Handover Management based on the Number of Frame Retransmissions for TCP over WLANs

Kazuya Tsukamoto[†], Shigeru Kashihara[††], and Yuji Oie[†††]
[†]Japan Society for the Promotion of Science (JSPS) Postdoctoral (PD) Research Fellow.
[†] [†††]Department of Computer Science and Electronics, Kyushu Institute of Technology.
680-4, Kawazu, Iizuka, 820–8502 Japan, TEL:+81-948-29-7687, FAX:+81-948-29-7652
[††]Graduate School of Information Science, Nara Institute of Science and Technology.
8916-5, Takayama-cho, Ikoma, 630-0192 Japan, TEL:+81-743-72-5213, FAX:+81-743-72-5219
Email: [†]kazuya@infonet.cse.kyutech.ac.jp, [††]shigeru@is.naist.jp, [†††]oie@cse.kyutech.ac.jp

***Abstract*— In ubiquitous networks based on Wireless Local Area Networks (WLANs), mobile nodes (MNs) will experience performance degradation at the handover. To achieve seamless and efficient communication, it is necessary for the MNs to satisfy the following three requirements: (I) initiation of handover processes based upon quick perception of the change in the wireless link quality, (II) elimination of a communication interruption due to handover processes, (III) selection of an optimal WLAN. In this paper, we propose a unified handover management scheme for TCP communication that can satisfy all of the above three requirements. Our proposed scheme employs both a cross-layer approach for obtaining the number of frame retransmissions to satisfy (I) and (III) and multi-homing to satisfy (II). Through simulation experiments, we demonstrate that our proposed scheme can quickly and reliably perceive the deterioration of the wireless link quality and select the optimal WLAN without communication interruption due to the handover process, thereby maintaining the best TCP performance at the handover.**

***Index Terms*— Seamless Handover, Cross-Layer, Multi-Homing, Frame Retransmission, TCP**

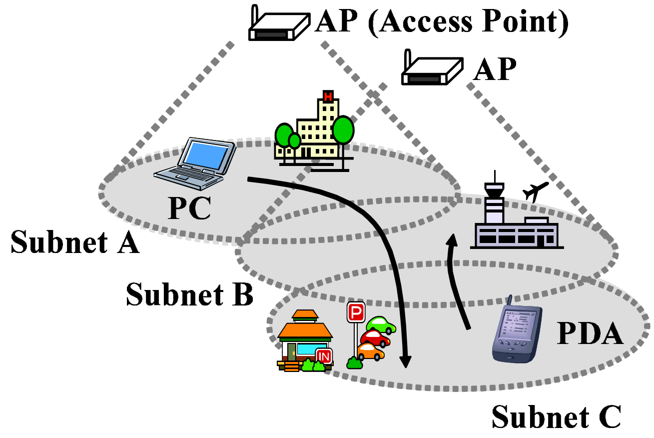

Fig. 1. Future ubiquitous mobile network based on WLANs.

I. INTRODUCTION

Wireless Local Area Networks (WLANs) based on IEEE 802.11 [1] have gained popularity due to their low cost, ease of installation, and broadband connectivity. WLANs are starting to cover not one spot but a wide area, such as a city, by using multiple access points (APs). Actually, many of these deployments have been progressing around the world [2] [3] [4]. In the near future, WLANs will continue to spread until they overlap to provide continuous coverage over a wide area, and they will be the underlying foundation of ubiquitous networks.

In such ubiquitous networks, mobile nodes (MNs) will be much more likely to traverse different WLANs, which are WLANs independently managed by different companies or organizations (different IP subnets), during TCP communication due to a WLAN's small coverage, as shown in Fig. 1. Therefore, an effective handover management scheme for achieving seamless and efficient communication at handovers is crucial.

The two most critical issues at handover are the potential changes in the IP address of an MN and degradation of the communication quality due to handover processes and radio characteristics. In the former issue, when an MN moves between WLANs consisting of different IP subnets, the IP address of the MN is changed, and the TCP communication is terminated. Many mobility management schemes such as Mobile IP (MIP) [5] and mobile Stream Control Transmission Protocol (mSCTP) [6] have been proposed to solve this issue. An MN employing the existing schemes can maintain the TCP connection at the handover between different WLANs.

However, in the later issue, communication interruption for several seconds by handover processes and the degradation of communication quality due to radio characteristics cannot be inherently avoided, regardless of the mobility management scheme used. The following three requirements are essential for achieving seamless and efficient communication: (I) initiation of handover processes based upon quick perception of the change in the wireless link quality, (II) elimination of a communication interruption due to handover processes, (III) selection of an optimal WLAN. Although none of the existing schemes can satisfy all the above requirements, we proposed a mobility management scheme considering (II) and (III) in [7]. Furthermore, in our previous work [8], we focused on a handover decision criterion that satisfied (I) and we showed not only that TCP performance under the existing decision criteria can be decreased drastically before the handover, but also the number of frame retransmissions obtained from the MAC layer (Layer 2) has the potential to serve as the handover decision criterion for avoiding performance degradation before the handover.

In this paper, we identify the issues of handover in more detail and then review the effectiveness of frame retransmission as a new handover decision criterion. After that, to satisfy all three requirements, we propose a unified handover management scheme that integrates our previous work for (I) [8] with that for (II and III) [7]. Finally, through simulation experiments, we show that the TCP performance degradation at a handover can be avoided by utilizing our proposed scheme.

II. ISSUES OF WLAN HANDOVER

When an MN executes the handover between WLANs consisting of different IP subnets, the following two issues should be solved:

- Termination of a connection by a change of IP address
- Performance degradation due to handover processes and deterioration of the wireless link quality

MIP has been proposed to solve the first issue. Although an MN can maintain the TCP communication at a handover under MIP, the communication quality decreases due to handover processes and reduction of the wireless link quality. More specifically, when the MN traverses between WLANs managed by different IP subnets, the MN has to execute handover processes consisting of the following five steps:

1. Detecting the MN's movement by the loss of Router Advertisement packets (beginning of the handover process)
2. Scanning for newly available APs
3. Establishing an association with the new AP
4. Updating the IP address binding by the Dynamic Host Configuration Protocol (DHCP)
5. Sending a Binding Update (BU) packet to both the Home Agent (HA) and Corresponding Node (CN)

In the MIP network, each MN detects its own movement by utilizing Router Advertisement packets, which are broadcast infrequently from an AP (The default interval is 1 second [9]). However, because its infrequency results in an increase of the handover decision latency, TCP goodput decreases drastically at the starting point of the handover. After that, the MN has to execute handover processes, which can be divided into two main parts. First, the link layer handover process (steps 2, 3 above) takes from approximately 50 ms to 400 ms, depending on the hardware used [10]. Next, the IP layer handover process (steps 4, 5) consists of both reconfiguration of the IP address using the DHCP mechanism (300 ms [11]) and sending a Binding Update (one-way delay). Considering the above discussion, the period for handover processes is more than 1 second. Because the MN cannot send or receive packets during this period, the performance of TCP communication degrades drastically.

Some enhanced protocols of MIP, such as Fast Handover Mobile IP (FMIP) [12], have been proposed to reduce the Link layer and IP layer handover processing period. However, because these schemes must deploy and manage special equipments such as the HA, it is extremely difficult for them to function in the current Internet, which is managed by different companies/organizations.

III. HANDOVER DECISION CRITERION

In our previous work [8], we showed that the number of frame retransmissions has the potential to serve as a new handover decision criterion. In this section, we review the effectiveness of the number of frame retransmissions.

A. Frame retransmission and packet loss

Frame retransmission occurs for the following two reasons: (i) reduction of signal strength and (ii) collision with other frames. In a WLAN, a sender can detect successful transmission by receiving an ACK frame in response to a transmitted data frame, that is, in the stop-and-wait manner. Therefore, when a data or an ACK frame is lost, the sender retransmits the same data frame until the number of frame retransmissions reaches a predetermined limit. Note that, with Request-to-Send (RTS)/Clear-To-Send (CTS), collisions between data frames, namely, a hidden terminal problem, never occur due to the exchange of the RTS/CTS frames. If the RTS/CTS mechanism is applied, the retransmission limit is set to 4: a data frame can be retransmitted a maximum of four times (the initial transmission and three retransmissions), if necessary. If the sender does not receive an ACK frame within the retransmission limit, the data frame is treated as a lost packet and the TCP sender finally retransmits the same packet by its retransmission control of TCP. Therefore, because data frames are inherently retransmitted before being treated as a lost packet, the number of frame retransmissions allows the MN to perceive the deterioration of the condition of a wireless link, and may enable the MN to determine when the handover process should be started before packet loss actually occurs.

B. TCP goodput and the number of frame retransmissions

We demonstrate the effectiveness of the number of frame retransmissions through simulation experiments using ns-2 (ver 2.27) [13]. In our simulations, an MN in a WLAN of only 11 Mb/s establishes a TCP (NewReno) connection with a Corresponding Node (CN) for file transfer communication with packets of 1500 bytes. Furthermore, we employ the RTS/CTS mechanism due to the large packet size. We investigate how the distance between the MN and AP affects both the number of frame retransmissions and the TCP goodput. The MN can obtain the number of frame retransmissions by using an ACK packet of the TCP flow transmitted from itself. Note that a TCP ACK packet is a data frame in a WLAN.

Figure 2 shows the change in the TCP goodput, and Fig. 3 shows how many retransmissions each frame experiences. The legend "Retransmission: *n*" in Fig. 3 indicates the ratio of frames suffering *n* retransmissions of all frames. Note that the "Retransmission:0" indicates the ratio of the successful frames without any retransmissions. On the other hand, "Packet

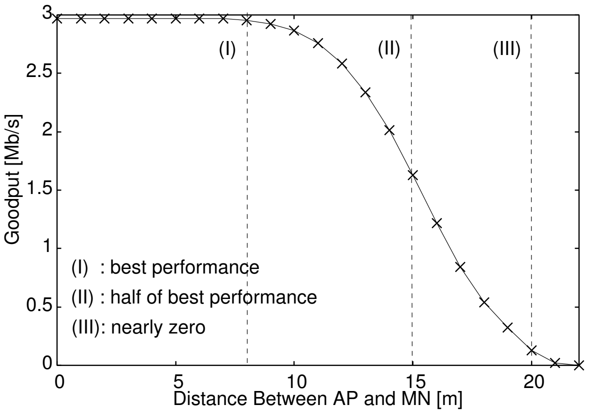

Fig. 2. Goodput.

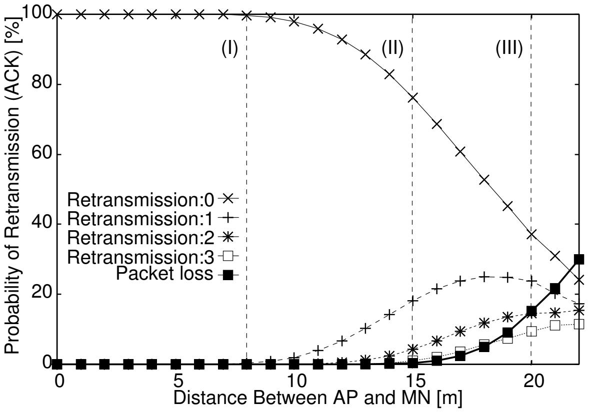

Fig. 3. Frame retransmission ratio.

loss” indicates the ratio of packet loss occurring after the fourth retransmission fails. As shown in Fig. 2 and 3, frame retransmissions begin to occur at around 8 m, and the TCP goodput also begins to decrease soon after the occurrence of the frame retransmissions. This result shows that degradation of the TCP goodput begins even when a frame retransmission occurs at least once. Therefore, we suggest that the degradation of the TCP goodput could effectively be avoided by exploiting the frame retransmissions when the MN executes a handover.

IV. PROPOSED HANDOVER MANAGEMENT SCHEME

We propose unified handover management to avoid the performance degradation at the handover. The handover management should satisfy the following three requirements: (I) initiation of handover processes based upon quick perception of the change in the wireless link quality, (II) elimination of a communication interruption due to the handover processes, and (III) selection of an optimal WLAN.

A. Initiation of handover processes based upon quick perception of the change in wireless link quality

As mentioned in Sec. III, our proposed scheme considers the number of frame retransmissions as the handover decision criterion. However, the information held in each layer cannot be accessed from different layers due to the concept of the traditional layered architecture. So, in this paper, we suppose that the benefit of introduction of cross-layer approach [14] is greater than its cost paid for benefits [15], and thus employ the cross-layer approach to achieve the interaction between these layers. Figure 4 illustrates our concept of the handover management mechanism. As illustrated in Fig. 4, in our proposed scheme, the Handover Manager (HM) at the Transport layer perceives the deterioration of the wireless link quality based on the number of frame retransmissions obtained from the MAC layer. Note that our proposed scheme can be applied only to both end-to-end hosts.

B. Elimination of communication interruption due to handover processes

As mentioned in Sec. II, the handover processing period, in which an MN cannot send or receive packets, cannot be avoided when the MN traverses WLANs with different IP subnets. Our proposed scheme allows the MNs to traverse multiple WLAN interfaces, i.e., to support multi-homing. As a result, the multi-homing MN can eliminate the communication interruption due to the handover process by establishing the connection with a new AP before the degradation of the condition of the current AP begins. For example, when an MN, communicating with a CN via one WLAN interface (IF1), finds a new AP, it starts to establish a new connection to the AP via the WLAN interface (IF2) in advance. Therefore, the MN never experiences the interruption period during handover.

C. Selection of an Optimal WLAN

Figure 5 shows how the optimal WLAN is selected during the handover period. When the number of frame retransmissions of the current interface in the single-path transmission mode exceeds the predetermined threshold (Ret_Thr), the MN detects the deterioration of the wireless link quality and starts the handover process, as mentioned in Sec. IV-A. After that, the MN switches to the multi-path transmission mode, then starts a parallel transfer by utilizing all the available WLANs, and finally selects the optimal WLAN among them. Our proposed scheme employs the number of frame retransmissions obtained from each available WLAN as the criterion for selecting the optimal WLAN. In this paper, the number of retransmissions is measured by only one packet transmitted from each WLAN interface for parallel transfer, and the HM updates the parameters Ret_IF1/Ret_IF2 (Fig. 4). Upon a comparison of these parameters, the MN selects the WLAN with the smallest number of retransmissions as the optimal WLAN and returns to single-path transmission mode. In this way, our proposed scheme can execute the handover considering the condition of all available WLAN interfaces. Furthermore, because the MN selects the optimal WLAN based on only one packet transmitted from each WLAN in the multi-path transmission mode, the network load due to parallel transfer can be extremely limited.

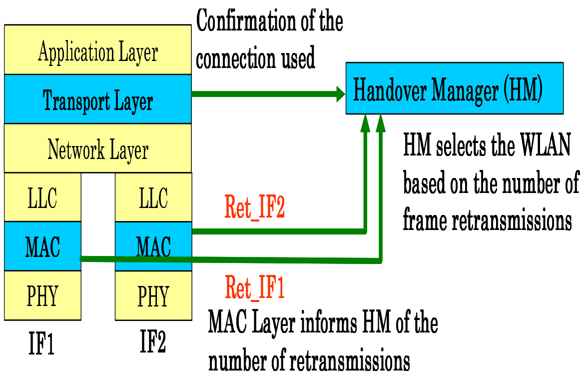

Fig. 4. Cross-layer architecture.

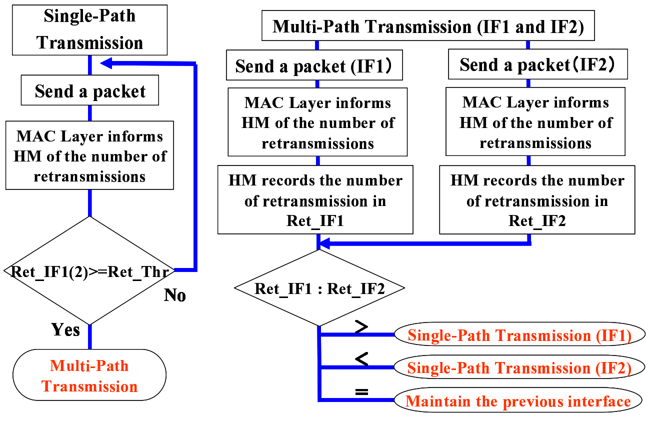

Fig. 5. Handover management procedures.

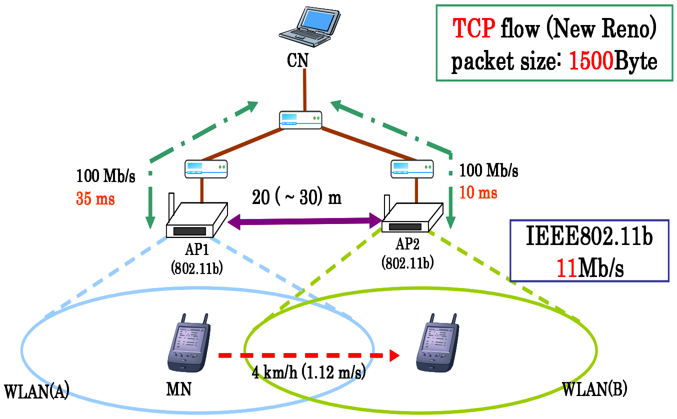

Fig. 6. Simulation model.

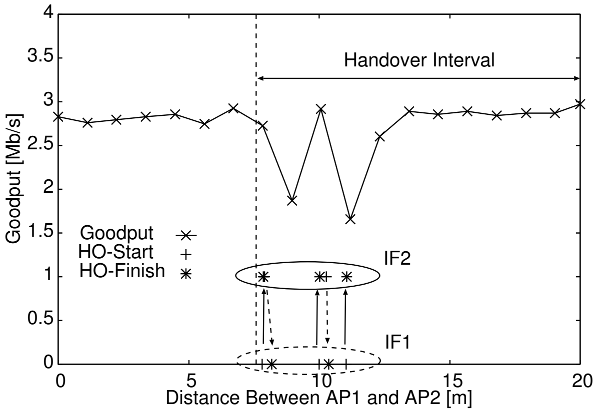

Fig. 7. Goodput (Ret_Thr = 1).

V. SIMULATION

We now show how our proposed scheme can maintain the TCP goodput when the MN executes a handover.

A. Simulation model

We employ a realistic model to evaluate the effect of an MN's movement from WLAN(A) to WLAN(B), as shown in Fig. 6. The MN first establishes a TCP (NewReno) connection with a CN via WLAN(A) for file transfer communication, with packets of 1500 bytes. Simulations are conducted for a period of 60 s, in which the MN located just under AP1 starts to move toward the AP2 of WLAN(B) at 35 s. The MN is moving at a walking speed of 4 km/h. The one-way delay between the CN and the MN is different in each WLAN, because we assume that each WLAN consists of different IP subnets: The delay via WLAN(A) is 35 ms and that via WLAN(B) is 10 ms. Through the ns-2 simulations [13], we evaluate how our proposed scheme can avoid the degradation of TCP performance at a handover when the retransmission threshold (Ret_Thr) and the distance between APs are varied.

B. Simulation results

We present the simulation results for two cases. In the first case, we examine the effect of the frame retransmission threshold (Ret_Thr). In the second case, we examine how the distance between two APs affects the TCP goodput.

1) Effect of frame retransmission threshold (Ret_Thr): We examine the TCP goodput when the MN located under AP1 moves toward AP2, as shown in Fig. 6. Note that the distance between the two APs is fixed at 20 m. Figures 7-10 shows how the TCP goodput varies with the value of Ret_Thr, which is set from 1 to 4. "HO-Start" in each figure indicates the time when the MN starts the parallel transfer, and is where the MN perceives the deterioration of the wireless link condition. Furthermore, "HO-Finish" indicates the time when the MN finally switches to the optimal WLAN. These two points are placed in either an IF1 or IF2 area in each figure. The area denotes the interface handling these processes. For example, the "HO-Start" placed in the IF1 area in the figures indicates that the MN perceives the deterioration of the WLAN(A)'s condition (IF1's condition) and starts the parallel transfer. On the other hand, the "HO-Finish" placed in the IF2 area indicates that the MN selects the IF2 (WLAN(B)) as the optimal interface. Each arrow indicates the transition in the optimal WLAN interface between IF1 and IF2. Note that in the initial state, IF1 is treated as the optimal WLAN.

In Fig. 7, the MN begins to execute the handover at around

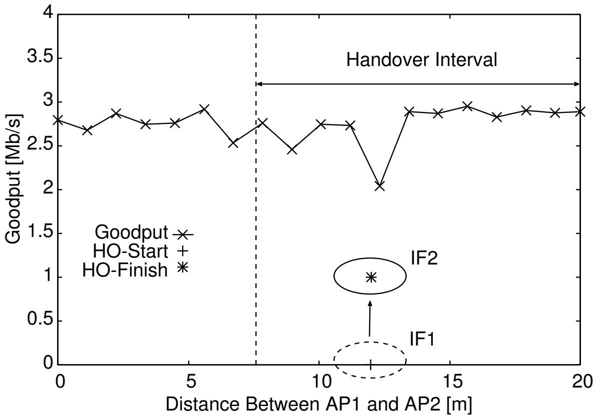

Fig. 8. Goodput (Ret_Thr = 2).

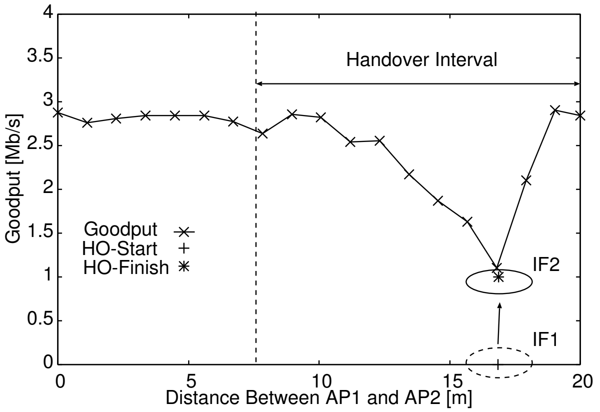

Fig. 9. Goodput (Ret_Thr = 3).

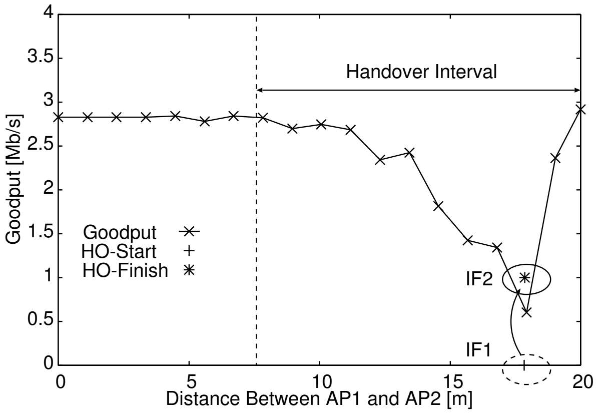

Fig. 10. Goodput (Ret_Thr = 4).

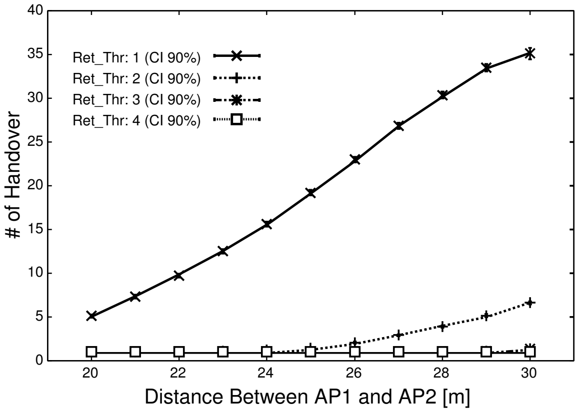

Fig. 11. Number of handover occurrences.

8 m, which is the distance from AP1, when Ret_Thr is set to 1. As shown in Fig. 2, the TCP goodput begins to degrade drastically just after 10 m. That is, the MN can start the handover process before the degradation of TCP performance actually occurs. Then, although the MN first selects WLAN(B) as the optimal WLAN, the optimal WLAN moves back and forth between these two WLANs several times until the condition of WLAN(B) becomes stable. As a result, the goodput performance degrades during this unstable period. Beyond 11 m, the communication becomes stable without the degradation of TCP goodput and the MN selects WLAN(B). This means that the handover from AP1 to AP2 is completed. On the other hand, in Fig. 10, the MN begins to execute the handover beyond 17 m, when the Ret_Thr is set to 4. Therefore, the goodput decreases to a very low value (0.5 Mb/s) due to the large latency of the handover decision. Note that, when Ret_Thr is set to 3, the goodput also decreases for the same reason (Fig. 9).

From these results, we can see that the MN can quickly perceive the deterioration of the wireless link condition with a small Ret_Thr, i.e., 1. In this case, multiple handovers occur because the handovers start too early and the condition of WLAN(A) has not yet deteriorated. On the contrary, when Ret_Thr is set to a relatively large value, handovers do not easily occur. In this case, the proposed scheme cannot avoid the performance degradation because the proposed scheme begins to execute the handover after the drastic deterioration of the wireless link quality. From these results, we can see that Ret_Thr strongly affects the TCP goodput, so that Ret_Thr should be determined carefully.

From Fig. 8, we can see that the MN begins to execute the handover at around 12 m, when Ret_Thr is set to 2. The MN can quickly perceive the deterioration of the wireless link condition and can appropriately select the optimal WLAN without multiple changes. As a result, the proposed scheme can maintain the excellent goodput even at handover.

2) Effect of change in the distance between APs: Next, we focus on how the number of handovers affects the goodput performance at the handover. The number of handovers depends on the distance between two APs. So far, the distance between AP1 and AP2 is fixed to 20 m. In this case, because the condition of WLAN(B) begins to be stable where that of WLAN(A) begins to be unstable, the number of handovers does not increase drastically. When the distance between APs is lengthened, the number of handovers increases and then

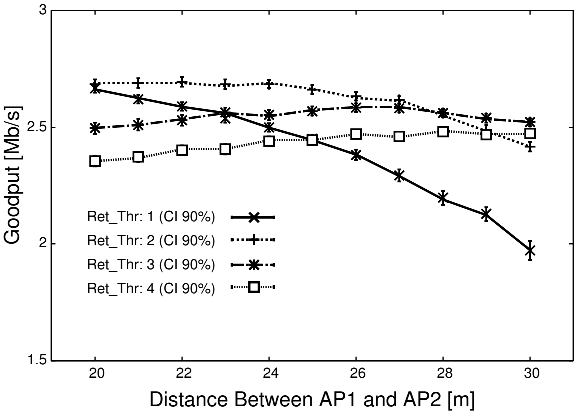

Fig. 12. Average goodput.

goodput decreases. Therefore, in this simulation, we focus on the relation between the number of handovers and the goodput of our proposed scheme when the distance between APs varies from 20 m to 30 m. Furthermore, in order to show the dependency between an appropriate Ret_Thr and the distance between APs, we examine the change in the number of handovers and TCP goodput when the value of Ret_Thr is set from 1 to 4. Figure 11 shows how often the handover occurs, and Fig. 12 shows the change in the goodput. Note that the 90% confidence interval (CI) of each value in Fig. 11 and Fig. 12 is examined closely to clarify the impact of the proposed scheme. With the increase in the distance between APs, the number of handovers increases up to around 35 times, and the goodput also decreases drastically when Ret_Thr $= 1$. Similarly, when Ret_Thr is set to 2, the number of handovers begins to increase and the goodput begins to decrease beyond 28 m. In contrast, when Ret_Thr is set to 3, our proposed scheme can select the optimal WLAN with only one time handover for any distance and can maintain the goodput, which is greater than that of Ret_Thr $= 2$.

From these results, we demonstrate that an appropriate Ret_Thr depends upon the distance between APs. Furthermore, if Ret_Thr is set to an appropriate value, our proposed scheme can select the optimal WLAN with a small number of handovers and can avoid the drastic degradation of the goodput at handover for any distance between APs.

VI. CONCLUSION AND FUTURE WORK

In this paper, we proposed the unified handover management scheme based on the number of frame retransmissions. The proposed scheme can satisfy the following three requirements in order to avoid the performance degradation at a handover: (I) initiation of handover processes based on quick perception of the change in the wireless link quality, (II) elimination of the communication interruption due to handover processes, (III) selection of an optimal WLAN. In our proposed scheme, the MN starts handover processes based on the number of frame retransmissions to perceive the deterioration of the wireless link quickly. Then, a communication interruption due to handover processes can be eliminated by utilizing multihoming. Finally, our proposed scheme can select the optimal WLAN through consideration of the condition of all available WLANs by allowing the MN to transmit only one data packet over each WLAN simultaneously.

Simulation results showed that the goodput deteriorates for two reasons: (a) an increase of the number of handovers with a small Ret_Thr, and (b) the large perception latency of the deterioration of the wireless link condition with a large Ret_Thr. Finally, we have shown that the Ret_Thr should be set to an appropriate value to achieve seamless and efficient communication during the handover.

In our future work, we plan to propose dynamic decision algorithms of the handover threshold to avoid the performance degradation at any distance between APs. Moreover, we are carrying out the implementation of our proposed scheme in the Linux kernel. After the completion of that work, we plan to report the experimental results.

ACKNOWLEDGMENT

This work was supported in part by a grant from the Cisco University Research Program Fund at Community Foundation Silicon Valley, in part by the Japan Society for the Promotion of Science, Grant-in-Aid for JSPS Fellows (17-6551) and Scientific Research (S)(18100001), and in part by the Ministry of Public Management, Home Affairs, Posts and Telecommunications, Japan.

REFERENCES

[1] IEEE 802.11, 1999 Edition, http://standards.ieee.org/getieee802/download/802.11-1999.pdf
[2] Wireless Philadelphia, http://www.phila.gov/wireless/index.html/
[3] Wireless London, http://wirelesslondon.info/HomePage
[4] WIFLY (Taipei), http://www.wifly.com.tw
[5] C. Perkins, "IP Mobility Support for IPv4," IETF RFC3344, Aug. 2002.
[6] S. J. Koh, et al., "Mobile SCTP for Transport Layer Mobility," *draft-reigel-sjkoh-sctp-mobility-04.txt*, Internet draft, IETF, Jun. 2004.
[7] K. Tsukamoto, et al., "Mobility Management of Transport Protocol Supporting Multiple Connections," *Proc of ACM MobiWac2004*, pp. 83-87, Oct. 2004.
[8] K. Tsukamoto, et al., "Impact of Layer 2 Behavior on TCP Performance in WLAN," *Proc. of IEEE VTC2005 fall*, Sep. 2005.
[9] D. Johnson, et al., "Mobility Support in IPv6," IETF RFC3775, Jun. 2004.
[10] A.Mishra, et al., "An Empirical Analysis of the IEEE802.11 MAC Layer Handoff Process," *Proc. of ACM SIGCOMM Comp. Commu. Rev.*, Vol.33, Issue 2, Apr. 2003.
[11] A. Dutta, et al., "Application Layer Mobility Management Scheme for Wireless Internet," Proc. of IEEE 3G Wireless, 2001
[12] P. McCann, et al., "Mobile IPv6 Fast Handovers for 802.11 Networks," IETF RFC4260, Nov. 2005.
[13] The Network Simulator 2 ns-2, http://www.isi.edu/nsnam/ns/
[14] S. Shakkottai et al., "Cross-Layer Design for Wireless Networks," *IEEE Commu. Mag.*, Vol. 41, No. 10, pp. 74-80, Oct. 2003.
[15] Vikas Kawadia, et al., "A Cautionary Perspective on Cross-Layer Design," *IEEE Wireless Communications*, Vol. 12, No. 1, pp. 3-11, Feb. 2005.