A Study on the Performance Enhancement of the Macro Handover in HMIP According to Protocol Layers

Jongjung Woo*, Chi-Hyun Ahn**, *Member, KIMICS*

***Abstract*—The Network-based handover still has problems such as the transmission delays and the packet losses in the case of macro mobility, though technological advances have been made in the wireless and mobile communication. For end-to-end handover, the link bandwidth has been reduced in the wireless network due to its burst errors and congestion control. To overcome such problems, we propose a new scheme of the macro handover according to the protocol layer. The proposed macro handover is implemented on the network layer to partially substitute wired signaling for wireless signaling, to flexibly employ buffers, and on the transport layer to postpone its retransmission time. We have performed extensive simulation using ns-2 and the result shows that our proposed scheme outperforms the other existing schemes in terms of transmission delay, packet loss, and data transfer rate during the handovers.**

***Index Terms*—HMIP, network-based handover, end-to-end handover, protocol layer**

I. INTRODUCTION

The basic infrastructure of mobile systems has recently been an IP-based wireless communication. The key technology in wireless IP networks is to ensure that IP addresses are seamlessly maintained during handovers [1],[2]. In 2004, the IETF (Internet Engineering Task Force) proposed the MIP (Mobile IP) for supporting mobile devices on the Internet [3]. However, the MIP poorly performs during the macro handover, an inter-domain movement of a MN (mobile node). To overcome this problem, the IETF also proposed two extensions: the FMIP (Fast handover for MIP) in 2005[4] and the HMIP (Hierarchical MIP) in 2007[5]. However, these extensions still have some problems, including increasing the traffic signals and packet loss by the former, and a triangular routing problem when the MN moves fast by the latter. Moreover, many studies have proposed various wireless TCP protocols for utilizing the TCP in wireless networks, but these proposed methods still have problems such as burst errors and reduction of the link's bandwidths due to their congestion control.

This paper proposes a new scheme that improves the handover performance according to protocol layers. On the network layer, our scheme substitutes the wired signaling for wireless signaling, and then it employs a buffer for improving the quality of multimedia traffic service. On the transport layer, it postpones its retransmission time, which reduces the packet loss and the handover delay and it prevents the networks from operating under the TCP control mode.

The remainder of this paper is organized as follows. Section 2 discusses related works. Section 3 presents our proposed scheme in detail. In section 4, we describe mathematical analysis and simulation results. Finally, conclusions are provided in section 5.

II. RELATED WORKS

The IP handover technology consists of a network-based handover and an end-to-end handover in the protocol layer.

Typical network-based protocols are MIP, HMIP, and FMIP [3-5]. The HMIP employs the MAP (Mobility Anchor Point) that allows it to play a role of regional HA (Home Agent), which improves its performance by reducing its outgoing signals when a micro handover occurs; however, it increases the signaling, and it causes a delay because the MN configures 2 CoA (Care of Address) and transfers 2 BU (Binding Update) messages. It is difficult to offer real-time multimedia services because it does not have any buffering capability. The FMIP acts when a link change occurred in layer 2 is detected before a layer 3 handover takes place and it establishes the handover environment in advance. Thus, it can reduce the time that is spent on completing a layer 3 handover, and it performs the configuration of the nCoA (new CoA) or the pre-establishment of the tunneling link between the PAR and the NAR before the L2 handover. However, the FMIP cannot optimize the routing path because it sends a BU message after the handover, and if the MN moves fast, then a triangle routing problem can occur.

The wireless link has high BER (Bit Error Rate) characteristic because of its path losses, fading, noises and interferences. The use of the wired TCP on a

---

Manuscript received September 21, 2009; revised September 18, 2009; accepted October 8, 2009. This work was supported by the Sungshin Women's University Research Grant of 2009.

* First Author: Jongjung Woo is with School of Information Technology, Sungshin W. University, Seoul, 136-742, Korea (E-mail: jwoo@sungshin.ac.kr)

** Chi-Hyun Ahn is with the Department of Information & Communication, Open Cyber University, Seoul, 110-776, Korea (E-mail: jfsodmes@chol.com)

wireless link makes it difficult to identify the reasons of packet losses, congestion or the characteristics of the wireless links. To solve these problems, various mechanisms such as Freeze-TCP [7], SNOOP [8] has been proposed. The Freeze-TCP does need not to establish a connection in the CN and the BS (Base Station) because the flow control is performed in received node, but the MN periodically performs the ZWP (Zero Window Probes). For this reason, the transmission period of the ZWP is exponentially increased. Thus it can be difficult for the MN to resume transmission for a long period of time after the re-connection occurs. The SNOOP is divided into wired and wireless connection based on the BS. When packets are lost, the SNOOP improves the TCP performance through the local re-transmission in the wireless link to avoid the fast re-transmission or the congestion control in the CN. The SNOOP inserts a SNOOP module into the TCP window with maximum size by transforming the network layer software in the BS. Thus, for a single error, the SNOOP can improve the performance by carrying out local re-transmissions. However, during burst errors, the SNOOP causes the CN to generate unnecessary re-transmissions because re-transmission timeout is generated and it operates in the flow control mode.

III. PROPOSED HANDOVER

3.1 The Network Layer

In our proposed scheme, the PAR (Previous AR) sends its neighbor AR the nCoA configuration information (interface ID and prefix) in advance. The address is pre-configured before the handover takes place, and then the MN moves and receives data at the same time. After its movement, the MN requests the registration to the NAR (New AR). The NAR can be directly connected with the MN, and it can transfer the MN by tunneling data from the PAR. This scheme can remove critical factors of delay of the handover by configuring addresses and performing the DAD (Duplicate Address Detection).

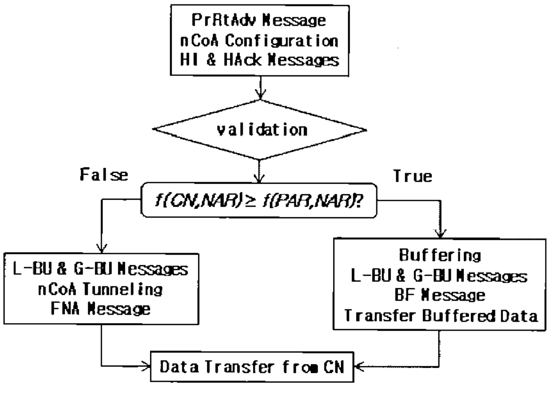


Fig. 1. Behavior of the Proposed Scheme on the Network Layer

This scheme operates at the network layer, as is shown in Figure 1. Before the handover occurs, the PAR shares the prefix with the NAR and the NMAP through a PrRtAdv (Router Solicitation for Proxy) message, thereby meeting configuration requirements for a new address. Hence, the PrRtAdv message includes a MAC address and other information which the PAR provides its neighbor NAR. The time for the address configuration is 40~60% of the Delta SNR (Signal-to-Noise Ratio) at the beginning point of CST (Cell Search Threshold). The NAR configures the nLCoA (new on-Link CoA) and the nRCoA (new Regional COA) of the MN. The configured nCoA performs the DAD in the NAR by sending the NMAP a HI (Handover Initiate) message. After that, the validation through HAck (Handover Ack) message is informed to the NAR which in turn registers the nCoA. If the handover is occurred by L2 trigger, the PAR sends the NAR a HI message requesting the nCoA. So the NAR sends the PAR a HAck message including the nCoA and the channel for MN's new addresses assignments. The PAR sends the MN a BU message including the nCoA and the NAR sends the CN a BU message on behalf of the MN. Next, the PAR determines the use of the buffer according to the following conditions.

$$f(CN, NAR) \geq f(PAR, NAR) \quad (1)$$

$$f(CN, PAR) \geq f(PAR, NAR) \quad (2)$$

,where f(a, b) means a hop count between a and b. When compared to the distance between the PAR and the NAR, (1) represents the distance between the CN and the NAR as being far and (2) represents the distance between the CN and the PAR as being far. Figure 2 is divided into three groups, and it represents a signal diagram of our scheme according to above conditions. The most upper part is not dependent on any of these conditions

The middle part on the dotted box in Figure 2 is a diagram that is used for signaling when both conditions are satisfied. The PAR sends the MN an nCoA by exchanging HI and HAck messages with the NAR. Next, the PAR performs the buffering management by sending the NAR a BR (Buffering Request) message. Buffering management avoids out-of-packet sequences that are transmitted after the binding updates. The packet that is transmitted to the CN during the disconnection of the MN is forwarded into the NAR through the tunneling. The PAR that receives an nCoA from the NAR performs the binding update by sending a BU message to the NMAP and the CN. After the handover is completed, the MN sends the NAR a BF (Buffering Forward) message which represents the completion of the MN's movement and transmission of the buffered data after the buffering has been stopped. Next, the CN sends the binding updated packets to the nRCoA. The NMAP passes the nLCoA packets, and finally the MN can receive packets.

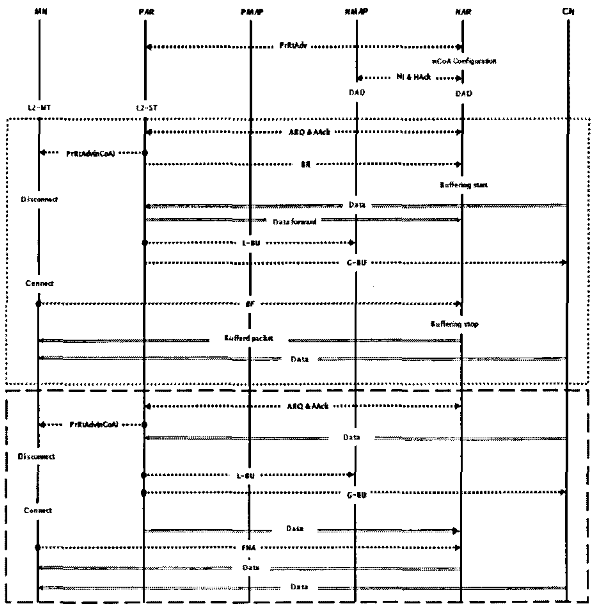


Fig. 2. Handover Signaling

The bottom part on the dashed box in Figure 2 is a diagram for signaling when neither (1) nor (2) are satisfied. The PAR informs the NMAP and the CN of the binding updates through BU messages with valid addresses. During disconnection of the MN, the PAR forwards the received packets into the MN via tunneling to the nCoA. After the handover is completed, the MN sends the NAR a FNA (Fast Neighbor Advertisement) message. The FNA message represents the completion of the MN's movement and requests the transmitted data by tunneling after comparing the addresses. The CN acts the binding update and sends the data to the nRCoA of the MN. Note that the ARQ (Address ReQuest) and the AAck (Address Ack) in Figure 2 are a message which the PAR requests the NAR for an nCoA and a message which the NAR requests the PAR for an nCoA.

3.2 The Transport Layer

During burst errors, our scheme processes the serial numbers of lost packets and received packets in a batch mode as is shown in Figure 3, and it re-transmits the lost packets faster than existing schemes because our scheme flexibly configures the ACK option of the SACK depending on the degree of burst errors, that is, the errors among the lost packets and the transmitted packets. Our scheme configures the ACK option by using the serial number of the smaller one. It uses a 1-bit flag of the ACK option that identifies the kinds of lost packets or transmitted packets. Through this process, the MN sends the BS an occurrence of burst errors. The BS works in a re-transmission mode, thereby sending the CN a ZWA message which delays a transfer for FT time. The FT is the time that is derived by multiplying the number of lost packets with the round trip time from the BS to the MN. At this point, the BS re-transmits the lost packets to the MN by using a wireless link. After the re-transmission, the MN sends an ACK to the CN, which begins the transmission process again. If the CN receives an ACK massage before the FT, the CN begins transmission again as soon as the CN received an ACK message. Next, the CN quickly recovers the network. If an ACK message is not arrived yet before the FT, the CN recognized the timeout and it will work in the flow control mode.

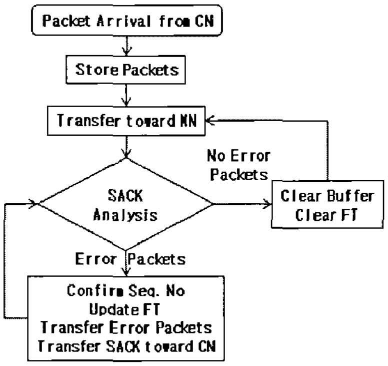


Fig. 3. Behavior of the Proposed Scheme on the Transport Layer

IV. SIMULATION RESULTS

4.1 Mathematical Analysis

If the Poisson distribution is applied to the mobile system, the traffic load for a cell (L) can be equal to Equation (3), which consists of traffic for nCoA configuration (l_a), traffic for binding update (l_u), a traffic for buffering (l_b), and traffic for wired messages (l_w).

$$L = l_a + l_u + l_b + l_w \qquad (3)$$

,where l_a is the same for all schemes. Assuming that the number of the MN is n, n message transfers and $2nP$ processes are needed for achieving the signaling load of the nCoA configuration. While $l_b = l_w = 0$ in the HMIP, $l_w \neq 0$ and $l_b \neq 0$ in the FMIP and our proposed scheme.

In the HMIP, the MN receives a message from the NAR after the handover, and then it configures the nCoA. Next, the MN receives a packet after it has sent a BU message to the MMAP and the CN. There is no traffic before the handover, but a traffic load is required for completing the binding updates and the DAD operations ($4n+10nP$). Therefore, the total traffic loads of the HMIP (L_H) are as follows:

$$L_H = \frac{\lambda_a}{V_m}(n+2nP)+\frac{\lambda_u}{V_m}(4n+10nP) \qquad (4)$$

,where λ_a and λ_u are probabilities of the channel connections for the address configuration and the binding update, respectively. V_m is the transmission rate on the wireless network; P means all processes generated by the MN or the AR.

The FMIP handover is based on the tunnel. If the mobile device receives a L2 trigger on the tunnel based handover, the PAR and the NAR exchange their messages to generate a bi-directional tunnel. When L2 is updated, the PAR forwards a packet for the MN into the NAR. Finally the NAR transfers this packet into the MN. Therefore this scheme requires a message process ($2n+4nP$) for tunneling on the wired network. In addition, after the handover is completed, the traffic loads are needed for the nCoA configuration (l_a), for the binding updates and the DAD operations ($4n+10nP$). Thus, the total traffic loads of FMIP (L_F) are as follows:

$$L_F = \frac{\lambda_a}{V_m}(n+2nP) + \frac{\lambda_u}{V_m}(4n+10nP) + \frac{w}{V_w}(2n+4nP) \quad (5)$$

Hence, w and V_w mean probabilities of the channel connections for the wired network and the transmission rate on the wireless network, respectively.

In the proposed scheme, most of signals' traffic is generated on the wired network. The messages on the wireless network are considered nCoA configurations (l_a) before the handover; a BF (FNA) message is forwarded into the NAR ($n+2nP$) after the handover. In addition, this scheme performs a process for buffering message on the wired network; this process of the PAR and the BU message configures the nCoA. Thus the total traffic loads of the FMIP (L_N) are as follows:

$$L_N = \frac{\lambda_a}{V_m}(n+2nP) + \frac{\lambda_u}{V_m}(n+2nP) + l_b + \frac{\omega}{V_w}(6n+10nP) \quad (6)$$

where l_b is $w(2n + 2nP)/V_w$ in the buffering scheme and n/V_w in the forwarding scheme.

According to Eq. (4) ~ (6), the total traffic of each scheme depends on l_u; the reason is that l_b and l_w (in L_F and L_N) were produced on the wired network, are very small when they are compared with l_a and l_u that were produced on the wireless network because of $V_w \gg V_m$; however, l_a is the same for all schemes. Consequently, lu of the proposed scheme is reduced to about $(1 + 2P)/(4 + 10P)$ of the other schemes because the proposed scheme substitutes wired signaling for wireless signaling.

4.2 Simulation Environments

The network environment for the simulation consists of a MAP router, an AR, and four cells. Each cell operates as an AP for the MN which randomly moves as a receiver. Our simulation is performed by Network Simulator2 2.1b6 [10] on the Redhat 8.0. We use Mobiwan [11] for HMIPv6 operation.

The MN moves at the fixed speed of 5m/s. The TCP traffic is set as 0.5 ~ 64 sec, and it is set into 2 sec for initial retransmission time-out. For the TCP, the window size is 30 segments with the maximum size being 500 bytes long. Transmission rates are 100 Mbps and 2 Mbps, and the link latencies are 10 ms and 200 ms respectively, for each wired and wireless network. The number of links between the CN and the NM is 1 ~ 12 and the buffer of the AR is 10 Mbytes long. We ignore the error rates of the end-to-end handover on the wired network. The error rate of the retransmission is 6.02×10^{-4}. We take advantages of the 2-state Markov model for the characteristic of the burst errors on the wireless network.

4.3 Simulation Results

Figure 4 shows simulation results for *f(CN, NAR)/f(PAR, NAR)* versus the ratio of the delay times, which is the delay times of the FMIP and our scheme divided by that of the HMIP respectively. Hence, the delay is calculated from the time which the MN is cutoff in the PAR area to the time which it starts a handover and then it receives a packet from the CN through the new MAP. Compared with the HMIP, the bigger than 1 the *f(CN, NAR)/f(PAR, NAR)* is, the better is transmission delay in our proposed scheme. The simulation result shows that ours can reduce the transmission time by 5.6~20.1%. But there is a slight difference in transmission delay between ours and the FMIP because the FMIP does not handle the BU messages and the binding updates during a handover. Therefore, the farther from the NAR the CN is, the better our proposed scheme performs.

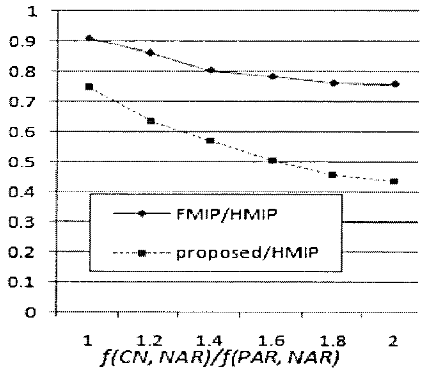


Fig. 4 Ratio of delay time vs. relative distance of CN

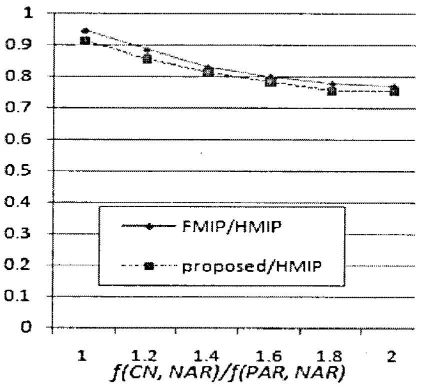


Fig. 5 Ratio of lost packets vs. relative distance of CN

Figure 5 shows simulation results for *f(CN, NAR)/f(PAR, NAR)* versus the ratio of the lost packets, which are the lost packets of the FMIP and our scheme divided by those of the HMIP respectively. The packet loss is the number of packets that are discarded by the MN during a handover and during the forwarding process of the BU messages towards the CN after its handover has taken place. Compared with the HMIP, our scheme can reduce packet loss by average 47.5%. The greater the *f(CN, NAR)/f(PAR, NAR)* is, the better it can reduce packet loss by about 25.3 ~ 47.5%. Moreover, ours can reduce packet loss by 73.5% more than that of the FMIP. Therefore, the farther from the NAR the CN is, the better our proposed scheme performs.

Figure 6 shows simulation results for the ratio of the data transfer rates versus mean bad period. Hence, the ratio of data transfer rate means is calculated by dividing the data transfer rate of SACK and our scheme by that of SNOOP respectively. The simulation result shows that the transfer rates of SACK and those of ours are superior to those of SNOOP. However, while the SACK and ours are similar for the MBP less than 64 ms, they have great differences when the MBP is more than 128ms. Thus, the transfer rate of ours is better than that of the SACK because it is possible for the CN to operate in the TCP flow control mode when burst errors frequently occur.

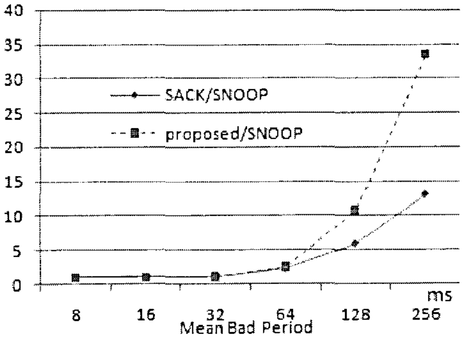


Fig. 6 Data transfer rate vs. mean bad period

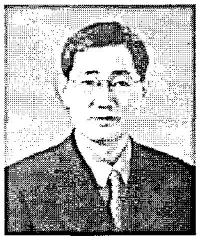
IV. CONCLUSIONS

The HMIP causes a lot of packet losses because it is inevitably accompanied by its transmission delays due to its macro handover, and the binding update becomes late due to the long distance between the CN and the NAR. On the other hand, the wireless network experiences packet losses, transmission delay, and the temporarily broken transmissions because the BS is changed when a handover in the wireless link occurs or when the TCP congestion control or the retransmissions are performed.

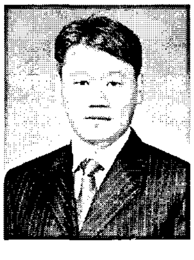
This paper proposed a new scheme of the macro handover according to protocol layers in order to overcome the aforementioned problems. Our scheme is implemented on the network layer to selectively exploit wired signaling instead of wireless signaling, which has higher error rates, to configure the nCoA for the MN. It also performs a binding update before the handover. Moreover, the NAR determines whether or not the buffering is employed. By the way, our scheme on the transport layer protects the retransmission time of the SNOOP, exploits the ZWA, and keeps networks from operating under the TCP flow control mode, thereby reducing the delay time of the DAD and increasing QoS of real time multimedia. Eventually our scheme can effectively manage burst errors and retransmission errors, causes of packet losses. The simulation results with an NS-2 show that our scheme is better than the other existing schemes on both the network and the transport layers.

REFERENCES

[1] ITU-T Recommendation Q.1706/Y. 2801, "Mobility Management Requirements for Next Generation Networks," November 2006.
[2] ITU-T SG19 Draft New Recommendation Q/Y.MMF, "Generic Framework of Mobility Management for Next Generation Networks," Working in Progress, 2007.
[3] D. B. Johnson and C. Perkins, "Mobility support in IPv6," RFC3775, Jun. 2004.
[4] R. Koodi, "Fast Handovers for Mobile IPv6," RFC 4068, July 2005.
[5] Hesham Soliman, Claude Castelluccia, Karim Malki, Ludovic Bellier, "Hierarchical Mobile IPv6 mobility management (HMIPv6)," RFC 4140, 2007
[6] A. Bakre, B. R. Badrinath, "I-TCP: Indirect TCP for Mobile Hosts," Proc. of the 15th Int,l conference of distributed Computing systems, June 1995.
[7] K. Brown and S. Sigh, "M-TCP: TCP for mobile cellular networks," ACM computer communication review, vol. 1 27, no. 5, October 1997.
[8] Tom Goff, James Moronski, Vipul Gupta, "Freeze-TCP: A true end-to-end TCP enhancement mechanism for mobile environments," 1995.
[9] H. Balakishnan, S. Seshan, and R, H, Kartz, "Improving reliable transport and handoff performance in cellular wireless networks," ACM Wireless Networks, vol. 1, December 1995.
[10] NS-2, http://www.isi.edu/nsnam/ns
[11] "Mobiwan: NS-2 extensions to study mobility in WideArea IPv6 Networks", http://www.inrialpes.fr/planete/pub/mobiwan



Jongjung Woo received the B.S. degree in Department of Electronics from Kyungpook National University in 1982, and M.S. and Ph.D. degree in Department of Electrical & Computer Engineering from University of Texas at Austin, in 1990 and 1993, respectively. He has worked as a professor in School of Information & Technology at Sungshin W. University since 1993. His current research interests include embedded software, mobile computing, computer architecture, parallel processing, and e-learning.



Chin-Hyun Ahn received the B.S. degree in Department of Electronics from Hannam University in 1998, and M.S. and Ph.D. degree in Department of Electronics from Kyunghee University, in 2000 and 2007, respectively. He has worked as an inviting professor in Department of Information & Communication at Open Cyber University since 2005. His current research interests include Mobile IP, Wireless TCP, and mobile communication.