTP signals such as ADDIP, DELETEIP, and Set-Primary, are implemented in a logical block named AMM (Address Management Module). Figure 1 presents the interaction between AMM and mSCTP, IP address acquisition module, and link layer. Receiving signals from the link layer and the IP address acquisition module, AMM determines when to trigger ADDIP, DELETEIP, and Set-Primary and informs it to mSCTP. mSCTP at MN then sends out ADDIP, DELETEIP, or Set-Primary messages to its peer mSCTP at CN in order to change the end point mapping or the primary path for the SCTP association.

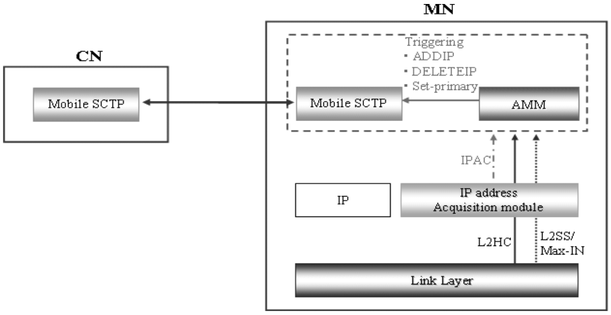


Fig. 1 Signaling between AMM and mSCTP, IP address acquisition module and link layer.

Link layer sends out following three types of signals to AMM in order to inform AMM about an L2 handover completion or changes of link signal strength:

1) L2HC (L2 Handover Completion): the L2 handover is completed for the interface specified in the signal.

2) Max-IN (Interface with Maximum signal strength): the interface providing maximum signal strength has been changed to the one specified in the signal.

3) L2SS (L2 Signal Strength): one of the L2 signal strength changes shown in Fig. 2 has occurred for a certain interface; L2SS specifies the interface for which the change has occurred and the ‘type of signal strength change.’ Note that the figures labeled on the arrows denote the ‘type of signal strength change.’

IP address acquisition module sends out IPAC (IP address Acquisition Completion) signal when an IP address acquisition for an interface is completed. The IPAC signal indicates the interface ID and the acquired IP address for that interface.

In order to store the information collected from the signals from the link layer and the IP address acquisition module, AMM maintains an Address Table as shown in Fig. 3. The SS (Signal Strength) field of the Address Table indicates the current signal strength of the interface, and the meaning of the value of this field is shown in Table 1. This field is updated by encoding the ‘type of signal strength change’ specified in L2SS signal as shown in Table 2. The H flag in the Address Table indicates whether the L2 handover is

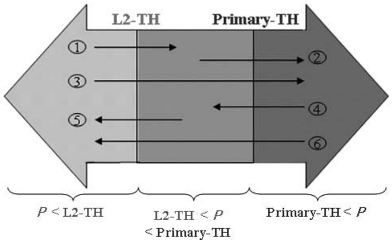


Fig. 2 Types of L2 signal strength changes (ρ: signal strength).

Interface ID	SS	H Flag (Connected(1) or not(0))	IP address
⋮	⋮	⋮	⋮

Fig. 3 Address Table in AMM.

Table 1 The values of the SS field in the Address Table.

SS	Signal Strength ρ
0	$\rho <$ L2-TH
1	L2-TH $< \rho <$ Primary-TH
2	Primary-TH $< \rho$

Table 2 Mapping between the value of the SS field and the 'type of signal strength change' in L2SS signal.

Type of Signal Strength Change in L2SS	SS field in Address Table
5, 6	0
1, 4	1
2, 3	2

completed for the corresponding interface. Receiving L2HC signal for a certain interface, H flag of corresponding entry in the Address Table can be set. The IP address field of the Address Table is filled when IPAC signal for the corresponding entry comes in from the IP address acquisition module. In addition to Address Table, AMM also maintains information such as the interface corresponding to the current primary path and the interface with maximum signal strength.

2.1.2 Operations of AMM

mSCTP at MN starts ADDIP for a certain IP address when both the L2 handover and the IP address acquisition of the corresponding interface are completed. That is, by receiving either an L2HC or an IPAC signal, if both the IP address field and the H flag are set for a certain entry of the Address Table, AMM triggers mSCTP to start ADDIP for the corresponding interface.

When AMM receives L2SS with the 'type of signal strength change' being equal to 4 or 6 for the current primary path interface, the primary path should be replaced. AMM first checks whether there is an alternative interface ready to be used as the new primary path. If one is found, it immediately triggers Set-Primary to mSCTP in order to replace the primary path with that alternative interface. In order for an interface to be a primary path interface, it should satisfy the following three conditions:

1) It is the interface with maximum signal strength and the signal strength is greater than the 'Primary-TH.' Note that even the interface with the maximum signal strength may not provide the signal strength higher than the Primary-TH.
2) Link layer handover for the interface is completed.
3) IP address acquisition for the interface is completed.

If there is no such interface, AMM postpones triggering Set-Primary until a path which satisfies all three of the above conditions appears. In order to avoid frequent changes of primary path during handover, the primary path is not replaced until a path which is stable enough is found even though the current one becomes inadequate. In the proposed scheme, a stable path is defined as the path that satisfies the above three conditions. The status of an interface may be transformed to being stable by incoming L2SS, L2HC, or IPAC signals. Having SS being equal to 0 or 1 for the current primary path indicates that the current primary path has become inadequate. Therefore, in this case, AMM triggers Set-Primary to mSCTP as soon as interface satisfying all three conditions of the primary path interface shows up.

If AMM receives an L2SS signal with the 'type of signal strength change' being equal to 5 or 6 for a certain interface, AMM triggers DELETEIP to mSCTP in order to start DELETEIP for that interface. If the interface happens to be the current primary path interface, AMM searches an alternative interface to serve as the primary path. If there is no interface ready to replace the primary path, DELETEIP triggering should be postponed. In this case, whenever Set-Primary is triggered afterwards, DELETEIP for the current primary path interface should be triggered together.

2.2 Operation of CN

In the proposed scheme, a CN could have multiple IP address bindings for an MN while the MN transiting the overlapping area of neighboring cells. In SCTP, when there are multiple IP address bindings for a destination end point, retransmissions could be done through alternative paths. In the proposed scheme, alternative paths are merely considered as the candidates of primary path and they are not used for retransmissions since the alternative paths exists temporarily, i.e., while an MN transiting a cell overlapping area. That is, retransmissions as well as the original transmissions are done through the primary path only in the proposed scheme.

If CN receives an ADDIP from MN, CN augments the end point IP address list for the corresponding association with the IP address specified in ADDIP message. When CN receives DELETEIP, it removes the IP address specified in DELETEIP from the end point IP address list of the corresponding association. When CN receives Set-Primary, the destination IP address of data transmission for the corresponding association is replaced with the one specified in Set-Primary message. In addition, in order to have fast recovery of losses that have occurred during handover, the proposed scheme requires CN to send out a probe packet, which is the first packet that has not been transmitted yet, through the new primary path, upon receiving Set-Primary from MN. The retransmission timer value for this probe packet is set to the default initial value since it is the start of transmission on a new path. Receiving the acknowledgement for the probe packet, CN can figure out all the losses occurred during handover. CN, then, start transmitting those lost packets followed by new packets in sequence under the slow start congestion control on the new primary path. It is worth to note that starting transmission on a new path with slow start is another aspect of taking advantage from the fact that the transport layer is aware of handover instances. Since the proposed scheme start transmission on a new path with slow start, not only it is friendly to the existing flows on the new path but also it can grab available bandwidth on the new path fast due to the exponential growth of window size.

Note that without the probe packet proposed in our scheme, the recovery of losses incurred by handover is delayed. When CN receives Set-Primary, if receiver window (rwnd) is not greater than the amount of outstanding packets(i.e., packets that are transmitted but not acknowledged yet), it cannot transmit new packet through the new primary path due to rwnd limitation. According to the original SCTP error control, CN has to wait until the retransmission timer for the last packet transmitted through the old primary path expires before it starts (re)transmission on the new primary path. On the other hand, if rwnd is greater than the amount of outstanding packets when CN receives Set-Primary, CN can transmit up to 2 new packets to the new primary path since the congestion window (cwnd) of the new path is 2. Note original SCTP always start the transmission on a certain path with slow start and the initial default value of cwnd is 2 packets. If losses have occurred during handover, CN would receive duplicate acknowledgements (dupACKs) for the packets transmitted on the new primary path. However, fast retransmit of the packets that are lost during handover cannot happen, because the maximum number of dupACKs that CN could receive is 3 (recall that dupACKs do not increment cwnd or decrement the amount of outstanding packets) whereas 4 dupACKs are required for fast retransmit to happen. Hence, the retransmission of the packets that are lost during handover can occur after the retransmission timer expires.

3. Simulation

In this section, we present the simulation model and compare the performance of proposed scheme and TCP over Mobile IP (hereinafter we will refer it as TOM) through the numerical results of the simulation. For the performance comparison purpose, TOM is specifically chosen since it is the representative data transmission framework based on network layer mobility support. The comparison of proposed scheme to the original mSCTP is not performed since plain mSCTP cannot cope with the mobility on its own due to the reasons explained in the introduction.

3.1 Simulation Model

The simulation was implemented using ns-2 simulator proposed by U.C. Berkeley. For the proposed scheme, the ns-2 SCTP node module implemented in [21] is patched. The simulation was run on RedHat Linux 7.3 with the v2.4.18 kernel.

For the simulation network model, we use a two-cell wireless network as shown in Fig. 4. MN randomly moves around the two cells according to the Random Waypoint mobility model [22]. Since the mobility across a general multicell wireless network consists of handovers between two neighboring cells, we believe simple two-cell wireless network model suffices for the simulation purpose. The wireless channel is assumed to be 802.11b WLAN with 2 Mbps capacity and negligible propagation delay. All of the wired links are assumed to have 10 Mbps link capacity with varying values of propagation delay. The coverage radius of each wireless cell is assumed to be 300 meters, and the distance between two neighboring cells is 520 meters. Therefore, the longest distance across the overlapping area between neighboring cells is 80 meters.

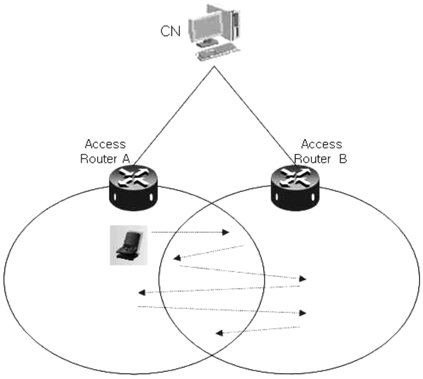


Fig. 4 Simulation network model.

As for the performance metric, the elapsed time for MN to download the 48 Mbytes of file from the CN is measured, and it is denoted as file transfer time. Handover latency, which is defined as the length of time interval between the instance receiving the last packet from the old path and the instance receiving the first packet from the new path when handover happens, is also measured. The path acquisition time, wired link delay from CN to access routers and moving speed of MN are varied as simulation parameters. The path acquisition time is defined as the time to complete both the L2 handover and the IP address acquisition for a wireless subnet. For L2 handover, both the soft and the hard handovers are experimented.

3.2 Simulation Results

Determining optimal Primary-TH is a problem since, as explained in Sect. 2.1, the optimal Primary-TH value depends on the moving speed of MN and the delay between MN and CN, and the link layer at MN does not have information on these. Reflecting the reality, if we assume the moving speed of MN from 2 m/sec to 28 m/sec (7 km/hour to 100 km/hour) and the delay between MN and CN being 5 msec to 100 msec, the optimal Primary-TH values would lie between 3.653e-10 W and 3.794e-10 W by the formula (2). We first checked whether repetitive ping pong movement within the cell overlapping area results in unnecessary primary path changes for any of the Primary-TH values within this range. Even though numerical results are not presented due to space limitation, we observed that for all the combination of moving speed and delay, none of the Primary-TH values in the given range incurred undesirable primary path changes for the ping pong movement of MN.

The value of Primary-TH is also varied when mobile node moves around the entire wireless cell area randomly in order to examine the impact of using Primary-TH that is lower than the minimum value determined by formula (2) to the performance of proposed scheme. Figure 5 shows the file transfer time for varying values of Primary-TH, and it is

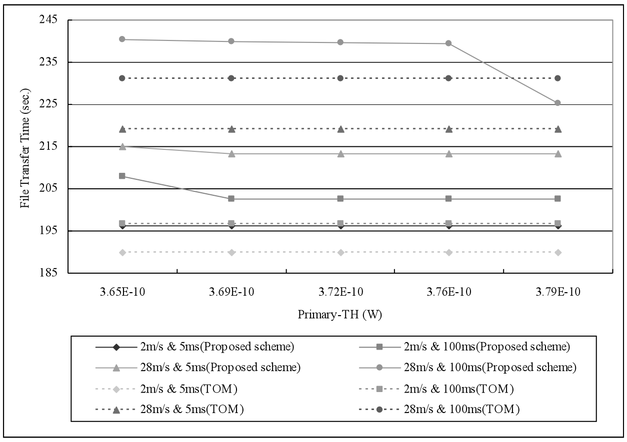


Fig. 5 File transfer time for varying values of primary-TH.

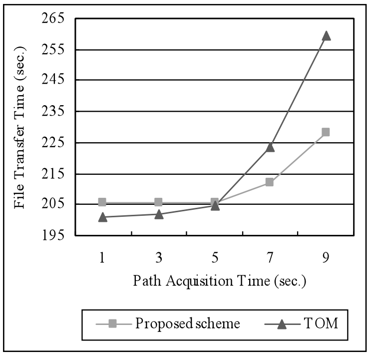


Fig. 6 File transfer time for different path acquisition time.

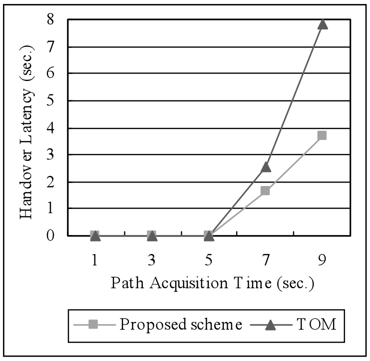


Fig. 7 Handover latency for different path acquisition time.

shown that the file transfer time grows when the Primary-TH is smaller than the minimum value. If Primary-TH is smaller than the minimum value, MN frequently moves out of the cell overlapping area before primary path is replaced and the amount of losses during handover becomes large resulting in longer recovery time after handover and ultimately a longer file transfer time. For the Primary-TH values greater than the minimum value, file transfer time is kept about the same. Based on these results, the Primary-TH value of the proposed scheme is set to 3.794e-10 W in the rest of the experiments.

Figures 6 and 7 show the file transfer time and the handover latency respectively for changing path acquisition time. For this experiment, soft handover is assumed for L2 handover, and the moving speed of mobile node is set to 15 m/sec. For path acquisition time less than or equal to 5 seconds, performance of TOM is better than the proposed scheme. In this case, both TOM and the proposed scheme can start transmitting data to the new path while MN is transiting the cell overlapping area, and the chance for MN to successfully receive all the data transmitted through the old path before it leaves the overlapping area is very high. That is, the impact of handover is minimal in this case. Under these circumstances, the performance of proposed scheme is slightly worse than TOM due to SCTP's higher header overhead as well as the impact of slow start used in the proposed scheme. Note TCP in TOM is not aware of the handover, and maintains the congestion window size of the previous path, which is, in most of the cases, larger than the initial window size of the slow start, when it starts transmitting on the new path. On the other hand, the proposed scheme always starts transmitting to the new path with the initial window size of the slow start.

One thing to note is that even though starting transmission on a new path with slow start seems to be adversary to the performance in this experiment, in which no external traffic exists, it may actually be beneficial in real world. For example, when the new cell has high traffic load, the incoming traffic flow with large window may suffer bust of packet losses. With slow start, injection of new traffic flow to a path begins cautiously with a small starting window size while allowing the incoming flow to grab the available bandwidth in the new path fast, i.e., with the exponential window size increment. The numerical results in Figs. 6 and 7 do not reflect those aspects of slow start since no external traffic is assumed in the experiment.

As the path acquisition time becomes longer, the time to start transmitting data through new path is delayed and as a result the amount of data, which are transmitted through the old path and not being able to be delivered to MN while it is transiting the overlapping area, increases. That is, amount of losses caused by handover increases. Moreover, changing the data delivery path may not even happen while MN is transiting the overlapping area if the path acquisition time becomes larger than the MN's overlapping area transiting time. The amount of losses caused by handover grows even larger in this case. Since TCP in TOM is not aware of handover, it reduces the transmission window if handover causes packet losses. Furthermore, if retransmission time out occurs due to the losses during handover, transmission through the new path may not start even after handover is completed since TCP waits for the current retransmission timer to expire. On the other hand, the proposed scheme starts transmitting data to the new path as soon as the handover is completed as explained in Sect. 2.2. Therefore, the proposed scheme always shows smaller handover latency as presented in Fig. 7. Mainly due to the impact of handover latency, it also shows shorter file transfer time than TOM

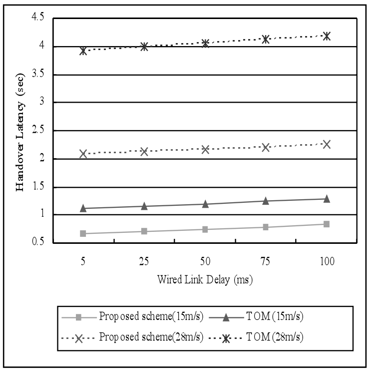


Fig. 8 Handover latency for varying wired link delay.

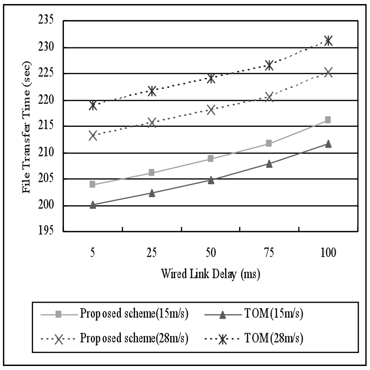


Fig. 9 File transfer time for varying wired link delay.

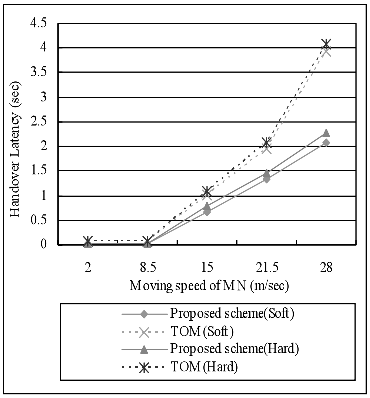


Fig. 10 Handover latency for different moving speed of MN and different handover approaches.

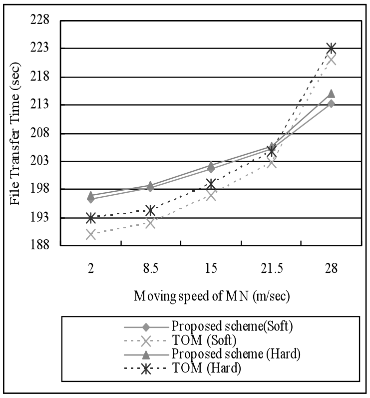


Fig. 11 File transfer time for different moving speed of MN and different handover approaches.

when path acquisition time is larger than 5 seconds as presented in Fig. 6.

Figures 8 and 9 present the handover latency and file transfer time respectively for varying wired link delay. Two different node moving speeds, 15 m/sec and 28 m/sec, are experimented. As shown in Fig. 8, the handover latency grows as the wired link delay increases in both TOM and the proposed scheme since it takes longer for Set-Primary to arrive at CN as the wired link delay increases. With the same reasons explained in Fig. 7, the proposed scheme always has shorter handover latency. Besides, the difference between TOM and the proposed scheme in handover latency is bigger when the moving speed of MN is higher. For higher moving speed, the number of repetitive retransmission time out is larger and it results in longer retransmission timer value. Since CN in TOM has to wait for the retransmission timer to expire before it (re)transmits to the new path, it results in longer handover latency.

As shown in Fig. 9, the file transfer time also grows as the wired link delay increases. It is due to the increase of amount of losses occurred during handover as well as the increased handover latency. Recall that if the link delay is large, the number of outstanding packets is large and as a result losses caused by handoff is also large. One thing to note from Fig. 9 is that when moving speed of MN is 28 m/sec the proposed scheme performs better than TOM whereas the opposite results are shown when moving speed of MN is 15 m/sec. Due to the SCTP packet format, the proposed scheme actually imposes more header related overhead than TOM, and this is mainly the reason that TOM outperforms the proposed scheme when moving speed of MN is 15 m/sec. When moving speed of MN is 28 m/sec, the advantages of the proposed scheme, i.e., faster recovery of losses occurred during handover as well as the shorter handover latency diminishe the adversary impact of higher header overhead.

Figures 10 and 11 show the handover latency and the file transfer time respectively for different moving speed of MN. For L2 handovers, both hard and soft handover cases are experimented. Soft and hard handovers make difference with respect to that from which data path an MN can receive data after L2 handover for a new path is completed within the cell overlapping area. In soft handover, the MN keeps receiving data from both of the old and new paths while it is transiting the cell overlapping area. Whereas, in hard handover, the MN receives data only through the new path once the L2 handover for the new path is completed.

In Fig. 10, as the moving speed of MN becomes faster, handover latency increases in both the proposed scheme and TOM. If two MNs with different moving speed start transiting the cell overlapping area at the same time, the faster

MN should escape from the overlapping area earlier, i.e., the faster MN stops receiving packets from the previous path earlier. Since the path acquisition time is not affected by the moving speed of MN, the time to start receiving packets through the new path is almost the same regardless of the moving speed. Therefore, handover latency becomes larger as the moving speed becomes faster. Furthermore, due to the same reason explained for Fig. 7, the proposed scheme always has shorter handover latency than TOM for all moving speeds.

For both TOM and the proposed scheme, the handover latency is smaller when soft handover is deployed in L2. The time instance for an MN to receive the first data packet from the new path is not affected by the L2 handover schemes, but the time instance for an MN to receive the last data packet from the old path is earlier when hard handover is deployed resulting in longer handover latency.

Figure 11 shows the file transfer time for different node moving speed. It is shown that the proposed scheme outperforms TOM when the moving speed of MN is over 21.5 m/sec. Through the numerical results illustrated in Fig. 6, it is shown that the relative performance gain of proposed scheme compared to TOM, with respect to the file transfer time, becomes greater as the ratio of path acquisition time to cell overlapping area transiting time becomes larger. Since the cell overlapping area transiting time becomes smaller as the moving speed becomes faster, the proposed scheme shows better performance than TOM when the moving speed of MN is relatively faster.

Comparing the results of hard and soft handovers, file transfer takes shorter in soft handover for both of the schemes. Since hard handover incurs more packet losses due to the large handover latency, it takes longer to recover them and results in longer file transfer time.

For the proposed scheme though, the performance variances depending on L2 handover schemes is very small compared to the TOM case. It is due to the effective error recovery scheme deployed in the proposed scheme. First, the proposed scheme finds out the losses that have happened during handover more quickly than TOM. In the proposed scheme, CN immediately sends out the probing packet in order to find out the losses that have happened during the handover when it finds out that there was a handover. On the other hand, TCP in TOM is not aware of handover and has to wait till it receives an ACK through the new path, which may get delayed due to various reasons, i.e., the losses of packets or ACKs transmitted through the old path or the delay of a packet transmission to the new path owing to the erroneous waiting for RTO. Second, the error and congestion control of SCTP itself is more efficient than TCP with respect to a couple of aspects. It uses SACK mechanism and it allows transmission to continue while the sender receives duplicate ACKs. It has been shown that both of these properties enhance the performance [23], [24]. Therefore, the increased handover losses caused by the hard handover are recovered more promptly in the proposed scheme and the impact of the underlying L2 handover scheme is very small in the proposed scheme.

4. Conclusions

We propose a transport layer mobility supporting scheme, which is based on mSCTP and utilizes the link layer radio signal strength information. The proposed scheme specifically addresses the following aspects:

- Determining when to add or delete IP addresses of MN for handovers based on the link radio signal strength information
- Initiating the change of data delivery path and selecting a new data path by MN in case of handovers based on the link radio signal strength information
- Error and congestion control mechanisms to reduce the handover latency, losses, and loss recovery time

The simulation results show that the proposed scheme is very competitive compared to the traditional network layer mobility supporting approach. Especially, when the moving speed of mobile node is fast or new path acquisition takes long, it shows better performance than TOM.

In order to empirically study the performance of proposed scheme, we plan to build a testbed to experiment the proposed scheme. We will also investigate the solutions to deal with the security issues when mobility is handled at the transport layer and compare them with the solutions proposed for the network layer mobility supporting approaches.

References

[1] C. Perkins, “IP mobility support for IPv4,” RFC3344, Aug. 2002.
[2] A. Campbell, J. Gomez, S. Kim, Z. Turanyi, C.-Y. Wan, and A. Valko, “Comparision of IP micro-mobility protocols,” IEEE Wireless Commun. Mag., vol.9, no.1, Feb. 2002.
[3] C. Perkins, “Mobile IP regional registration,” Internet-Draft, draft-ietf-mobile-ip-reg-tunnel-04.txt, March 2001.
[4] R. Ramjee, et al., “IP-based access network infrastructure for next-generation wireless data network: HAWAII,” IEEE Pers. Commun., vol.7, no.4, pp.34–41, Aug. 2000.
[5] A.T. Campbell, et al., “Design, implementation, and evaluation of cellular IP,” IEEE Pers. Commun., vol.7, no.4, pp.42–49, Aug. 2000.
[6] A. Misra, et al., “IDMP-based fast handoffs and paging in IP-based 4G mobile networks,” IEEE Commun. Mag., vol.40, no.3, pp.138–145, March 2002.
[7] A. Bakre and B.R. Bardinath, “I-TCP: Indirect TCP for mobile hosts,” Proc. IDCS, May 1995.
[8] J. Hu and K. Yeung, et al., “Hierarchical cache design for enhancing TCP over heterogeneous networks with wired and wireless links,” GLOBECOM 2000, Nov. 2000.
[9] K. Brown and S. Singh, “M-TCP: TCP for mobile cellular networks,” Computer Communication Review, vol.27, no.5, Oct. 1997.
[10] T. Goff, et al., “Freeze-TCP: A true end-to-end enhancement mechanism for mobile environments,” INFOCOM 2000, March 2000.
[11] A.C. Snoeren and H. Balakrishnan, “An end-to-end approach to host mobility,” ACM/IEEE MobiCom, Aug. 2000.
[12] W. Xing, H. Karl, and A. Wolisz, “M-SCTP: Design and prototypical implementation of an end-to-end mobility concept,” 5th Workshop The Internet Challenge, Technology and Applications, Oct. 2002.
[13] R. Stewart and Q. Xie, et al., “Stream control transmission protocol,” RFC 2960, Oct. 2000.

[14] E. Kohler, M. Handley, S. Floyd, and J. Padhye, “Datagram congestion control (DCCP),” Internet-Draft, draft-ietf-dccp-spec-04, June 2003.

[15] M. Riegel and M. Tuxen, “Mobile SCTP,” Internet-Draft, draft-reigel-tuexen-mobile-sctp-03, Aug. 2003.

[16] J. Rosenberg and H. Schulzrinne, et al., “SIP: Session initiation protocol,” RFC 3261, June 2002.

[17] P. Vixie, S. Thomson, Y. Rekhter, and J. Bound, “Dynamic updates in the domain name system,” RFC 2136, April 1997.

[18] R. Stewart, “Stream control transmission protocol (SCTP) dynamic address reconfiguration,” Internet Draft, draft-ietf-tsvwg-addip-sctp-08.txt, Sept. 2003.

[19] S. Thomson and T. Narten, “IPv6 stateless address autoconfiguration,” RFC 2462, Dec. 1998.

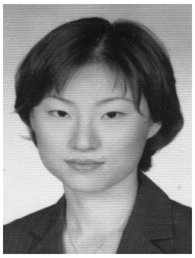
[20] T.S. Rappaport, Wireless Communications, principles and practice, Prentice Hall, 1996.

[21] http://pel.cis.udel.edu/#download

[22] C. Bettstetter and C. Wagner, “The spatial node distribution of the random waypoint model,” Proc. WMAN’02, March 2002.

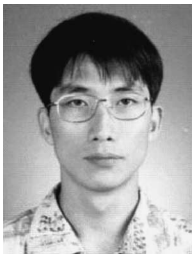
[23] S. Vangala and M.A. Labrador, “The TCP SACK-aware snoop protocol for TCP over wireless networks,” Proc. IEEE Vehicular Technology Conf., Oct. 2003.

[24] S. Fu and W. Ivancic, “Effect of delay spike on SCTP, TCP reno, and eifel in a wireless mobile environment,” International Conf. on Computer Commun. and Networks, pp.575–578, Oct. 2002.



Moonjeong Chang received the B.S. and M.S. degrees in Department of Computer Science from Ewha Womans University in 2001 and 2003, respectively. Since 2003, she has been a Ph.D. student in Department of Computer Science at Ewha Womans University. Her current research interests include multi-hop cellular architecture and mobile SCTP. E-mail: mjchang@ewha.ac.kr

Meejeong Lee received her B.S. degree in computer science from Ewha Womans University, Seoul, Korea in 1987, M.S. degree in computer science from University of North Carolina, Chapel Hill in 1989, and Ph. D degree in computer engineering from North Carolina State University, Raleigh in 1994. In 1994, she joined the Department of Computer Science and Engineering, Ewha Womans University, Seoul, Korea, where she is currently an associate professor. She has been engaged in research in the field of computer communications and networks, and performance modeling and evaluation. Her current research interests focus on routing and transport protocols for multimedia traffic transmission over wireless mobile networks and Internet. She is a member of IEEE, ACM, and KICS (Korean Information Communication Society), and serving as an editor of KISS (Korean Information Science Society) and KIPS (Korean Information Processing Society), and as an associate electronic publication editor of JCN (Journal of Communications and Networks). E-mail: lmj@ewha.ac.kr



Seokjoo Koh received B.S. and M.S. degrees in KAIST in 1992 and 1994 respectively. He also received Ph.D. degree in Industrial Engineering from KAIST in 1998. Since 1998, he had worked for the Protocol Engineering Center in ETRI. He has actively been participated in the project of ECTP (Enhanced Communications Transport Protocol) in ITU-T SG17, and in the project of Mobility Management for Systems Beyond 3G in ITU-T SSG. He has also been involved in the IETF in the area of the mobile SCTP. He has now joined the Department of Computer Sciences in KNU (Kyungpook National University) since March 2004. His current research interests include Internet Mobility in Wireless Mobile Communications Networks.