Mobile IPv6망에서 Smooth 핸드오프 패킷의 과도기간 분석 및 단축

정회원 이 동 욱, 김 종 원*

Analysis and Reduction of Transient Time Periods for Smooth Handoff Packets in Mobile IPv6 Networks

Dong-Wook Lee, Jong-Won Kim *Regular Members**

요 약

본 논문에서는 Mobile IPv6 망에서 Smooth 핸드오프가 패킷의 전송에 끼치는 영향을 분석한다. 분석결과 핸드오프는 단말의 관점에서 불안정구간(UTP), 묵음구간(UTP), 핸드오프구간(HTP)으로 구분됨을 보인다. 또한 정확한 구간의 값을 구하기 위해 큐잉 모델을 제시한다. 수치해석의 결과 불안정구간, 묵음구간, 핸드오프구간은 핸드오프되는 패킷의 큐잉 지연과 패킷이 전송되는 링크의 대역폭에 영향을 받음을 보인다. 한편, 핸드오프의 수행기간이 길어지면 핸드오프로 인한 응용의 피해는 증가한다. 우선순위 스케줄링 기법을 이용하면 기존의 IP망에서 최선형 서비스를 위한 FIFO 스케줄링 방식보다 더 짧은 과도기간을 가짐을 확인한다.

Key Words : Mobile IP, smooth handoff, packet sequence analysis, transient time periods, and queuing analysis.

ABSTRACT

In the paper, we investigate the impact of handoff on the packet delivery in the Mobile IPv6 networks, where the smooth handoff is adopted. That is measured by an 'unstable time period (UTP)', a 'silence time period (STP)', and a 'handoff time period (HTP)' in the mobile node's perspective. Then, we propose a queuing model to get the exact value of the handoff transient time. The numerical results show that the queuing delay for the handoff packets and the involved link (or route) capacities affect the estimated UTP, STP, and HTP. On the other hand, the damage of application caused by handoff will increases when the handoff transient time becomes longer. We show that the priority scheduling method can achieve shorter STP and UTP than the FIFO scheduling method that is generally used in best-effort IP network.

Ⅰ. Introduction

Mobile IP has been recommended by the Internet Engineering Task Force (IETF) as the baseline protocol to support the mobility for Internet services [1]. Initial version of the Mobile IP is known to cause packet losses during the handoff. It can be patched by the route optimization extension [2], which includes binding caching and smooth handoff options [2]. IETF Mobile IPv6 [3] adopts the route optimization with smooth handoff as a basic option.

* 광주과학기술원 정보통신공학과 네트워크 미디어 연구실 ({dulee, jongwon}@kjist.ac.kr)
¶ 본 연구는 2002년도 한국학술진흥재단의 지원에 의하여 연구되었음 (KRF-2002-041-D00448).
※본논문은 2003년 4월 JCCI 학술대회에서 우수논문으로 선정되어 게재 추천된 논문입니다.

Unfortunately, the smooth handoff introduces packet sequence disruption (i.e., out-of-order sequence) between the directly delivered new packets and re-routed old packets [4,5]. These out-of-sequence packets bring in several unwanted effects. For TCP congestion control, they create duplicated ACK's and invoke unnecessary packet re-transmissions. It is hard to realize real-time applications on top of the Mobile IP, since the required buffering to compensate transient behaviors can waste limited latency budget and thus can degrade the overall quality. Thus, lots of schemes have been proposed to address this problem, which range from a simple modification of TCP to resource-demanding 'refuge proxy [4]' scheme. In [6], a buffered packet scheme is proposed. In the wireless ATM area, several works have explored the seamless handoff issue [7], where they mostly focus on the protocol aspects for the seamless handoff and less attention is paid on the quantitative analysis of the above transient behaviors. In this paper, we investigate the impact of handoff rate, traffic load, and involved queuing scheme on the packet sequence in the Mobile IPv6. The impact is measured by an 'unstable time period (UTP)' and a 'silence time period (STP)'. The UTP explains the time duration of out-of-sequence packets while the STP reflects the blackout duration of a mobile node (MN) after the MN initiates handoff. The 'handoff time period (HTP)' is the transient time duration until the handoff completes, which can be derived from the analysis result on the UTP and STP. The numerically calculated analysis result show that the queuing delay for the handoff packets (i.e., affected by background traffic) and the involved link (or route) capacities affect the estimated UTP, STP, and HTP. Moreover, two schemes such as priority and FIFO forwarding are compared to show that the priority method can remarkably reduce those transient time during the handoff of mobile terminal.

The remainder of this paper is organized as follows. In Section 2, we explain how the out-of-sequence packets are encountered, and define the UTP, STP, and HTP. The queuing delay involved in the forwarding of handoff packets is introduced in Section 3. Section 4 provides the numerical result. Final Section5 gives the discussion.

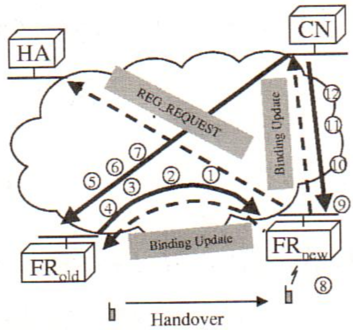


Figure 1. Example packet sequence under the smooth handoff in a Mobile IPv6 network.

II. Transient Time Periods

Under the Mobile IPv6, a mobile node (MN) can move out of the reach in the midst of a data transfer [3]. Let FR$_{old}$, and FR$_{new}$ be the FR's where the MN had stayed at and moves to, respectively in Fig.1. After the handoff, the MN will experience the blackout period and there is no packet exchange. At the same time, a number of packets would be delivered to the FR$_{old}$ rather than the FR$_{new}$. These packets that lose their route are called as astray packets. In Mobile IPv6, FR$_{old}$ forwards the astray packets for the smooth handoff.

When the MN initiates the handoff, it sends a *BIND* (binding update message) to its CN via the FR$_{new}$. The MN also sends another *BIND* message to the FR$_{old}$ to announce a new MN location [3]. Before the CN receives the *BIND* message, it continues to route packets to the FR$_{old}$; this packet stream is denoted as an *old stream* throughout the paper. After receiving the *BIND* message, the CN re-routes the packets to the FRnew; we call this packet stream as a ***new*** *stream*. Also, according to the smooth handoff procedure, the packets in the old stream is

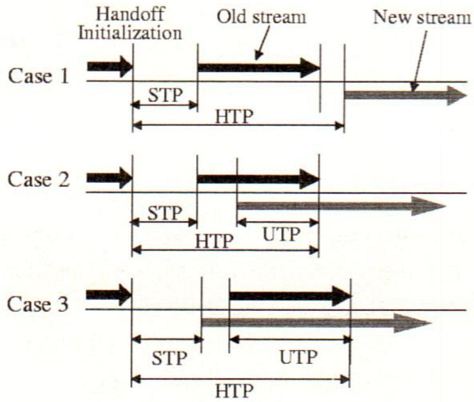


Figure 2. Possible packet orderings observed at the MN during the handoff (for the traffic from the CN to the MN).

forwarded to the MN by the FR_{old}. Thus, both old and new streams can be mixed together at the FR_{new}[1]

Under a Mobile IPv6 networks with smooth handoff, the MN must wait for a time period to receive packets after the handoff initialization. We define this blackout period as the STP (silence time period) during which there is no packet exchange with the CN. Also, if we know when a new stream arrives to the FR_{new} earlier than the last packet of an old stream, we can estimate the number of out-of-sequence packets. We thus define this period as the UTP (unstable time period) during when the packet sequence could be mis-ordered. Finally, we define HTP (handoff time period) as the total transient time duration until the handoff completes.

Depending on the location of FR_{old} and FR_{new}, the timing between the old and new streams can be classified into three cases[2]. As shown in Fig. 2, we can associate the transient time periods

1) In Fig. 1, the astray packets (from 1 to 8), which have already been in-flight until the binding change of CN, are to be forwarded to the FRnew. These astray packets are causing the out-of-sequence packets, since packet 9 reaches the FRnew earlier than the astray packets.

2) Throughout in this paper we mostly consider the one-way traffic from the CN to the MN. Note however that similar analysis can be conducted for the reverse traffic.

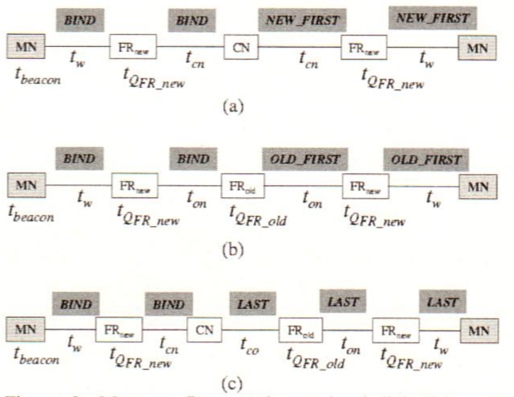


Figure 3. Message flows and associated link delays for various packet streams (for the case when the MN moves from a FR_{old} to a FR_{new}); (a) The first packet of the new stream (for T_{f_new}), (b) the first packet of the old stream (for T_{f_old}), and (c) the last packet of the old stream (for T_{l_old}).

(i.e., STP, UDP, and HTP). The 'Case 1' illustrates the situation where the MN moves to a far-away (in network routing sense) FR_{new} from the CN. In contrast, 'Case 3' represents the movement of the MN to the FR_{new} close to the CN. Note that these two cases are kind of extreme examples. Usually we expect to see timing of 'Case 2', since typical movement of the MN is localized while the distance between the MN and the CN is large.

Thus, by setting the time when the MN leaves the FR_{old} to zero, the STP is

$$STP = min(T_{f\_old},\ T_{f\_new}), \qquad (1)$$

where T_{f_new} is the time when the first packet of new stream arrives to the MN and T_{f_old} is the time when the first packet of the old stream is delivered to the MN. That is, for 'Case 1' and 'Case 2', the old stream arrives to the MN earlier than the new stream and T_{f_old} determines the STP. Also, as shown in Fig. 2, the mix of old and new streams can happen in two cases ('Case 2' and 'Case 3'). It is clear that the out-of-sequence packets are generated in the new stream until all packets of the old stream are transferred to the MN. Thus, in all cases of Fig. 2, the UTP is determined by

$$UTP = max(0,\ T_{l\_old}-T_{f\_new}), \tag{2}$$

where T_{l_old} is the time when the last packet of the old stream is delivered to the FR_{new}. Thus, the whole transient time period after handoff, the HTP (depicted in Fig. 2), can be denoted by

$$HTP = UTP + STP + |T_{f\_new} - T_{f\_old}|. \tag{3}$$

To obtain these timings, in this work, we focus on the movement between FR's since it is the most comprehensive among three cases. We assume that the *BIND* message is not delayed in the nodes but served first. It is also assumed that there is no packet loss in links while we address the queuing loss in our analysis.

The link delays between CN - FR_{old}, CN - FR_{new}, and FR_{old} - FR_{new} are denoted by t_{co}, t_{cn}, and t_{on}, respectively. A new handoff connection is started when the MN moves from the FR_{old}. After the departure, there is a delay (which contributes to the part of blackout period) until the moved MN detects the FR_{new} and sends a *BIND* message to it. We denote this delay as t_{beacon}, which is mainly affected by the waiting time due to periodical router advertisements. The propagation delay of wireless link, denoted by t_w, also affects the above delay.

As shown in Fig. 3(a), T_{f_new} is affected by the link delay for the *BIND* message and the queuing delays for packet itself. It equals to $t_{beacon}+2t_{cn}+t_{QFR\_new}+2t_w$, where t_{QFR_new} is the queuing delay of the first packet at FR_{new}. From Fig. 3(b), T_{f_old} is represented by $t_{beacon}+2t_{on}+t_{QFR\_old}+t_{QFR\_new}+2t_w$, where t_{QFR_old} is the queuing delay of the last packet at FR_{old}[3]. Similarly T_{l_old} is $t_{beacon}+t_{cn}+t_{co}+t_{QFR\_old}+t_{on}+t_{QFR\_new}+2t_w$ as shown in Fig. 3(c).

From Eq. (1), the STP is the minimum of T_{f_new} and T_{f_old}. Although it may vary depending on the actual network configuration, we need to focus on the t_{QFR_old} and the t_{beacon} time for the STP analysis. Also, if we substitute the above results into Eq. (2), then we can represent the UTP by

$$UTP = max(0, (t_{co}-t_{cn}) + t_{QFR\_old} + t_{on}). \tag{4}$$

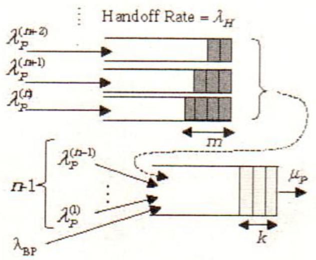
The main causes of out-of-sequence packets are the link delay difference between FR_{old} and FR_{new} from the CN, the queuing delay at the FR_{old}, and the link delay t_{cn}. Note that, the UTP can be reduced effectively by decreasing t_{QFR_old}, since $(t_{co}-t_{cn})$ and t_{cn} depend on the underlying network. Similarly, we can obtain the HTP from Eq. (3).

Ⅲ. Analysis and Reduction – Prioritized Queuing

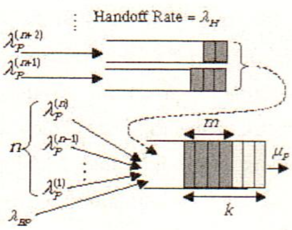
In this section, we perform queuing analysis to estimate the variable t_{QFR_old}, which is t_Q for simplicity, in Eq. (4). The packet delay is mainly constrained by scheduling method, and FIFO (First-In-First-Out) scheduling method is generally used in the best-effort IP network. Under the FIFO scheduling method, the packets from the old stream have the same priority to the background traffic. If we apply the prioritized queuing for the astray packets in a router during the smooth handoff, we can reduce the transient period effectively. The prioritized queuing for handoff connections implies that the astray packets are served first at the FR_{old}. Thus the astray packets from handoff connections don't compete with the background traffic.

After a MN moves out of FR_{old}'s reach, it initiates the handoff process. Since the FR_{old} has lost the contact with the MN, the FR_{old} stores the astray packets in a temporary buffer until the *BIND* message arrives as shown in Fig. 4(a). With the arrival of the binding update message for the *n*th connection in Fig. 4(b), all packets (the number of packets are *m*) residing in the temporary buffer will move to the system buffer. To the system buffer at the FR_{old}, this movement can be modelled as a bulk arrival. After this

3) However, in the queuing analysis of Section 3, we approximate t_{QFR_old} with the average queuing delay without differentiating the queuing delay of first and last packets of the old stream.



(a)



(b)

Figure 4. Queuing system model at the FR_{old} immediately (a) before and (b) after the arrival of the binding update for a handoff connection.

point, all subsequent packets of this handoff connection directly enters the system buffer without going through the temporary buffer and the number of on-going handoff connections becomes n. Note however that n decreases by one after the last packet of any on-going handoff connections arrives into the system buffer of the FR_{old}.

This situation can be explained by a state transition diagram for the system buffer, which leads us to a special case of G/M/1 queuing type with bulk arrival and others. By calculating the steady state probability from the state transition diagram, we can get the average queuing delay $\overline{t_Q}$ of packets at the FR_{old} as follows.

We assume that the arrival rate of a new handoff connection, the packet arrival rate of all the i'th handoff connection, and the background packet arrival rate are the Poissoin Process with λ_H, $\lambda_P^{(i)}$, and λ_{BP}, respectively. The packet service time is exponential distribution with $1/\mu_P$. We assume that the traffics from all handoff connection are heterogeneous for simplicity. Then, $\lambda_P^{(i)}$ becomes λ_P. A new handoff connection is started when a MN leaves a FR_{old}. Then, the time duration when the FR_{old} must save packets in the temporary buffer, t_{buf}, equals to $t_{beacon}+t_w+t_{on}$. The average number of packets to be buffered is thus $L_{buf} = t_{buf} \cdot \lambda_P$. If we denote by g_m the probability that the number of stored packets is m, g_m can be given by $g_m = p[bulk = m] = \dfrac{L_{buf}^m}{m!} e^{-L_{buf}}$.

By taking the number of on-going handoff connections (n) and the packet number at the system buffer (k) as the state variables (i.e., n,k), we have two-dimensional state transition diagram as shown in Fig. 5. Here, N_H stands for the allowed total number of handoff connections and K is the maximum size of system buffer.

The a_m represents the bulk arrival of a handoff connection with bulk size m, where a_m is $\lambda_H \cdot g_m$. The b_i's stand for the arrival rates at which the bulk arrivals cause the buffer to be full. To fill the system buffer full from state (n,k), the bulk size must be bigger than K-k. That is, b_{K-k} is $\sum_{i=K-k}^{\infty} a_i$.

The c_n represents the packet arrival rate into the system buffer when there exist n handoff connections. For FIFO method, the handoff traffic competes with the background traffic and ,however, for the proposed priority scheme, the handoff traffic competes with the other handoff traffic without competing with the background traffic to be served. Thus, c_n is given by

$$c_i = \begin{cases} \lambda_{BP} + n\lambda_P, & \text{for FIFO} \\ n\lambda_P, & \text{for Priority,} \end{cases} \tag{5}$$

The d is the service rate for the handoff connection. Note that, after the FR_{old} receives the *BIND* message, the astray packets from the old stream are directly entering into the system buffer instead of the temporary buffer. At this time, we assume that the handoff connection begins with the bulk arrival. The handoff connection continues

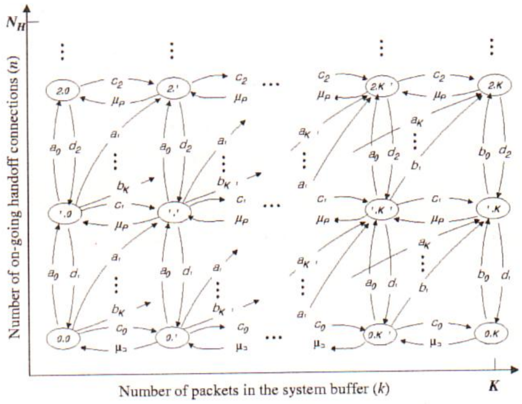


Figure 5. State transition diagram for the system buffer at the FR_{old}.

to generate traffic (i.e., astray packets) until the last packet from the old stream reaches to the FR_{old}. Thus, it is clear that holding time of the handoff connection, t_{hold}, is $(t_{cn}+t_{co}-t_{on})$. The handoff process is just an occupation process in a $M/M/\infty$ queueing system with the arrival rate λ_H and the service rate $1/t_{hold}$. Thus, the service rate of handoff connection d_n equals to n/t_{hold}.

The probability distribution of steady state, $\pi(n,k)$, can be obtained by matrix algebra [9], since the above Markov chain is irreducible and aperiodic. From Little's results [9], the average queuing delay of the packets obtained as $\overline{t_Q} = \dfrac{\bar{k}}{\lambda_{offered}(1-P_B)}$, where the average queue size $\bar{k} = \sum_{n=0}^{N_H}\sum_{k=0}^{K} k\pi(n,k)$, the offered load $\lambda_{offered} = \sum_{n=0}^{N_H}\sum_{k=0}^{K}(c_n + \sum_{i=0}^{i=\infty} ia_i)\pi(n,k)$, and the loss probability,

$$P_B = \frac{\sum_{n=0}^{N_H}\pi(n,K)c_n + \sum_{n=0}^{N_H}\sum_{k=0}^{K}\pi(n,k)\alpha(k)}{\lambda_{offered}},$$

where

$$\alpha(k) = \sum_{i=K-k+1}^{\infty}(i-(K-k))a_i.$$

IV. Numerical Results

In this section, we evaluate the transient time periods of MN under the smooth handoff in the Mobile IPv6 networks. To evaluate the mobile network environment, we refer to the parameters of [5] and [8], especially for the link delays. Basic parameters listed in Table 1 are used unless specified else and simulation environment is depicted in Fig 6.

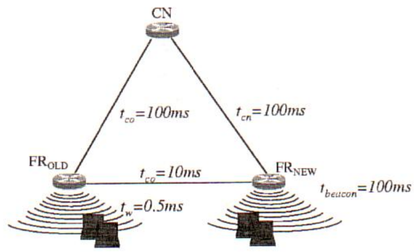


Figure 6. Simulation environment.

Fig. 7(a), (b), and (c) depict the UTP, the STP, and the HTP variation according to the variation of λ_{BP}, λ_H, and t_{beacon}, respectively. Fig. 7 (a) shows that the UTP increases exponentially as the background traffic rate increases. The UTP of priority method merely keeps a certain delay level that is less than 15*ms*. Fig. 7 (b) depicts that the STP increases as the handoff rate or t_{cn} increase. It is also observed that the STP of the priority scheme is significantly less than that of the FIFO method. Fig. 7 (c) represents that the HTP increases somewhat linearly with respect to the t_{beacon}, since the t_{beacon}, is much bigger than the queuing delay and the link delay.

V. Conclusion

We analyzed the impact of handoff on the packet sequence under the Mobile IPv6 with the route optimization extension. The impact, measured by the UTP, STP, and HTP, gives the estimation on several transient time periods after the handoff initiation. Through the analysis, it is observed that the UTP is a function of traffic load, t_{on}, t_{beacon},, and λ_H. The prioritized queuing scheme can effectively alleviate the delay effect caused by the

Table 1. System parameters for numerical analysis.

Parameters	Description	Values
K	Maximum system buffer size of the FR_{old}	200
N_H	Maximum number of handoff connections to be hold at the FR_{old}	50
μ_P	Packet service rate at the FR_{old}	1000 packets/s
λ_P	Packet generation rate of connection	50 packets/s
λ_H	Handoff rate of node	1-20 handoffs/s
t_{beacon}	Period of broadcasting the beacon message	10-500ms
t_{on}	Packet travel time along the link between the FR_{old} and the FR_{new}	1-10ms
t_w	Packet travel time over the wireless link at the FR_{new}	0.5ms
t_{co}	Packet travel time along the link between the corresponding node and the FR_{old}	50-200ms
t_{cn}	Packet travel time along the link between the corresponding node and the FR_{new}	50-200ms
λ_{BP}	Packet generation rate of background traffic at the FR_{old}	100-900packets/s

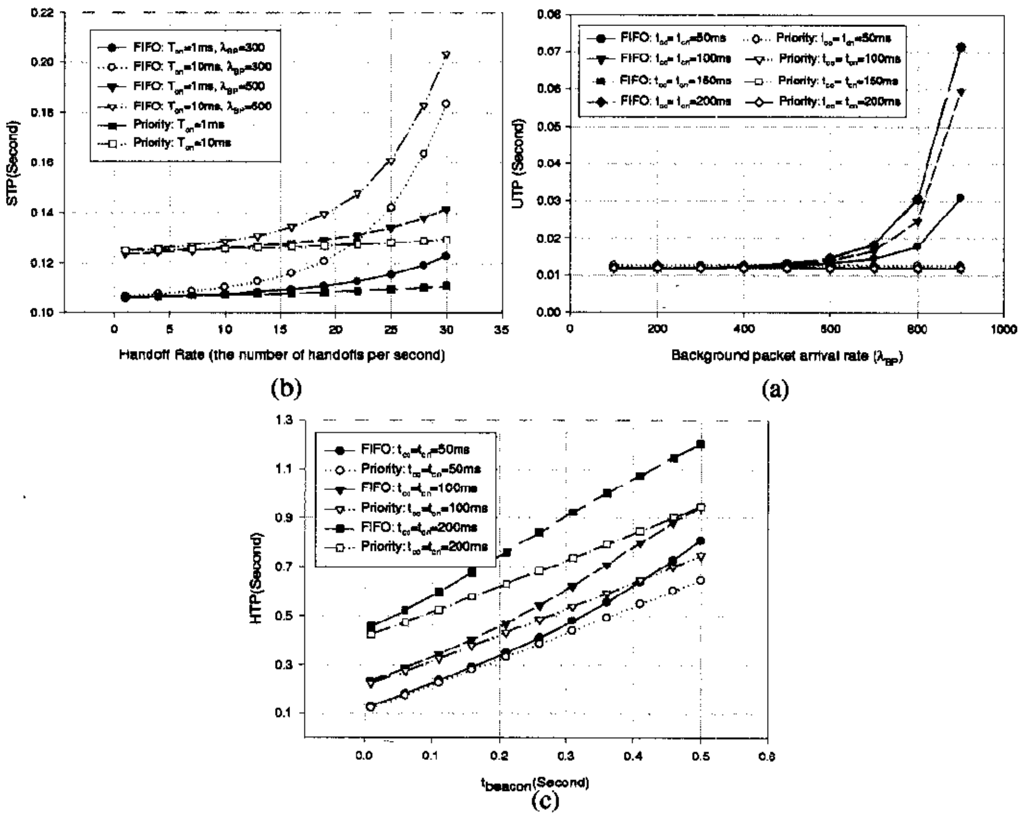


Figure 7. (a)UTP vs. the background traffic, (b)STP vs. the handoff rate, and (c) HTP vs. the t_{beacon}.

background traffic load. Currently we are working to investigate the overall impact of the transient time on TCP and streaming media.

References

[1] C. Perkins, "IP Mobility Support in IPv4," *Internet Engineering Task Force*, RFC 3344, Aug. 2002.

[2] C. Perkins and D. Johnson, "Route optimization in Mobile IP," *Internet Engineering Task Force*, Internet Draft, draft-ietf-mobileip-optim-11.txt, Sept. 2001.

[3] D. Johnson and C. Perkins, "Mobility Support in IPv6," *Internet Engineering Task Force*, Internet Draft, draft-ietf-mobileip- ipv12.txt, Mar. 2002.

[4] D. Guo, A. S. Acampora, and Z. Zhang, "Using

refuge proxy to prevent packet loss/out-of-sequence and path loop in micro-cell wireless networks," in *Proc. of IEEE GLOBECOM '97*, pp. 1338-1343, Nov. 1997.

[5] D. Eom, M. Sugano, M. Murata, and H. Miyahara, "Performance improvement by packet buffering in Mobile IP based networks," *IEICE Trans. Commun.* vol. E83-B, no. 11, pp. 2501-2512, Nov. 2000.

[6] D. Lee, G. Hwang, and C. Oh, "Performance evaluation of handoff procedure in Mobile IP networks," In *Proc. of APCC '2001*. pp. 513-516, Nov. 2001.

[7] C. Chiasserini and R. L. Cigno, "Handovers in wireless ATM networks: In-band signa ling protocols and performance analysis,'' *IEEE Trans. Commun.*. vol. 1, no. 1, pp. 87-100, Jan. 2002.

[8] P. Manzoni and D. Ghosal, "Impact of mobility on TCP/IP: An integrated performance study," *IEEE J. Select. Areas Commun.*, vol. 13, no. 5, pp. 858-867, June 1995.

[9] G. Bolch, S. Greiner, H. Meer, and K. S. Trivedi, *Queueing networks and Markov Chains*, John Wiley & Sons, 1998.

이 동 욱 (Dong-Wook Lee) 정회원

[redacted]

1997년 2월 : 충북대학교 컴퓨터공학과 학사

1999년 2월 : 광주과학기술원 정보통신공학과 석사

1999년 3월~현재 : 광주과학기술원 정보통신공학과 박사과정

〈주관심분야〉 Mobile Network, Network-adaptive Media Delivery, Performance Measurement, Multicast, and End-to-end QoS

김 종 원 (Jong-Won Kim) 정회원

[redacted]

1987년 2월: 서울대학교 제어계측공학과 학사

1989년 2월: 서울대학교 제어계측공학과 석사

1994년 2월: 서울대학교 제어계측공학과 박사

2001년 9월~현재: 광주과학기술원 정보통신공학과 부교수

2000년 7월~2001년 6월: 미국 InterVideo Inc., Fremont, CA, 개발자문

1998년 12월~2001년 7월: 미국 Univ. of Southern California, Los Angeles, CA, EE-Systems Department 연구조교수

1994년 3월~1999년 7월: 공주대학교 전자공학과 조교수

〈주관심분야〉 Networked Media Systems and Protocols focusing "Reliable and Flexible Delivery for Integrated Media over Wired/Wirelss Networks

(네트워크미디어:http://netmedia.kjist.ac.kr)