PAPER

A Unified Handover Management Scheme Based on Frame Retransmissions for TCP over WLANs

Kazuya TSUKAMOTO[†a)], **Shigeru KASHIHARA**[††], *Members*, *and* **Yuji OIE**[†], *Fellow*

SUMMARY In ubiquitous networks based on Wireless Local Area Networks (WLANs) with limited individual coverage, mobile nodes will be likely to traverse different WLANs during TCP communication. An effective handover management scheme for achieving seamless and efficient communication throughout the handover operation is therefore crucial. To achieve this, the following three requirements are essential: (i) early initiation of handover, (ii) elimination of communication interruption upon handover, (iii) selection of an optimal WLAN. The handover scheme proposed in this study employs frame retransmission over WLAN as an indicator of link degradation, and a handover manager (HM) on the transport layer obtains the number of frame retransmissions on the MAC layer using a cross-layer architecture in order to achieve (i) and (iii). Then, it also employs multi-homing in order to achieve (ii). Simulations demonstrate that the proposed scheme can satisfy all of the three requirements and is capable of maintaining TCP performance throughout the handover operation.

key words: *wireless LAN, handover, TCP, frame retransmission, cross-layer, multi-homing*

1. Introduction

With the proliferation in the number of mobile Internet users, a diverse range of wireless access network technologies, including cellular, wireless local area networks (WLANs), Bluetooth, and WiMAX, have emerged and have been developed to make ubiquitous Internet access a reality. WLANs based on the IEEE 802.11 (a/b/g) specification family [1] have gained popularity as being low-cost solutions that are easy to install and provide broadband connectivity, and such networks are being widely deployed in both private spaces (e.g., homes and workplaces) and as hot spots in public spaces (e.g., waiting areas and hotel lobbies). Internet access is currently provided independently by a large number of competing organizations or Internet service providers (ISPs). WLANs deployed by different companies are starting to overlap, not only at individual sites but over wide areas such as an entire city through the installation of multiple access points (APs). This situation is common to many metropolitan areas around the world [2]–[4], and WLANs can be expected to continue to spread until continuous coverage over a wide area is achieved through extensive overlap. Such a structure will fulfill an important part of the ubiquitous network concept: “ubiquitous WLANs.”

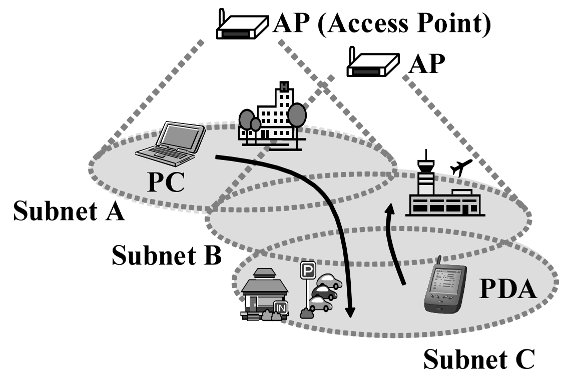

Fig. 1 Future ubiquitous mobile network based on WLANs (ubiquitous WLANs).

In ubiquitous WLANs, mobile nodes (MNs) can access the Internet through any available AP at any location. However, as MNs are very likely to move during transmission control protocol (TCP) communication, many handovers between WLANs with different Internet protocol (IP) subnets will need to be handled due to the small coverage of individual WLANs (Fig. 1). Therefore, an effective handover management scheme for achieving seamless communication is essential in order to provide transparent mobility for MNs in a ubiquitous WLANs environment.

The most critical issues at handover are potential changes in the IP address of the MN, and degradation of communication quality. When an MN moves between WLANs with different IP subnets, the IP address of the MN is changed. As a result, the TCP connection under the traditional Internet architecture is terminated by the handover. A number of existing handover management schemes, such as the Mobile IP (MIP) [5], [6] and the mobile stream control transmission protocol (mSCTP) [7], have already been proposed to resolve the problem associated with the change in IP address. An MN employing such a scheme can maintain a TCP connection upon handover between different WLANs. However, communication interruption cannot be avoided as an inherent function of Layer 2 and 3 handover processes, even when any existing mobility management schemes are employed [8], [9]. Further degradation of communication quality is also possible due to effects such as weaker signal strength and radio interference. Existing schemes are unable to address the problem of interruption and degradation of communication quality upon handover. The requirements

Manuscript received May 24, 2007.
Manuscript revised October 20, 2007.

[†]The authors are with the Dept. of Computer Science and Electronics, Kyushu Institute of Technology, Iizuka-shi, 820-8502 Japan.
[††]The author is with the Graduate School of Information Science, Nara Institute of Science and Technology (NAIST), Ikoma-shi, 630-0192 Japan.
a) E-mail: kazuya@infonet.cse.kyutech.ac.jp

DOI: 10.1093/ietcom/e91–b.4.1034

of any system implemented to prevent degradation of communication quality and thus achieve seamless handover are as follows:

(i) Handover must be initiated based on quick detection of changes in wireless link quality
(ii) Communication interruption due to handover must be eliminated
(iii) The optimal WLAN at any location must be selected

Although the handover management scheme in Ref. [10] considers points (ii) and (iii), and the enhanced scheme in Ref. [11] considers point (i), none of the schemes presented to date satisfy all the above requirements in one model. Note that the scheme that can satisfy all of the requirements is referred to in this paper as the unified handover management scheme. In particular, the communication performance immediately prior to a handover in the enhanced schemes considering responsive quality change detection has yet to be examined in detail.

In the present study, the issues arising due to WLAN handover are identified and a handover decision criterion satisfying point (i) above is developed. The number of frame retransmissions obtained from the media access control (MAC) layer (Layer 2) is employed as a new handover decision criterion for avoiding performance degradation before the handover. Through both simulation and experimental evaluation, the problems of existing decision criteria are highlighted and the effectiveness of frame retransmission as a handover decision criterion is demonstrated. A unified handover management scheme that integrates the handover decision criterion satisfying (i) with the scheme of Ref. [10] satisfying (ii) and (iii) is then proposed as a complete model, and simulations are performed to demonstrate how the degradation of TCP performance upon handover can be avoided under the new scheme.

2. Issues Arising from WLAN Handover

When an MN executes a handover between WLANs with different IP subnets, the following two issues arise.

- The connection is terminated due to the change in IP address
- Both the handover processes and the deterioration of the wireless link condition degrade communication performance

Popular application protocols, such as the file transfer protocol (FTP), simple mail transfer protocol (SMTP), and hypertext transfer protocol (HTTP), employ TCP as the transport protocol (Layer 4). TCP ensures reliability of end-to-end communication between the source and destination hosts, which can be identified by an IP address (Layer 3). Therefore, if the IP address of the MN is changed due to the movement of the MN, TCP communication is terminated.

Many handover management schemes, such as MIP, mSCTP, and other method [10], have already been proposed to solve this termination issue. However, although MNs employing such schemes can maintain TCP communication upon handover, the communication quality is reduced due to the period of handover processing. Specifically, when an MN employing MIPv6 [6] traverses between WLANs managed by different companies or organizations (different IP subnets), the handover process consists of the following five steps.

1. MN movement is detected by loss of router advertisement packets (initiates handover)
2. Newly available APs are scanned
3. An association between the MN and the new AP is established
4. The IP address binding by the dynamic host configuration protocol (DHCP) is updated
5. The binding update (BU) packet is sent to notify the home agent (HA) and corresponding node (CN) of the new IP address

Other schemes also involve a similar set of handover processes. In the MIP network, each MN detects its own movement using router advertisement packets, which are broadcast infrequently by an AP (e.g., the default advertisement interval in MIP is 1 s [5], [6]). The handover process is initiated on the basis of loss of router advertisement packets from connecting AP. However, as the infrequency of advertisement causes latency in the handover decision, the TCP goodput performance could be reduced dramatically immediately prior to handover. Therefore, to achieve a seamless and efficient handover, the handover decision criterion plays an important role in avoiding the performance degradation leading up to handover.

The remaining handover processes can be divided into two main parts: the Link layer handover process ((2), (3)) and the IP layer handover process ((4), (5)). The Link layer handover processing period varies from approximately 50 ms to 400 ms, depending on the hardware used [12]. The IP layer handover process involves both reconfiguration of the IP address using DHCP (avg. 300 ms [13]) and binding update (one-way delay). The handover processing period could therefore exceed 1 s. As the MN cannot send or receive packets in this period, the performance of TCP communication during handover is substantially degraded.

A number of enhanced versions of MIPv6, including fast handover mobile IP (FMIP) [14] and seamless mobile IP (S-MIP) [15], have been proposed as a means of reducing the duration of both the Link layer and IP layer handover processes. However, such schemes involve the deployment and management of special equipment such as a HA and a FA, and are thus extremely difficult to implement across different organizations and ISPs given the current business model of the Internet. The penetration of these schemes has been very slow, and will continue to be so. Comparison of handover performance between MIPv6 and FMIP using a real testbed network or simulation [8], [9] have shown that FMIP can offer shorter interruption time, yet it remains very difficult to reduce the interruption time to zero, even if the S-MIP or FMIP scheme can achieve zero-loss handover. Processing such as buffering and forwarding inherently causes communication interruption, and hence a degradation of TCP performance. The occurrence of even one packet loss due to communication interruption causes severe degradation of TCP performance upon handover.

In this paper, end-to-end handover management as proposed in Ref. [10] is used as a basis for the proposed scheme, allowing transparency to be achieved without the need for the deployment and management of special equipment as in MIP. The number of frame retransmissions is then employed as a new handover decision criterion [16] in order to avoid performance degradation in the lead-up to handover and to select the optimal WLAN. The number of frame retransmissions can be obtained from the MAC layer (Layer 2) using the cross-layer architecture [17]. To eliminate communication interruption due to the handover processes completely, a multi-homing MN equipped with two or more WLAN interfaces is suggested, as proposed in Ref. [18]. Finally, a unified handover management scheme utilizing both cross-layer and multi-homing architectures is proposed as a model for avoiding performance degradation upon handover.

3. Handover Decision Criteria

Existing handover decision criteria, including the number of frame retransmissions as employed in the proposed model, are evaluated below. The effectiveness of the number of frame retransmissions as a handover decision criterion is then examined by comparison with existing criteria through both simulation and practical experiments.

3.1 Upper-Layer Information (Packet Loss, SRTT, Jitter)

A number of existing technologies have been proposed to allow MNs to traverse different IP networks without interrupting communication. One example is MIP, which is being proposed as a standard by the Internet Engineering Task Force (IETF). In MIP networks, each MN detects its own movement using router advertisement packets, which are broadcast from an AP at infrequent intervals (typically once per second). Major movement detection mechanisms such as lazy cell switching (LCS) and eager cell switching (ECS) have been proposed in MIP [19], by which the movement of an MN is detected based on the loss of router advertisement packets. However, the infrequency of advertisement results in long handover decision latency of up to 3 s in LCS (worst case) and 1 s in ECS (worst case).

Another proposed protocol is mSCTP, the mobile extension of SCTP, which is newly equipped with functions to dynamically add or delete the IP addresses of MNs, thereby supporting mobility during handover. In mSCTP, the issues of handover decision are not discussed in detail [7].

A number of new enhanced protocols deal with the handover decision and involve new movement detection mechanisms allowing changes in the transmission condition to be detected quickly. In the methods proposed by Cunningham [20] and Kelly [21], the handover decision is based on the change in either the jitter or smoothed round trip time (SRTT) of a stream of packets. However, the jitter and SRTT may change dynamically due to a range of factors including congestion in a wired network and frequent and sudden transmission errors in a wireless network. These methods are therefore not appropriate for accurately perceiving changes in the transmission condition of wireless links.

3.2 Lower-Layer Information (Beacon Message, Signal Strength)

To solve the above issues, some new enhanced methods base the handover decision on the information obtained from the lower layer. S-MIP [15] employs Layer 2 messages (beacon messages), which are broadcast from an AP frequently (typically once per hundred milliseconds), as a handover decision criterion. However, as a waiting time for receiving a beacon message is necessarily incurred, the interruption time cannot be reduced to zero even using such enhanced methods, and any interruption will cause severe TCP performance degradation.

To detect the degradation of wireless link quality more quickly, the signal strength obtained from Layer 1 has become widely employed as a handover decision criterion [11]. The received signal strength indicator (RSSI) is a common index of the signal strength, assigned an integer value from 0 to 255, and the maximum value obtained from a given WLAN card is vendor-specific (e.g., 0-60 for Atheros, 0-100 for Cisco devices [22]). In addition, the RSSI is also used as a handover decision criterion for intra-domain handover, called roaming [23]. However, the RSSI can fluctuate abruptly due to various and complicated effects such as multi-path fading, intervening objects, and movement. Therefore, as setting an optimal threshold for the handover decision is very difficult, such a measure is not suitable as a handover decision criterion.

3.3 Frame Retransmission

The number of frame retransmissions is proposed in this study as a new handover decision criterion that is effective for wireless links. Frame retransmissions occur in response to reduced signal strength and collision with other frames. In a WLAN, a sender detects successful transmission by receiving an ACK frame in response to a transmitted data frame in a stop-and-wait manner. 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. With request-to-send (RTS)/clear-to-send (CTS), collisions between data frames, indicating a hidden terminal problem, never occur due to the exchange of RTS/CTS frames. If RTS/CTS is applied, the retransmission limit is set to 4 in the 802.11 specification [1], that is, a data frame will be retransmitted a maximum of four times (the initial transmission and three retransmissions) if necessary. Note that collisions may also occur due to interference, which will be address later, even if RTS/CTS is applied.

If the sender does not receive an ACK frame within the retransmission limit, the data frame is treated as a lost packet on the MAC layer. Data frames are thus inherently retransmitted before being treated as a lost packet. Therefore, the number of frame retransmissions allows the MN to quickly perceive deterioration in the condition of the wireless link, and may allow the MN to determine that handover processes should start before packet loss actually occurs.

The method proposed by Velayos [24] employs the number of frame retransmissions as a handover decision criterion. However, that scheme only considers the frame retransmissions caused by collision with frames transmitted from other MNs in a non-interference environment. Moreover, the effectiveness of the number of frame retransmissions as a criterion is only examined analytically. In the present study, the number of frame retransmissions caused by both MN movement and interference is investigated, and the effectiveness of this measure is demonstrated through both simulations and practical experiments.

4. Evaluation of Index Behavior

4.1 Packet Loss vs. Frame Retransmission

The effectiveness of upper-layer information (packet loss and round trip time) as a handover decision criterion is investigated by evaluating the TCP goodput performance upon handover through simulations. In the simulation (Fig. 2), an MN establishes a NewReno TCP connection for file transfer with a corresponding node (CN) via 802.11b WLAN [25], and then moves away from an AP. The TCP goodput performance in this situation is significantly degraded even before the packet loss ratio begins to increase (Fig. 3).

The TCP sender can normally transmit new TCP DATA packets by receiving a TCP ACK packet in response to a transmitted TCP DATA packet. Note that both TCP DATA/ACK packets are treated as data frames over a WLAN. If these packets experience frame retransmissions over a WLAN but successfully receives an ACK frame within the retransmission limit, the round trip time (RTT) between the CN and the MN increases due to the retransmission delay on a WLAN. In this case, the packet transmission speed, i.e., the TCP goodput performance, constantly degrades with the increase of the RTT. As a result, the TCP goodput performance degrades significantly before the packet losses actually occur.

As the packet loss ratio may change dynamically due to various factors, such as congestion in a wired network and frequent and sudden transmission errors in a wireless network, the setting of an optimal threshold for handover decision is quite difficult. Therefore, the degradation of WLAN link quality cannot be promptly and reliably detected using upper-layer information.

On the other hand, we investigate how the distance between the AP and the MN affects both the TCP goodput performance and the number of frame retransmissions (Fig. 4).

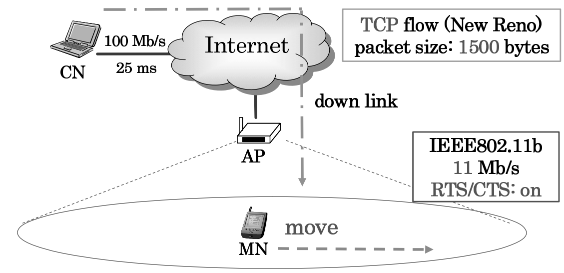

Fig. 2 Simulation model.

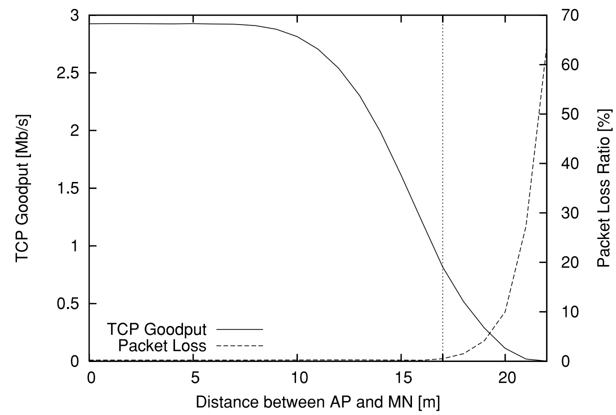

Fig. 3 Relationship between TCP goodput and packet loss ratio.

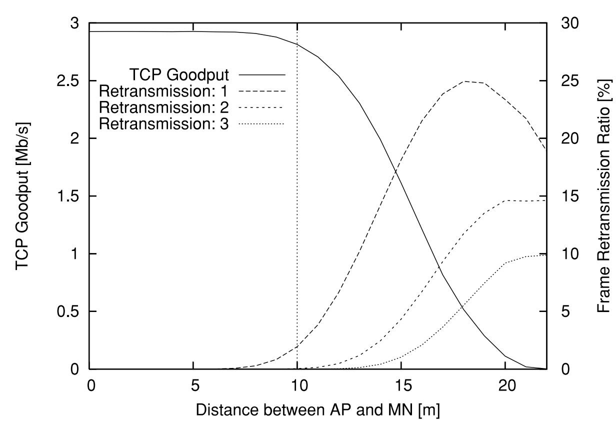

Fig. 4 Relationship between TCP goodput and frame retransmission ratio.

Note that we here employ down-link TCP communication. Thus we examine the number of frame retransmissions experienced by TCP ACK packet that is transmitted from the MN. Here, Retransmission:n indicates the ratio of frames for which n retransmissions are required to the total number of frames. Frame retransmissions begin to occur at around 8 m, and the TCP goodput begins to decrease soon after the occurrence of frame retransmissions. That is, the TCP goodput degrades due directly to frame retransmissions, and the degradation begins as soon as one frame retransmission is triggered. The number of frame retransmissions thus has the potential to serve as an effective handover decision criterion that avoids TCP performance degradation.

Comparison between the number of frame retransmissions and signal strength has yet to be examined in detail due to the complexity of factors affecting signal strength, including multi-path fading, radio interference, intervening objects, and movement. Existing simulators such as Network Simulator Version 2 (NS-2) [26], OPNET [27], and QualNet [28] have difficulty in dealing with unreliability and complexity of radio communications. In the present study, the relationship between frame retransmission and signal strength is evaluated through a practical experiment in a real environment.

4.2 Signal Strength vs. Frame Retransmission

The majority of previous studies have focused on the performance degradation during handover due to reduction and fluctuation (fading) of signal strength caused by movement. However, in the future ubiquitous network, performance degradation may occur as a result of various radio phenomena, such as multi-path fading and noise caused by intervening objects and radio interference. As a result, the handover decision criterion is required to be able to perceive the performance degradation due to reduction of signal strength by MN movement and intervening objects, and that due to radio interference involving other APs. The effectiveness of signal strength and the number of frame retransmissions as handover decision criteria are evaluated here through extensive experiments in a real environment.

The first experiment is conducted to examine whether signal strength and the number of frame retransmissions can promptly and reliably detect the performance degradation due to reduction of signal strength by MN movement and intervening objects in an indoor environment. As shown in Fig. 5, an MN communicates with the CN via 802.11b WLAN. Note that we employ Windows XP on the MN and the CN. Namely, the TCP variant employed here is "Reno with SACK option". The transmission rate of the WLAN is fixed at 11 Mb/s, and the RTS/CTS mechanism is employed. An analyzer node (AN) captures frames transmitted over the WLAN using Ethereal 0.10.13 [29]. The AP is a Proxim ORiNOCO AP-4000 [30], and the WLAN card is a Proxim ORiNOCO 802.11a/b/g Combo Card Gold. Both the MN and AN are equipped with a WLAN card, for communication and frame capture, respectively. The characteristics of signal strength and the number of frame retransmissions are investigated in detail by analyzing captured frames. The communication considered is the download of a 10 MB file from the CN via FTP. TCP goodput is used as a performance measure for the FTP application. As a WLAN card with an Atheros chipset is employed, the RSSI value has a range of 0 to 60 [22].

In the experiment, the change in communication quality (TCP goodput performance for FTP, number of frame retransmissions, and signal strength) as the MN moves away from the AP is examined. The effects of both increase of distance between the AP and the MN and multi-path fading are investigated until the MN reaches a corner. Beyond that,

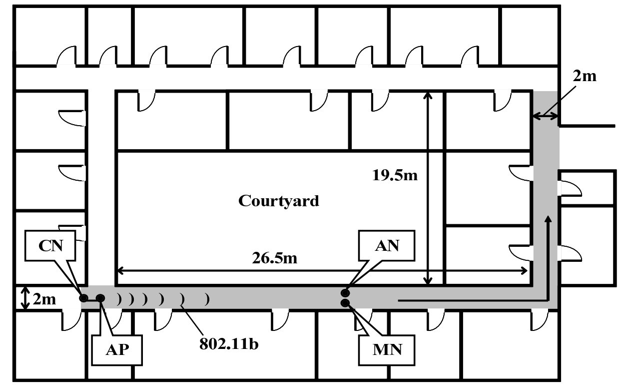

Fig. 5 Experimental environment for evaluating the effect of reduction of signal strength.

the effect of intervening objects between the AP and MN is also investigated. The average communication performance is obtained over 10 experimental runs at each distance with the MN stationary at each point.

Figure 6 shows the change in TCP goodput performance and RSSI in this scenario, and Fig. 7 shows how the TCP goodput performance and frame retransmission ratio change at 16 fixed points from 2 m to 46 m from the AP. In Fig. 6, the TCP goodput is constant out to a distance of 27.5 m, where the MN reaches a corner of the floor (line of sight is maintained). Beyond that, the TCP goodput performance drops dramatically, and then fluctuates markedly at distances greater than 40 m. However, the signal strength decreases monotonically with distance from the AP out to 27.5 m, even though the TCP goodput performance is not degraded at all. Beyond 27.5 m, although the signal strength drops abruptly due to the effect of intervening objects as in TCP goodput performance, however, the signal strength little fluctuates past 40 m in contrast with TCP goodput behavior. From these results, because quick perception of TCP goodput performance degradation is difficult when using signal strength [31], we can remark that the signal strength is not appropriate for a handover criterion enabling to avoid communication degradation.

The frame retransmission ratio, on the other hand, remains low out to 27.5 m, beyond which it increases corresponding to the decrease in the TCP goodput (Fig. 7). In particular, Retransmission: 2 and Retransmission: 3 begin to increase accompanying the decrease in TCP goodput, yet remain nearly zero out to 27.5 m. Beyond 40 m, the frame retransmission ratio fluctuates and the TCP goodput performance roughly follows that behavior. This experiment demonstrates that the degradation of TCP goodput begins in response to an increase in the frame retransmission ratio, showing that degradation of TCP goodput due to reduction in signal strength by MN movement and intervening objects can be detected by monitoring the number of frame retransmissions.

Next, the dynamic communication performance as the MN moves away from the AP is investigated. Figure 8 shows the change in TCP goodput, signal strength, and num-

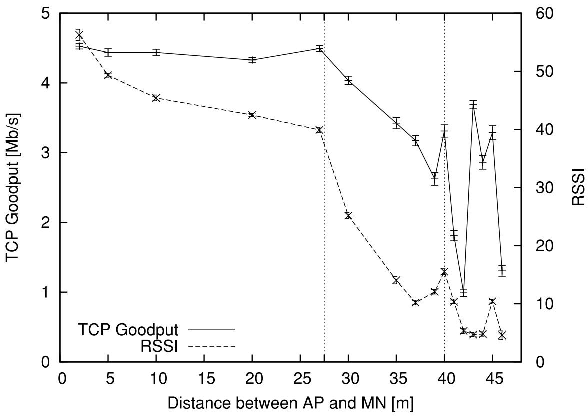

Fig. 6 Relationship between TCP goodput and signal strength.

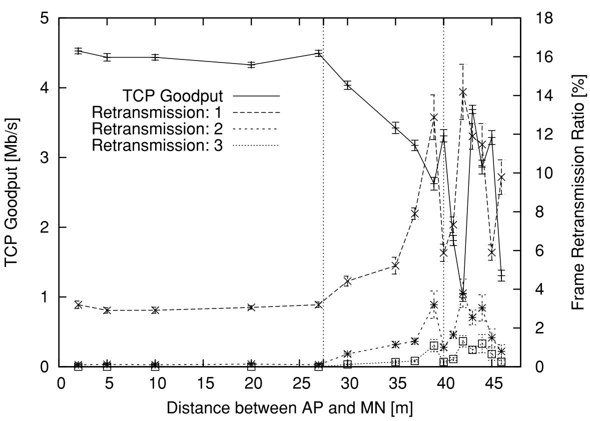

Fig. 7 Relationship between TCP goodput and frame retransmission ratio.

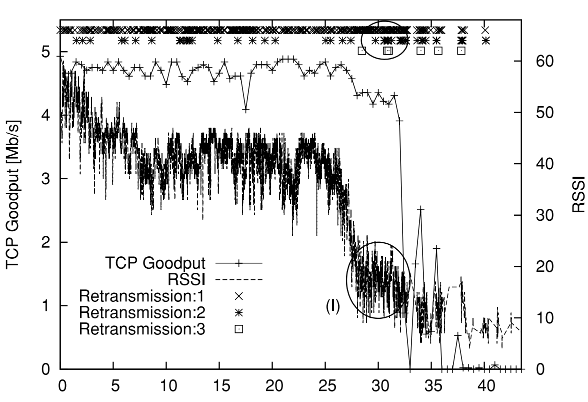

Fig. 8 Relationship between TCP goodput, signal strength, and frame retransmissions as MN moves away from an AP.

ber of frame retransmissions for FTP communication as the MN is moved. Retranamission: n indicates the occurrence distance of a packet that experienced frame retransmissions *n* times.

Comparing Fig. 6 with Fig. 8, it can be seen that RSSI fluctuates sharply and drops abruptly with MN movement. The RSSI fluctuates between approximately 10 and 22 in

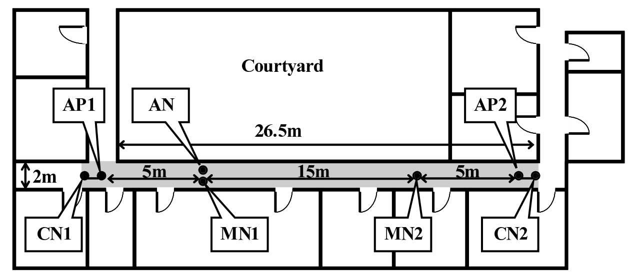

Fig. 9 Experimental environment for evaluating the effect of radio interference.

the region of drastic decrease of TCP goodput (Fig. 8 (I)), making it difficult to set a threshold for executing handover using signal strength as the basis for the handover decision criterion. In contrast, frame retransmissions occur frequently before the communication quality is degraded. Retransmission: 3 in particular increases immediately before the communication quality actually decreases. Using the number of frame retransmissions, it is therefore possible to detect deterioration of communication quality as the MN moves. These experiments thus demonstrate that the number of frame retransmissions has the potential to serve as an effective handover decision criterion for detecting the degradation of communication quality due to the reduction of signal strength by MN movement and intervening objects.

The effect of radio interference with other APs on signal strength and the number of frame retransmissions is examined using the environment shown in Fig. 9. The distance between AP1 and AP2 is set at 25 m, and the distance between each AP and MN (AP1-MN1 and AP2-MN2) is set at 5 m in order to maintain good communication quality and signal strength. Frame collisions due to radio interference occur frequently depending on the number of frames transmitted over the two wireless channels. FTP communication is tested, in which a large number of frames are transmitted. The communication performance of MN1 communication with CN1 via AP1 is monitored, and communication between MN2 and CN2 via AP2 causes radio interference.

The transmission rate of both WLANs is fixed at 11 Mb/s, with the auto rate fallback (ARF) function off and RTS/CTS active. The communication channel of AP1 is fixed to 14 (2471–2497 MHz), which is available in Japan in addition to the standard 13 channels (2400–2485 MHz at 5 MHz intervals) and is independent of channel 11 (i.e., radio interference never occurs). The channel of AP2 is changed from 11 to 14 in a series of experiments. The strength of radio interference increases as the channels of AP1 and AP2 become closer. In each experiment, MN1 downloads a 10 MB file from CN via AP1. Figures 10–12 show the change in TCP goodput, signal strength, and Retransmission: 1 as the channel of AP2 is changed. With AP2 on channel 11, TCP goodput is maintained, since no frame collision due to radio interference occurs between channel 14 (AP1) and channel 11 (AP2). However, the TCP goodput drops drastically as the channel of AP2 is set close to

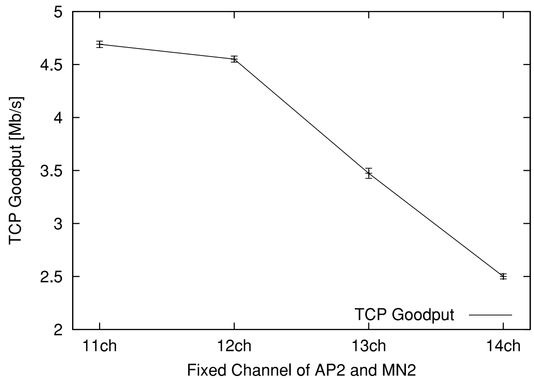

Fig. 10 TCP goodput under radio interference.

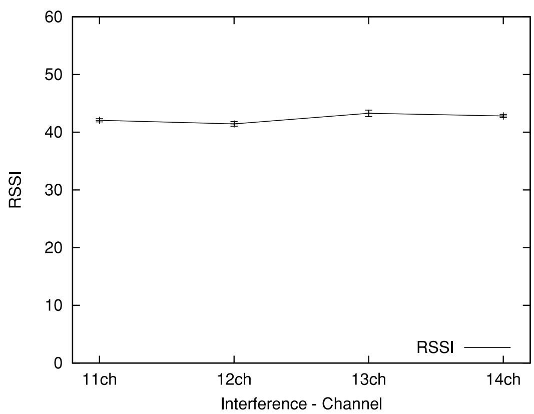

Fig. 11 Signal strength under radio interference.

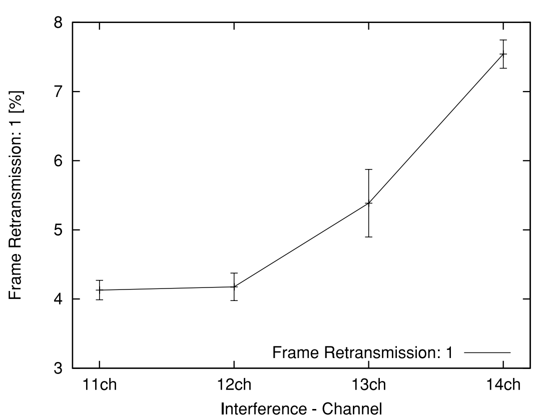

Fig. 12 Frame retransmission ratio under radio interference.

that of AP1, even though the signal strength remains unchanged. Signal strength is therefore unable to detect any degradation of TCP goodput due to radio interference. In contrast, the number of frame retransmissions increases as the channel of AP2 approaches that of AP1. In particular, when AP1 and AP2 are both on channel 14, the number of frame retransmissions increases dramatically due to failure of the carrier sense multiple access/collision avoidance (CSMA/CA) function.

These results demonstrate that signal strength is unable to detect performance degradation due to radio interference, whereas the number of frame retransmissions changes responsively and reliably due to degradation of communication quality caused by radio interference with other APs. An MN employing the number of frame retransmissions as a basis for the handover decision criterion can therefore promptly and reliably detect radio interference and execute handover to an AP that is not so affected.

5. Proposed Unified Handover Management Scheme

The experiments above show that the number of frame retransmissions is an effective basis for a handover decision criterion, allowing MNs to detect degradation of communication quality due to both reduction of signal strength and radio interference. In periods when an MN does not transmit a frame, and just after an MN enters a WLAN, the RSSI will be helpful and may be employed in conjunction with the number of frame retransmissions if necessary. The number of frame retransmissions thus plays an important role to satisfy the first (early initiation of handover) and the third (selection of the optimal WLAN) of the three requirements for eliminating performance degradation during handover. All three requirements including the second (elimination of communication interruption) can be addressed by incorporating a cross-layer architecture and multi-homing. The unified handover management scheme proposed in this study is presented below.

5.1 Early Initiation of Handover

Due to the conventional layered architecture, the information held in each layer cannot be accessed from different layers. In the proposed scheme, the cross-layer approach [17] is employed, the benefit of which is greater than the cost in this scenario [32], to allow interaction between layers. Figure 13 illustrates the proposed concept of the handover management mechanism. A handover manager (HM) on the transport layer perceives the deterioration of wireless link quality based on the number of frame retransmissions obtained from the MAC layer. Note that, as end-to-end handover management is focused on here in order to achieve transparency without the need for deploying and coordination of special equipments, the proposed scheme is applied to both end-to-end hosts only. Furthermore, the HM is implemented in the transport layer, coordinating various functions of the transport layer, such as congestion control, flow control, and retransmission control, as well as initiating handover.

5.2 Elimination of Communication Interruption upon Handover

A handover processing period, in which an MN cannot send or receive packets, cannot be avoided due to the time required for link layer and IP layer handover processes when

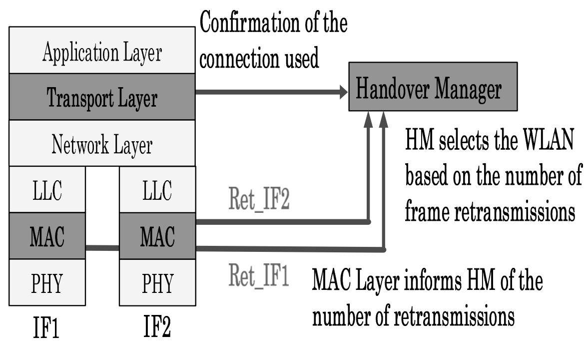

Fig. 13 Cross-layer architecture.

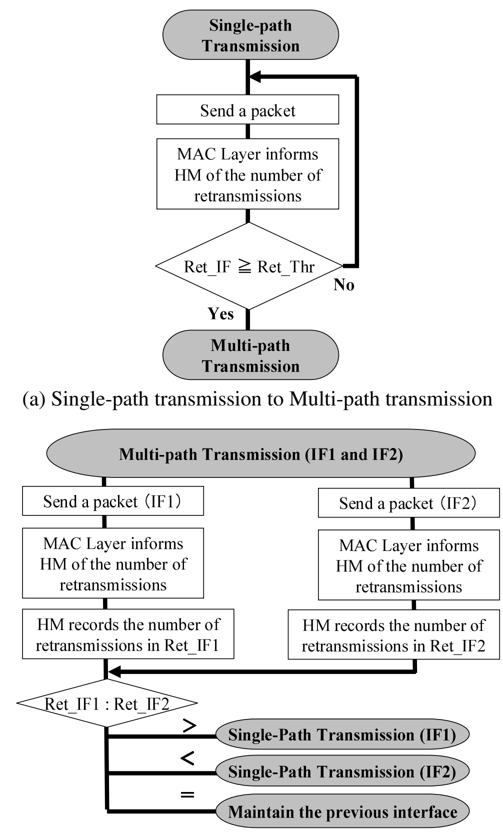

Fig. 14 Handover management (MN).

the MN traverses between WLANs with different IP subnets. In the proposed scheme, the MN is equipped with multiple WLAN interfaces (i.e., multi-homing) to reduce this handover processing time, effectively eliminates communication interruption due to handover by establishing a connection with a new AP before communication with current AP becomes degraded. For example, when an MN communicating with a CN via one WLAN interface (IF1) finds other available APs, a handover candidate AP is selected based on some criterion (e.g., signal strength, and frame retransmissions of probe packets), and a new connection is established to the handover candidate AP via the second WLAN interface (IF2) in advance. As the focus here is on communication quality upon handover, the choice of handover candidate AP is beyond the scope of the present study. MN maintains continuous communication via IF1 throughout this period in exploration of alternate APs and execution handover to the handover candidate AP.

5.3 Selection of Optimal WLAN

Figures 14(a), (b) shows how the optimal WLAN is selected during the handover period. When the number of frame retransmissions of the current interface in single-path transmission mode exceeds the predetermined threshold (Ret_Thr), the MN detects the deterioration of wireless link quality and starts handover. For example, if Ret_Thr is set to 1, the MN judges the deterioration of link quality from the occurrence of one retransmission. The MN then switches to multi-path transmission mode, and informs the CN of the change of transmission mode. After receiving the information, the CN starts parallel transfer over two WLANs. The MN finally selects the WLAN with the better performance. The proposed scheme employs the number of frame retransmissions, which is obtained for each of the available WLANs on the MN, as the criterion for selecting the best WLAN. In this scheme, the number of retransmissions is measured in response to the transmission of a single packet from each WLAN interface after parallel transfer, where each packet has a different sequence number in order to prevent duplicate ACKs. The number of frame retransmissions is recorded in the parameter Ret_IF1/Ret_IF2 (Fig. 13) on the HM. Upon comparison of the value of each interface, the MN selects the WLAN with the smallest number of retransmissions as the better WLAN, and returns to single-path transmission mode over that connection. Finally, the MN notifies the change in its own transmission mode to the CN. In this way, the proposed scheme can execute handover considering the condition of all available WLAN interfaces by comparing the condition of each connection in parallel transfer. As the MN selects the better WLAN based on only one packet transmitted from each WLAN in multi-path transmission mode, the additional network load due to the parallel transfer is extremely limited.

6. Evaluation and Discussion

The performance of the proposed scheme is evaluated through simulations. The main concern is how well the scheme can maintain TCP performance when the MN executes a handover. The advantages of the proposed scheme are discussed by comparing its features with existing handover management schemes.

6.1 Simulation Model

NS-2 (ver 2.27) is employed for simulation, with shadowing and fading effects applied in order to introduce random/burst packet losses in a wireless link. A realistic model is used to evaluate the effect of the movement of an MN

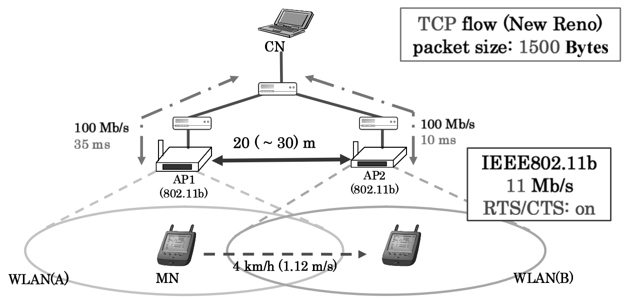

Fig. 15 Simulation model.

from WLAN(A) to WLAN(B), as shown in Fig. 15. We assume that the channel of AP1 and AP2 is completely independent, namely, there is no radio interference effect. The MN first establishes a NewReno TCP connection with a CN via WLAN(A) for file transfer communication with a packet size of 1500 bytes. Each simulation is conducted for an analytical period of 60 s, in which the MN located just under the AP1 starts to move toward to AP2 of WLAN(B) at 35 s. The MN moves at a walking speed of 4 km/h. The one-way delay between the CN and MN is different on different WLANs, assuming that each WLAN is managed by different organizations (i.e., different IP subnets). The delay via WLAN(A) is 35 ms and that via WLAN(B) is 10 ms. These simulations test how well the proposed scheme can avoid degradation of TCP performance upon handover with changes in the retransmission threshold (Ret_Thr) and the distance between AP1 and AP2. As stated in Sect. 5.2, the choice of handover candidate AP is beyond the scope of the present study. That is, we assume that AP2 is selected as the candidate AP.

6.2 Simulation Results

The simulation results are presented below for varying frame retransmission threshold (Ret_Thr) and the distance between APs.

6.2.1 Effect of Frame Retransmission Threshold (Ret_Thr)

The TCP goodput performance when the MN located under AP1 moves toward AP2 with the distance between the two APs fixed at 20 m is shown in Figs. 16–19 for Ret_Thr values of 1 to 4. The TCP goodput is calculated from the volume of effective data received at the TCP receiver in 1 s intervals. “HO-Start” in each figure indicates the time at which the MN starts the parallel transfer, that is, when the MN perceives deterioration of the wireless link condition based on the number of frame retransmissions. “HO-Finish” indicates the time at which the MN finally switches to the optimal WLAN. These two points occur on either interface (IF1 or IF2) as indicated in the figure. Arrows indicate the transition to the optimal WLAN interface. In the initial state, IF1 is treated as the optimal WLAN and the MN communicates with the CN via WLAN(A).

Figure 16 shows that the MN executes the handover at

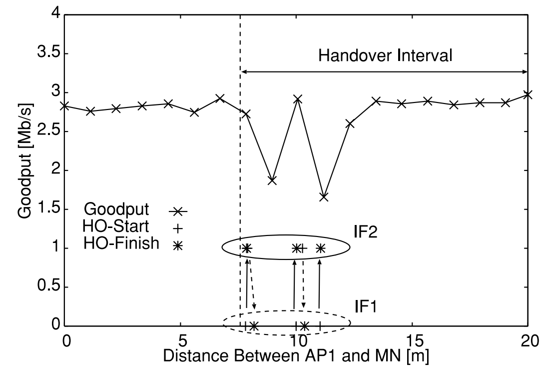

Fig. 16 Goodput performance using proposed scheme (Ret_Thr = 1).

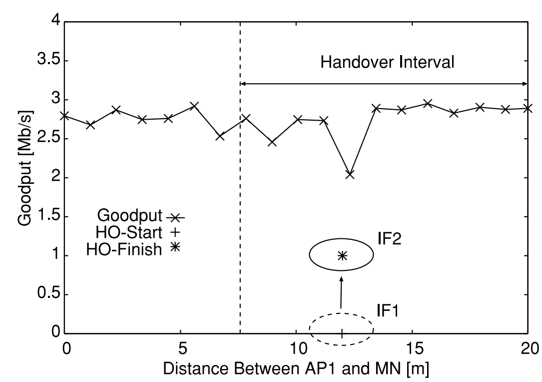

Fig. 17 Goodput performance using proposed scheme (Ret_Thr = 2).

a distance of approximately 8 m from AP1 when Ret_Thr is set to 1. As shown in Fig. 3, the TCP goodput performance begins to degrade at 10 m. Therefore, the MN in this case initiates handover before 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. The goodput performance is somewhat lower during this unstable period. Beyond 11 m, the MN selects WLAN(B) and the communication becomes stable with fully recovered TCP goodput performance, indicating that handover from AP1 to AP2 has been completed.

With the Ret_Thr set to 4 (Fig. 19), the MN initiates handover at 17 m from AP1, and the goodput near the changeover is quite low (0.5 Mb/s) due to the high latency of the handover decision. After MN selects WLAN(B) as the optimal WLAN, communication quality is successfully recovered.

From these results, it can be seen that the MN can quickly perceive deterioration of the wireless link condition when the value of Ret_Thr is small (1), although multiple handovers occur because the MN starts the handover too early, that is, while the condition of the current WLAN is still better than that offered by other available WLANs. On the contrary, when Ret_Thr is set to a high value (3 or 4), handovers do not occur often and the current connection is

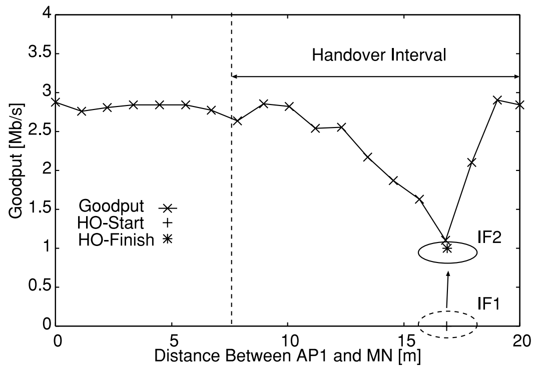

Fig. 18 Goodput performance using proposed scheme (Ret_Thr = 3).

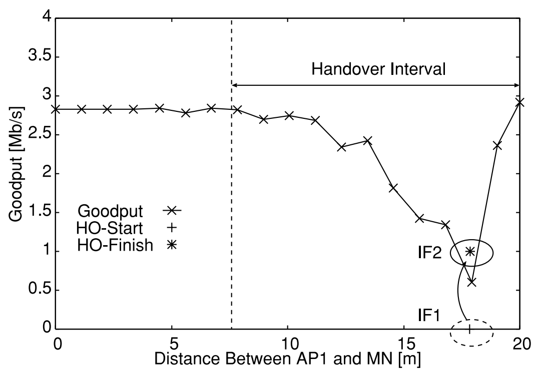

Fig. 19 Goodput performance using proposed scheme (Ret_Thr = 4).

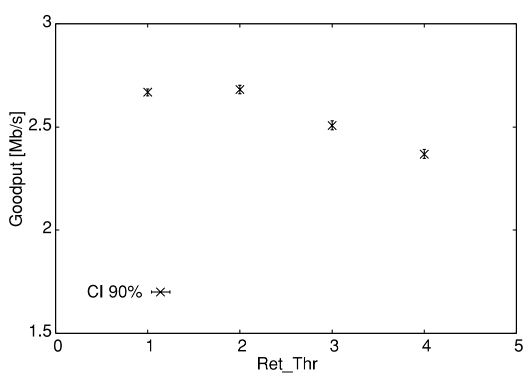

Fig. 20 Variation in average goodput with Ret_Thr (handover interval).

maintained despite a drop in wireless link quality. The value of Ret_Thr therefore has a strong effect on the TCP goodput performance achieved by the proposed scheme and should be selected carefully.

With Ret_Thr set to 2 (Fig. 17), the MN executes the handover at around 12 m, and selects WLAN(B) as the optimal WLAN without alternation. In this case, the proposed scheme quickly perceives deterioration of the wireless link condition and appropriately selects the better WLAN in a single handover. With this value of Ret_Thr, the proposed scheme maintains excellent goodput performance throughout the handover operation.

The effect of Ret_Thr on TCP goodput and the number of handovers was clarified by considering 400 simulations to obtain the 90% confidence interval (CI) of both average goodput and the number of handover occurrences. Figure 20 shows the confidence interval of goodput for several values of Ret_Thr corresponding to the handover interval in Figs. 16–19. It is assumed that the handover should be executed after frame retransmissions begin to occur. Therefore, the handover interval is defined as 8–20 m from AP1. Fig. 20 shows that an Ret_Thr value of 2 achieves the best goodput performance under the proposed scheme. The high goodput at Ret_Thr = 1 and lower performance at Ret_Thr ≥ 3 confirm the results above. Figure 21 shows that handover occurs several times with Ret_Thr set at 1, but only once with Ret_Thr set at higher values. However, as the handover occurs only five times, the goodput performance is ntjot reduced by a large amount.

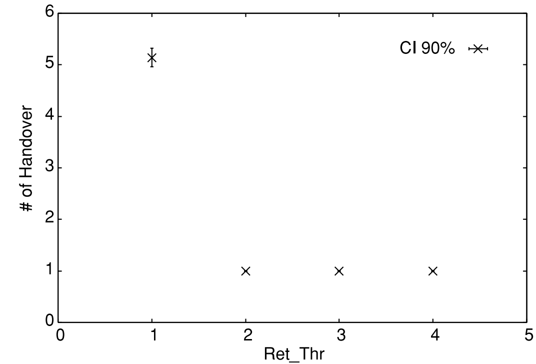

Fig. 21 Variation in number of handovers with Ret_Thr.

6.2.2 Effect of the Change in the Distance between APs

The distance between the two APs affects how many times alternating handovers are made in the handover interval. With the distance between APs fixed to 20 m in the above simulations, the condition of WLAN(B) becomes stable where that of WLAN(A) becomes unstable, and the number of handovers does not increase drastically. As the APs becomes more distant, however, the number of handovers can be expected to increase, causing goodput to decrease. The relationship between the number of handovers and the goodput performance of the proposed scheme as the distance between APs is increased from 20 m to 30 m is examined through additional simulations. To show the relationship between Ret_Thr and the distance between APs, 400 simulations are conducted with each value of Ret_Thr (1-4). Figure 22 shows the number of handovers, and Fig. 23 shows the change in goodput. With Ret_Thr set at 1, increasing the distance between APs causes the number of handovers to increase to 35, with a corresponding decrease in goodput. When Ret_Thr set at 2, the number of handovers increases and the goodput decreases beyond 28 m. However, with Ret_Thr set at 3, the proposed scheme selects the better WLAN in a single handover, regardless of distance between APs. For the region beyond 28 m, this value of

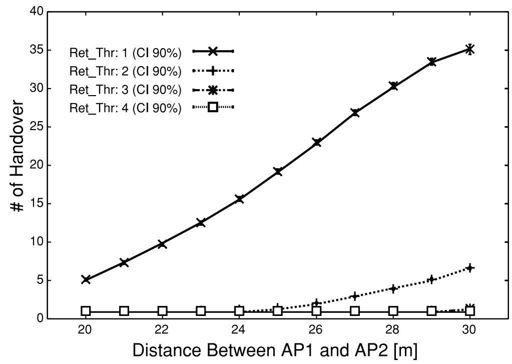

Fig. 22 Variation in number of handovers with distance between APs.

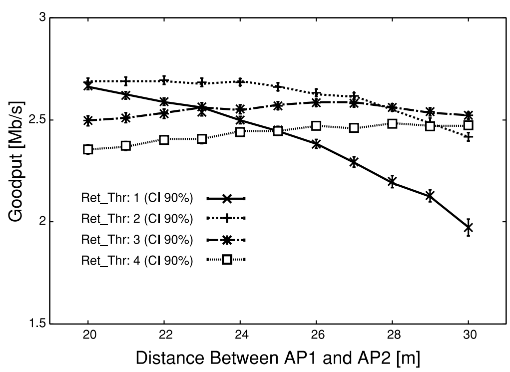

Fig. 23 Variation in average goodput with distance between APs.

Ret_Thr achieves better goodput performance than Ret_Thr = 2.

These results demonstrate that the appropriate value of Ret_Thr depends on the distance between APs, yet with Ret_Thr set to an appropriate value, the proposed scheme selects the optimal WLAN in a minimal number of handovers, avoiding degradation of goodput through the handover operation with any distance between APs to which the value of Ret_Thr is applicable.

6.3 Comparison with Existing Handover Management Schemes

The proposed scheme is compared with the exiting mobility management schemes listed in Table 1. Existing handover management scheme handle either Layer 2 handover (L2 HO) or Layer 3 handover (L3 HO). SyncScan [33] monitors beacon messages transmitted on each channel by regularly switching to each channel, and makes the handover decision based on the signal quality received from multiple APs. Upon this architecture, SyncScan minimizes the handoff processing periods associated with authentication, reassociation, and so on. MultiScan [34] employs multihoming (M-H) and eliminates the handover processing time by exploiting multiple WLAN interfaces. Velayos [24] uses the number of frame retransmissions as a handover criterion and minimizes the IEEE 802.11b handoff time.

Table 2 lists the handover decision criteria used by these handover management schemes. The enhanced protocols of MIP and mSCTP employ lower-layer information as the handover decision criterion to reduce time of handover processes. However, as MIP does not employ M-H and the issues of handover decision are not clearly described in mSCTP, communication interruption due to handover processes still occurs, thereby degrading communication performance. SyncScan, MultiScan, and Velayos also employ lower-layer information, but cannot handle L3 HO. In future ubiquitous WLANs, as MNs are very likely to traverse WLANs (requiring handover) with different IP subnets, it is necessary for the handover management scheme to handle L3 HO to provide transparent mobility for MNs. That is, the L2 HO management schemes will not be practical in such networks.

Table 1 Architecture and target layer of handover management schemes.

	L2 HO	L3 HO	M-H	C-L
MIP	×	○	×	×/(○ ([14] [15])
mSCTP	×	○	○	×/(○ ([11]))
SyncScan	○	×	×	×
MultiScan	○	×	○	×
Velayos	○	×	×	×
Proposed	×	○	○	○

Table 2 Handover decision criterion of handover management schemes.

	Lower Info.		Upper Info.
	Beacon/RSSI	F-retrans.	P-loss/SRTT/Jitter
MIP	×/(○ ([15])	×	○
mSCTP	×/(○ [11])	×	○
SyncScan	○	×	×
MultiScan	○	×	×
Velayos	×	○	×
Proposed	×	○	×

The proposed scheme, on the other hand, supports multi-homing to avoid communication interruption during handover, and employs the number of frame retransmissions as a handover decision criterion to quickly detect the change in wireless link quality and to reliably select the optimal WLAN. The number of frame retransmissions can be obtained from Layer 2 by virtue of the cross-layer (C-L) architecture. As a result, we can remark that only the proposed scheme has the three functions (handover decision using lower Info, multi-homing, cross-layer) for achieving seamless and efficient L3 HO that can avoid degradation of communication quality.

7. Conclusion and Future Work

A unified handover management scheme based on the number of frame retransmissions was proposed. The proposed scheme satisfies the requirements of early initiation of handover based on quick perception of change in wireless link quality, elimination of communication interruption due to the handover operation, and selection of the optimal WLAN. Early detection of the deterioration of wireless link quickly is achieved by employing the number of frame retransmissions obtained from Layer 2 using a cross-layer architecture. Communication interruption due to the handover operation is eliminated by providing two WLAN interfaces on the mobile node (multi-homing). The optimal WLAN is selected by maintaining a parallel connection using two interfaces, allowing the condition of the two best available WLANs to be compared by transmitting single packets over each simultaneously. The network load due to this multi-path transmission is very minor.

Simulation results showed that the goodput performance deteriorates due to an increase of the number of handovers (small threshold value) and increased latency of quality detection (large threshold value). With the distance between APs fixed at 20 m, the proposed scheme with a retransmission threshold set at 2 achieves excellent goodput performance throughout the handover, where deterioration of the wireless link condition is detected early and the optimal WLAN is selected in a single handover. The most appropriate threshold value was found to depend on the distance between APs, suggesting that a dynamic decision algorithm for the handover threshold will be helpful for increasing the robustness of the proposed scheme with respect to variations in AP intervals.

In this paper, we employed WLAN APs whose transmission rate is fixed to 11 Mb/s in order to focus on the effect of the MN movement. Development of the proposed scheme that can apply to the WLAN APs with a function of Auto Rate Fallback (ARF) is currently underway. Furthermore, the frequency of the number of frame retransmissions may depend on the frame size. As we focused on FTP download communication, the TCP ACK packet that is transmitted from the MN and is treated as a small DATA frame over WLAN was used for acquiring the number of frame retransmissions. However, because the frame size transmitted from the MN becomes quite large under FTP upload communication such as file exchange, we will examine the effect of the difference of the frame size.

The proposed scheme was evaluated through extensive simulations of FTP communication. Practical implementation of the scheme using the Linux kernel is currently being carried out. A cross-layer architecture based on the number of frame retransmissions has also been proposed for VoIP (UDP) [18], and implementation of that architecture has been successful. Demonstrations of the scheme were presented at MobiHoc 2006 [35]. The proposed cross-layer architecture is therefore highly practical from an implementation perspective, and the results of implementation of the handover scheme presented in this study will be presented in the near future. Moreover, we will examine how the occurrence of the number of frame retransmission is different between TCP and UDP communication by using both implementations.

To successfully send packets over a wireless radio link, the sender in any wireless network retransmits data frames when a data or an ACK frame is lost. The characteristics of frame retransmissions described in this paper are thus theoretically applicable in other wireless access networks such as 802.11n and WiMAX. However, as various parameters such as bandwidth and propagation delay differ substantially among these technologies, the application of a handover management scheme based on frame retransmissions needs to be discussed, and further enhancements to the scheme considering their differences may be needed.

Acknowledgment

This work was supported in part by the Japan Society for the Promotion of Science, Grant-in-Aid for Scientific Research (S)(18100001) and in part by the Ministry of Internal Affairs and Communications, Japan.

References

[1] IEEE 802.11, 1999 Edition, Available at 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, ed., "IP mobility support for IPv4," IETF RFC3344, Aug. 2002.
[6] D. Johnson, C. Perkins, and I. Arkko, "Mobility support in IPv6," IETF RFC3775, June 2004.
[7] M. Riegel and M. Tuexen, ed., "Mobile SCTP," IETF Internet Draft, draft-riegel-tuexen-mobile-sctp-07.txt, Oct. 2006.
[8] Y. Kim, D. Kwon, K. Bae, and Y. Suh, "Performance comparison of mobile IPv6 and fast handovers for mobile IPv6 over wireless LANs," Proc. IEEE VTC2005-fall, CD-ROM, Sept. 2005.
[9] N. Montavont and T. Noël, "Analysis and evaluation of mobile IPv6 handovers over wireless LAN," Mobile Netw. Appl., vol.8, no.6, pp.643–653, Dec. 2003.
[10] K. Tsukamoto, Y. Hori, and Y. Oie, "Mobility management of transport protocol supporting multiple connections," Proc. ACM MobiWac2004, pp.83–87, Oct. 2004.
[11] M. Chang, M. Lee, and S. Koh, "Transport layer mobility support utilizing link signal strength information," IEICE Trans. Commun., vol.E87-B, no.9, pp.2548–2556, Sept. 2004.
[12] A. Mishra, M. Shin, and W. Arbaugh, "An empirical analysis of the IEEE802.11 MAC layer handoff process," ACM SIGCOMM Computer Communication Review, vol.33, no.2, pp.93–102, April 2003.
[13] A. Dutta, F. Vakil, J. Chen, and M. Tauil, "Application layer mobility management scheme for wireless Internet," Proc. IEEE 3G Wireless, pp.379–385, May 2001.
[14] P. McCann, "Mobile IPv6 fast handovers for 802.11 networks," IETF RFC4260, Nov. 2005.
[15] R. Hsieh, Z.G. Zhou, and A. Seneviratne, "S-MIP: A seamless handoff architecture for mobile IP," IEEE Infocom2003, pp.1774–1784, March 2003.
[16] S. Kashihara, K. Tsukamoto, Y. Kadobayashi, and Y. Oie, "A simple heuristic for handover decisions in WLANs," IETF Internet Draft, draft-shigeru-simple-heuristic-wlan-handover-00.txt, March 2006.
[17] S. Shakkottai, T.S. Rappaport, and P.C. Karlsson, "Cross-layer design for wireless networks," IEEE Commun. Mag., vol.41, no.10, pp.74–80, Oct. 2003.
[18] S. Kashihara and Y. Oie, "Handover management based upon the number of retries for VoIP in WLANs," Proc. IEEE VTC2005-Spring, CD-ROM, May 2005.
[19] C. Perkins, Mobile IP, Design Principles and Practices, Addison-Wesley, 1998.
[20] G. Cunnigham, P. Perry, and L. Marphy, "Soft, vertical handover of streamed video," Proc. IEE 3G Mobile Communication Technologies (3G 2004), Oct. 2004.

[21] A. Kelly, G.M. Muntean, P. Perry, and J. Murphy, "Delay-centric handover in SCTP over WLAN," Trans. Automatic Control and Computer Science, vol.49, no.63, pp.211–216, May 2004.
[22] K. Muthukrishnan, N. Meratnia, M. Lijding, G. Koprinkov, and P. Havinga, "WLAN location sharing through a privacy observant architecture," Proc. COMSWARE2006, pp.1–10, Jan. 2006.
[23] P. Khadivi, T.D. Todd, and D. Zhao, "Handoff trigger nodes for hybrid IEEE 802.11 WLAN/Cellular networks," Proc. QSHINE2004, pp.164–170, 2004.
[24] H. Velayos and G. Karlsson, "Techniques to reduce the IEEE 802.11b handoff time," Proc. IEEE ICC2004, vol.7, pp.3844–3848, June 2004.
[25] IEEE 802.11b, 1999 Edition, Available at http://standards.ieee.org/getieee802/download/802.11b-1999.pdf
[26] The Network Simulator 2 ns-2, http://www.isi.edu/nsnam/ns/
[27] OPNET, http://www.opnet.com/
[28] QualNet, http://www.scalable-networks.com/products/QualNet
[29] Ethereal, http://www.ethereal.com/
[30] Proxim, http://www.proxim.com/
[31] K. Tsukamoto, T. Yamaguchi, S. Kashihara, and Y. Oie, "Experimental evaluation of decision criteria for WLAN handover: Signal strength and frame retransmission," IEICE Trans. Commun., vol.E90-B, no.12, pp.3579–3590, Dec. 2007.
[32] V. Kawadia and P.R. Kumar, "A cautionary perspective on cross-layer design," IEEE Wireless Communications, vol.12, no.1, pp.3–11, Feb. 2005.
[33] I. Ramani and S. Savage, "SyncScan: Practical fast handoff for 802.11 infrastructure networks," Proc. IEEE INFOCOM2005, pp.675–684, March 2005.
[34] V. Brik, A. Mishra, and S. Banerjee, "Eliminating handoff latencies in 802.11 WLANs using multiple radios: Applications, experience, and evaluation," Proc. IMC 2005, pp.299–304, Oct. 2005.
[35] S. Kashihara, K. Tsukamoto, H. Koga, and Y. Oie, "Handover management based on the number of frame retransmissions for VoWLANs," ACM MobiHoc2006 Demo Session, May 2006.

Kazuya Tsukamoto received his D.E. degree in computer science from Kyushu Institute of Technology (KIT), Japan, in 2006. From April 2006 to March 2007, he was a Japan Society for the Promotion of Science (JSPS) Research Fellow at KIT. Beginning in July 2006 and ending in January 2007, he spent six months at the Department of Information and Computer Science, University of California, Irvine, on leave of absence from KIT. As of April 2007, he is an assistant professor with the Department of Computer Science and Electronics, KIT. His research interests include performance evaluation of computer networks and wireless networks. He is a member of the IEEE.

Shigeru Kashihara received the D.E. degree in information science from Nara Institute of Science and Technology (NAIST), Japan, in 2003. From January 2004 to February 2005 he was a postdoctoral researcher at KIT. He is currently an assistant professor in the Graduate School of Information Science, NAIST. His research interests include ubiquitous computing and networks. He is a member of the IEEE.

Yuji Oie received the B.E., M.E., and D.E. degrees from Kyoto University in 1978, 1980, and 1987, respectively. From 1980 to 1983 he worked at Nippon Denso Company Ltd., Kariya, Japan. From 1983 to 1990 he was with the Department of Electrical Engineering, Sasebo College of Technology. From 1990 to 1995 he was an associate professor in the Department of Computer Science and Electronics, Faculty of Computer Science and Systems Engineering, KIT. From 1995 to 1997 he was a professor at the Information Technology Center, NAIST. Since April 1997 he has been a professor in the Department of Computer Science and Electronics, Faculty of Computer Science and Systems Engineering, KIT. His research interests include performance evaluation of computer communication networks, high-speed networks, and queuing systems. He is a fellow of the IPSJ, and a member of the IEEE.