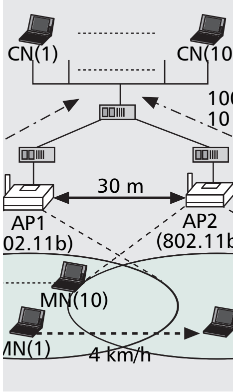
SERVICE-ORIENTED MOBILITY MANAGEMENT ARCHITECTURE FOR SEAMLESS HANDOVER IN UBIQUITOUS NETWORKS

SHIGERU KASHIHARA, NARA INSTITUTE OF SCIENCE AND TECHNOLOGY
KAZUYA TSUKAMOTO AND YUJI OIE, KYUSHU INSTITUTE OF TECHNOLOGY



In the conventional Internet architecture, an MN can never inherently avoid the degradation in communication quality during handover. To achieve seamless handover, we propose a service-oriented mobility management scheme to address application quality.

ABSTRACT

In ubiquitous wireless LANs, a mobile node is likely to move between many access points while using certain applications. However, in the conventional Internet architecture, an MN can never inherently avoid the degradation in communication quality during handover. To achieve seamless handover, we propose a service-oriented mobility management scheme to address application quality. In this article, we first clarify three requirements for achieving seamless handover. We then describe our concept of the service-oriented mobility management scheme, which satisfies all three requirements. Our main contribution is the proposal of a scheme of how to properly use the number of frame retransmissions as a new handover-decision criterion to accomplish seamless handover. Performance evaluations show that our proposed scheme can maintain application quality during handover.

INTRODUCTION

With the proliferation of mobile Internet users, diverse wireless access network technologies, such as wireless LAN (WLAN), WiMAX (worldwide interoperability for microwave access), and cellular networks appeared. WLANs, based on the IEEE 802.11 family [1](IEEE 802.11a/b/g), spread rapidly due to their low cost, simplicity of installation, and high data-transmission rates and are being set up in both private and public spaces. Furthermore, WLANs that are independently managed by different IP subnets, that is, different organizations or Internet service providers (ISPs), are starting to cover a wide area, such as an entire city, in a complementary way by using numerous access points (APs). Indeed, many wireless networks, for example, WIFLY in Taipei, Wireless Philadelphia, and Wireless London, are springing up around the world. In the near future, WLANs will continue to spread until they overlap to provide continuous coverage over a wide area and then will fulfill the important role of ubiquitous networks as ubiquitous WLANs.

In ubiquitous WLANs, a mobile node (MN) can access the Internet through an AP at any location. However, as the MN may move while using diverse applications, it will experience numerous handovers between different IP subnets because of the narrow coverage of a WLAN, and the communication may be disconnected due to a change in the IP address of the MN. To support mobility despite a change of IP address, mobility management schemes such as Mobile IP (MIP) [2] and mobile Stream Control Transmission Protocol (mSCTP) [3] were proposed. However, the degradation of communication quality during handover can never inherently be avoided whichever existing mobility management scheme is employed. More specifically, the MN cannot send or receive packets during handover processes due to layer 2 and 3 handover operations. Furthermore, communication quality is sensitive to changes in wireless link conditions. Therefore, to achieve seamless handover, the following three requirements should be satisfied:

- Initiation of handover based on prompt and reliable detection of change in wireless link condition
- Elimination of communication interruption due to handover processes
- Selection of an optimal WLAN

To provide service-oriented mobility in ubiquitous networks, we also must carefully consider application quality during handover. Most existing applications employ either Transmission Control Protocol (TCP, for non real-time applications) or UDP (User Datagram Protocol, for real-time applications) as a transport protocol. In non real-time applications such as file transfer (FTP), the important performance measure is goodput, while packet loss, RTT (round trip time), and jitter are important in real-time applications such as Voice over IP (VoIP). Because of this difference between non-real-time and real-time applications, adaptive execution of handover according to applications is required.

 1536-1284/07/$20.00 © 2007 IEEE

We first describe a new handover decision criterion that considers applications to promptly and reliably initiate handover. We then propose a service-oriented mobility management scheme satisfying the above three requirements and then evaluate its performance.

Handover Decision Criterion

To achieve seamless handover, a handover decision criterion is crucial to execute handover based on prompt and reliable detection of changes in the wireless link condition. However, in MIP networks, an MN detects its own movement by means of router advertisement (RA) packets, which are broadcast infrequently from an AP (typically one per second). This frequency increases the handoff decision latency, thereby causing degradation in application quality. On the other hand, although some enhanced MIP schemes, such as Fast Mobile IP (FMIP) [4] and Hierarchical Mobile IP (HMIP) [5], were proposed to improve communication performance during handover, they have not been examined in terms of handover decision criteria. In one recent study, although mSCTP was found to support mobility during handover, the issues of handover decision criteria were also not discussed in detail [3].

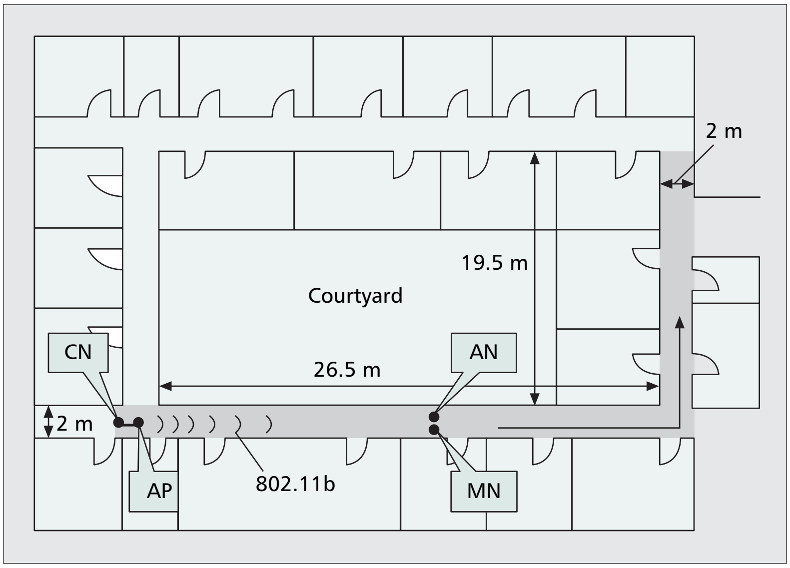
Effective handover decision criteria thus remain unanswered. In ubiquitous WLANs, the communication quality is often degraded, due to:

1. Reduction of signal strength.
2. Radio interference with other WLANs.

Proposing a handover decision criterion that detects both is essential.

Signal strength usually is used as the performance measure of the wireless link condition. However, it is very difficult for an MN to properly detect deterioration in communication quality, because signal strength fluctuates abruptly due to distance and interfering objects. In addition, signal strength is a difficult basis on which to set a threshold for handover, because the allowable range of received signal strength indicator (RSSI), a common index of signal strength, depends on each vendor. Cisco chooses 100 as RSSI-max, while the Atheros chipset chooses 60 [6]. Finally, in ubiquitous WLANs, as degradation in communication quality frequently occurs due to radio interference, an MN also must detect radio interference.

In this article, we focus on the number of frame retransmissions over a WLAN as a new handover decision criterion that can detect both items, 1 and 2, mentioned previously. We first outline the frame retransmission mechanism of IEEE 802.11 [1]. When a data or an acknowledgment (ACK) frame is lost over a WLAN, the sender (e.g., an MN) retransmits the same data frame to the receiver (e.g., an AP) until the number of frame retransmissions reaches a predetermined retry limit. If RTS/CTS (request to send/clear to send) is applied, the retry limit is set to four; otherwise, it is set to seven. Therefore, a data frame can be retransmitted a maximum of four or seven times (the initial transmission and three or six retransmissions), if necessary. If the sender does not successfully receive an ACK frame within the retry limit, it treats the data frame as a lost packet. Note that RTT and jitter also increase due to frame retransmissions. As a packet inherently experiences frame retransmissions before the occurrence of packet loss or the increase of RTT and jitter, it is therefore reasonable to suppose that the number of frame retransmissions can be a new handover decision criterion [7].



■ **Figure 1.** *Experimental model.*

Experimental Evaluation of Handover Decision Criterion

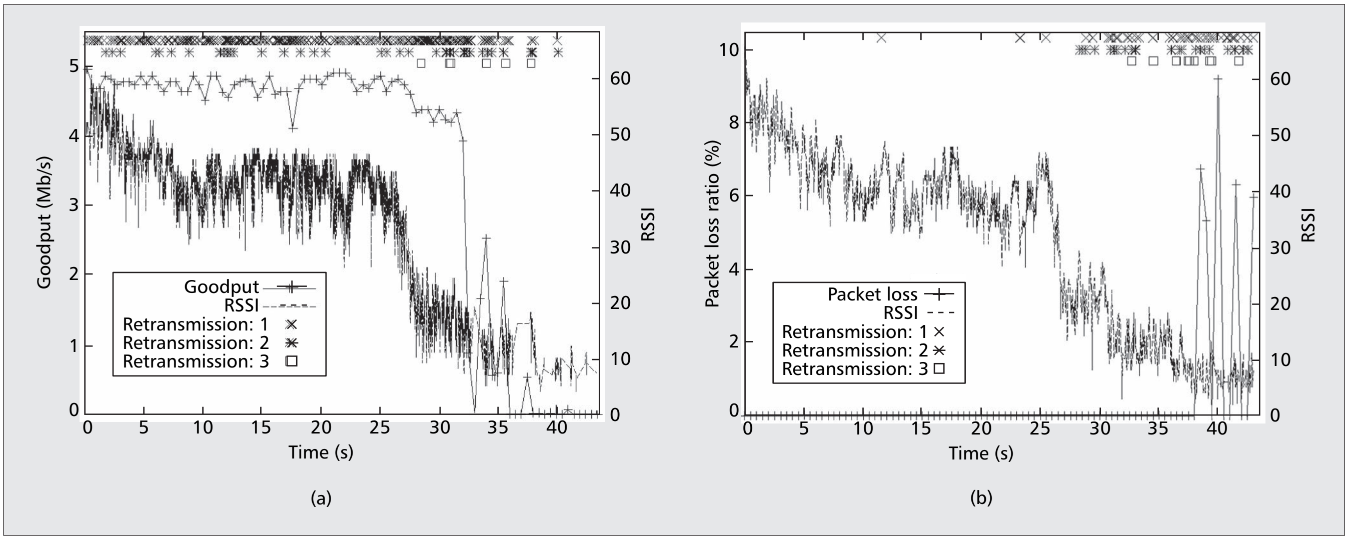
In ubiquitous WLANs, prompt and reliable detection of changes in the wireless link condition is essential for avoiding degradation in communication quality and for executing a handover to another, better WLAN. In this section, employing FTP and VoIP applications, we show how RSSI and frame retransmissions can promptly and reliably detect the performance degradation in each application due to the two factors (reduction of signal strength and radio interference, mentioned earlier) in a real environment [8].

Experimental Model

As illustrated in Fig. 1, an MN communicates with a correspondent node (CN) via an AP of 802.11b. The transmission rate of the WLAN is fixed at 11 Mb/s, that is, auto rate fallback is not employed, and the RTS/CTS mechanism is employed. The AP is Proxim ORiNOCO AP-4000, and the MN WLAN card is ORiNOCO 802.11a/b/g Combo Card Gold. To capture frames transmitted over the WLAN, Ethereal 0.10.13 is installed in an analyzer node (AN). In FTP, the MN downloads a 10-megabyte file from the CN (FTP server). In VoIP, the MN communicates with the CN by using a VoIP application (Gphone 2.0) for 60 s.

Experimental Results

We examine in detail how 1 and 2 affect RSSI, frame retransmissions, and application quality



■ **Figure 2.** *Change in communication quality, RSSI, and the number of frame retransmissions: a) FTP communication; b) VoIP communication.*

(FTP and VoIP). TCP goodput is used as the performance measure of FTP application, and packet loss ratio is used as that of the VoIP application. Although RTT and jitter usually also are considered as performance measures, we focus on packet loss rate, which is strongly related to application quality, because RTT and jitter are relatively small in this experiment.

We first investigate the communication performance for FTP and VoIP when the MN moves away from the AP. Figure 2a shows the change in TCP goodput, RSSI, and the number of frame retransmissions under FTP communication. Figure 2b shows the change in packet loss ratio, RSSI, and the number of frame retransmissions under VoIP communication. *Retransmission: n* indicates the occurrence time of a packet suffering *n* retransmissions.

From these figures, RSSI fluctuates greatly and drops drastically with the movement of the MN and intervening objects. Note that the range of RSSI is from 0 to 60 (Atheros chipset). In Fig. 2a the value of RSSI fluctuates approximately from 10 to 22, when the TCP goodput begins to decrease. On the other hand, in Fig. 2b, it ranges from approximately 4 to 8, when packet loss begins to occur. That is, the value of RSSI, when communication quality begins to decrease, is different depending on each application. Consequently, a threshold setting for handover is required for each application if RSSI is employed as a handover decision criterion. In contrast, frame retransmissions frequently occur soon before communication quality is degraded. In particular, Retransmission: 3 begins just before the communication quality actually decreases. From these results, we can see that the number of frame retransmissions has the potential to detect degradation in communication quality due to reduction of signal strength, irrespective of the kind of application.

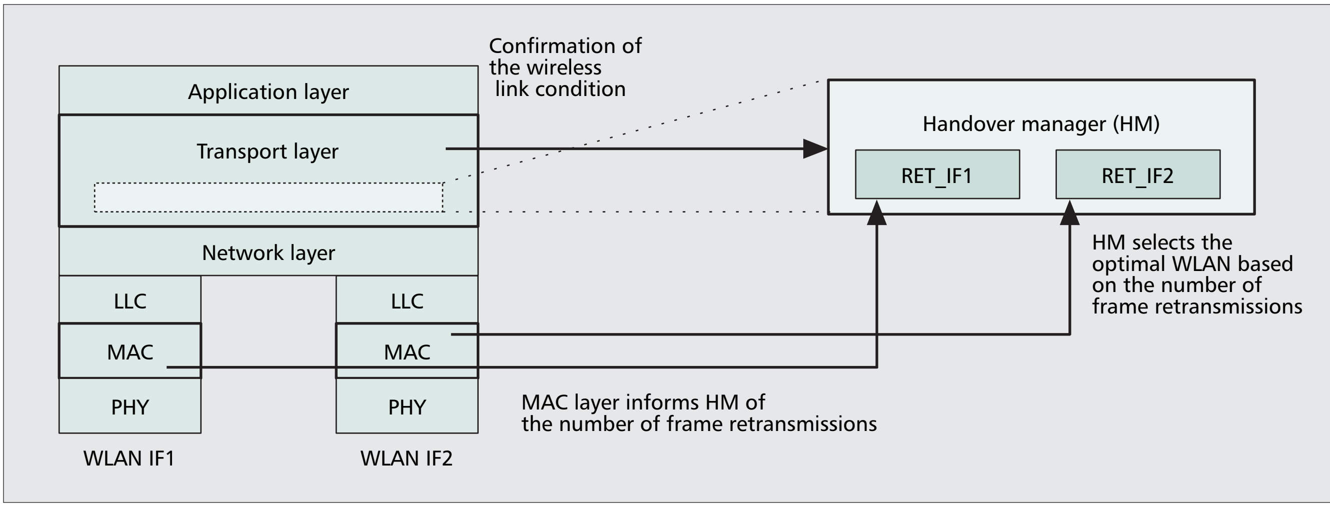
We next examine how radio interference affects RSSI, frame retransmissions, and application quality. However, since there is no space for a detailed discussion, we briefly explain the experimental results. The results demonstrate that TCP goodput drops drastically due to strong radio interference, while RSSI is not reduced at all. In contrast, the number of frame retransmissions increases in a radio interference environment. Therefore, we conclude that the number of frame retransmissions can be an optimal handover decision criterion to detect degradation of communication quality due to both 1 and 2 during communication. Note that when an MN transmits no frame, and just after an MN enters a WLAN, the RSSI is helpful and can be employed with the number of frame retransmissions, if necessary.

Advantages of Frame Retransmissions

The number of frame retransmissions has the following three advantages [9]:

- Detection of reduction of signal strength
- Detection of radio interference
- Ease of the threshold setting for handover

First, signal strength is degraded and fluctuates abruptly due to the increase of the distance from the AP and to any interfering objects. When this happens, a data packet experiences retransmissions before the occurrence of packet loss or the increase of RTT and jitter. Thus, an MN can promptly and reliably detect reduction of signal strength from the number of frame retransmissions. Next, degradation in communication quality due to radio interference cannot be detected from signal strength, because signal strength is not influenced at all by radio interference. On the other hand, frame retransmissions frequently occur due to collisions between transmitted frames in a radio interference environment. Therefore, an MN can detect degradation of communication quality due to radio interference from the number of frame retransmissions. Last, ease of the threshold setting for handover is noted here. As mentioned earlier, RSSI is measured in different ways by each vendor, so that it is extremely difficult to set an appropriate threshold for each WLAN card. On the other hand, as frame retransmissions can be handled



■ **Figure 3.** *Handover manager on the transport layer.*

in the same manner in all WLAN cards, we can simply set the same threshold by plain numbers (e.g., 1, 2, 3,…, *n*).

Handover Management Scheme

As mentioned earlier, seamless handover must meet three requirements. In this section, we propose a service-oriented mobility management scheme satisfying these three requirements when a handover manager (HM) is implemented on the transport layer and handles handover as illustrated in Fig. 3 [10]. Our proposed scheme can be applied only to both end hosts.

Initiation of Handover Processes Based on Prompt and Reliable Detection of Change in Wireless Link Condition

Our proposed scheme employs the number of frame retransmissions as a new handover decision criterion. Therefore, the MAC (medium access control) layer must inform the HM on the transport layer when an ACK frame is received or the number of data frame retransmissions reaches the retry limit. In traditional layer architecture, however, the information held in each layer cannot be accessed from different layers. We then propose a cross-layer approach that supports interaction between these layers. As illustrated in Fig. 3, in our concept of a handover management mechanism, the HM promptly and reliably detects the change in wireless link condition based on the number of frame retransmissions and executes handover processes.

Elimination of Communication Interruption Due to Handover Process

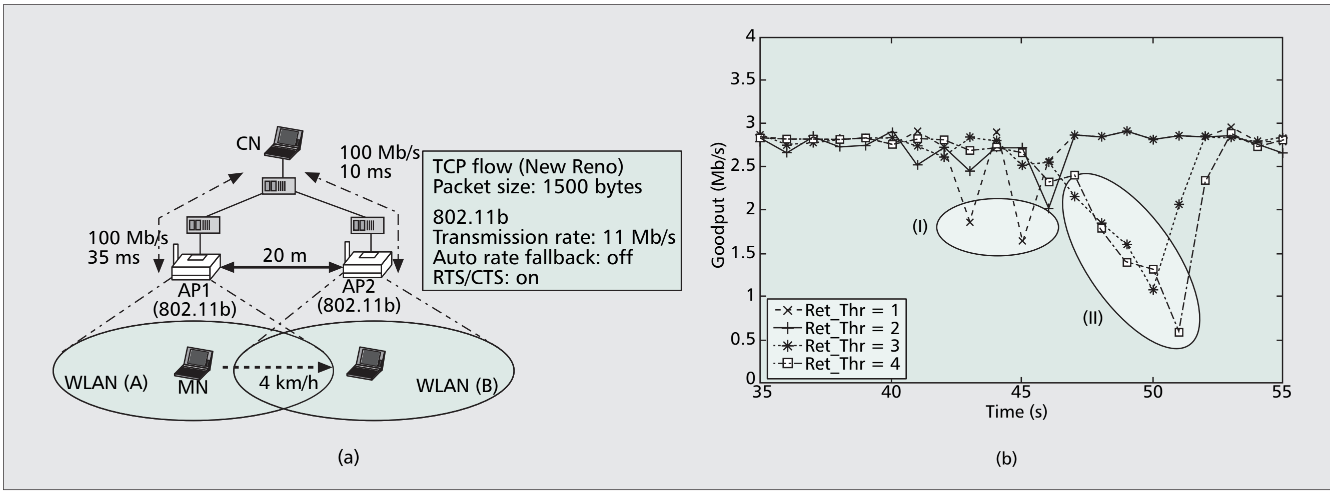
An MN cannot send or receive packets during layer 2 and 3 handover processes. To eliminate this interruption, an MN is equipped with multiple WLAN interfaces (i.e., a multi-homing approach). The multi-homing MN can thus preserve the communication by establishing an alternative connection with a new AP before handover.

Selection of an Optimal WLAN

We explain how an optimal WLAN is selected during handover in each application (non-real-time and real-time applications). First, we describe the common parts of both operations. As mentioned previously, an MN is connected with two different APs by using two WLAN interfaces in advance. The number of frame retransmissions is conveyed from each WLAN interface to the HM, and the latest information is recorded to the parameters (Ret_IF1 and Ret_IF2) on the HM (Fig. 3).

Next, we explain the handover management operation for FTP [11]. When the number of frame retransmissions on the current WLAN interface exceeds the predetermined retransmission threshold (Ret_Thr) on the HM, the HM detects deterioration of the wireless link condition and starts the handover processes. After that, the HM switches to multi-path mode and starts multi-path transmission by using two WLANs simultaneously. In multi-path transmission, by comparing the Ret_IF1 and Ret_IF2, the HM selects a WLAN with the smallest value as the optimal WLAN and then returns to single-path transmission. Note that because the optimal WLAN is selected by the result of only one packet transmitted from each WLAN interface in multi-path transmission, the network load due to multi-path transmission is extremely limited.

We finally explain the handover management operation for VoIP [10]. Whenever the HM receives the number of frame retransmissions, it compares the value (e.g., Ret_IF1) with the multi-path threshold (MPT). If Ret_IF1 exceeds MPT, the HM detects degradation in the wireless link condition and switches to multi-path transmission to prevent packet losses and to investigate the condition of the alternative WLAN. Because network load doubles in multi-path transmission, an operation by which to return to single-path transmission as quickly as possible is essential. Because packets are never retransmitted even if they are lost in real-time applications, the HM carefully selects an optimal WLAN, unlike FTP. The way to return to



■ **Figure 4.** *Simulation a) model; b) result for FTP communication.*

single-path transmission is as follows. We focus only on the operation for IF2 in multi-path transmission, because both of the WLAN interfaces perform the same operation. A packet may experience some retransmissions due to fluctuations of the wireless link condition even when handover is not required. Thus, to measure the stability of the wireless link condition, we provide a stability counter (SC) for each WLAN interface (SC_IF1, SC_IF2) and the single-path threshold (SPT) to return to single-path transmission on the HM. When the number of frame retransmissions on IF2 is zero (i.e., the sender successfully receives an ACK frame without any retransmissions), the HM increases SC_IF2 by one; otherwise the HM resets SC_IF2 to zero because it concludes that the wireless link condition is not yet stable. When SC_IF2 exceeds the SPT, the HM judges that the WLAN of IF2 has become stable and returns to single-path transmission. Through this mechanism, the HM can carefully select an optimal WLAN and prevent packet losses while properly switching between single-path and multi-path transmissions during handover.

Performance Evaluation

The performance of our proposed scheme is evaluated through simulation experiments. We implement the HM in Network Simulator Version 2 (v. 2.27). Our primary concerns are how the communication quality of each application can be maintained during handover by introducing our scheme.

FTP Communication

We examine the effect of the movement of an MN from WLAN(A) to WLAN(B), as illustrated in Fig. 4a. The MN first establishes a TCP connection with a CN via WLAN(A). Simulations are conducted for a period of 60 s, in which the MN located just under AP1 starts to move toward AP2 of WLAN(B) at 35 s. The MN moves at a walking speed of 4 km/h. The one-way delay to the CN for each WLAN is different, because each WLAN is assumed to be managed by different organizations (i.e., different IP subnets).

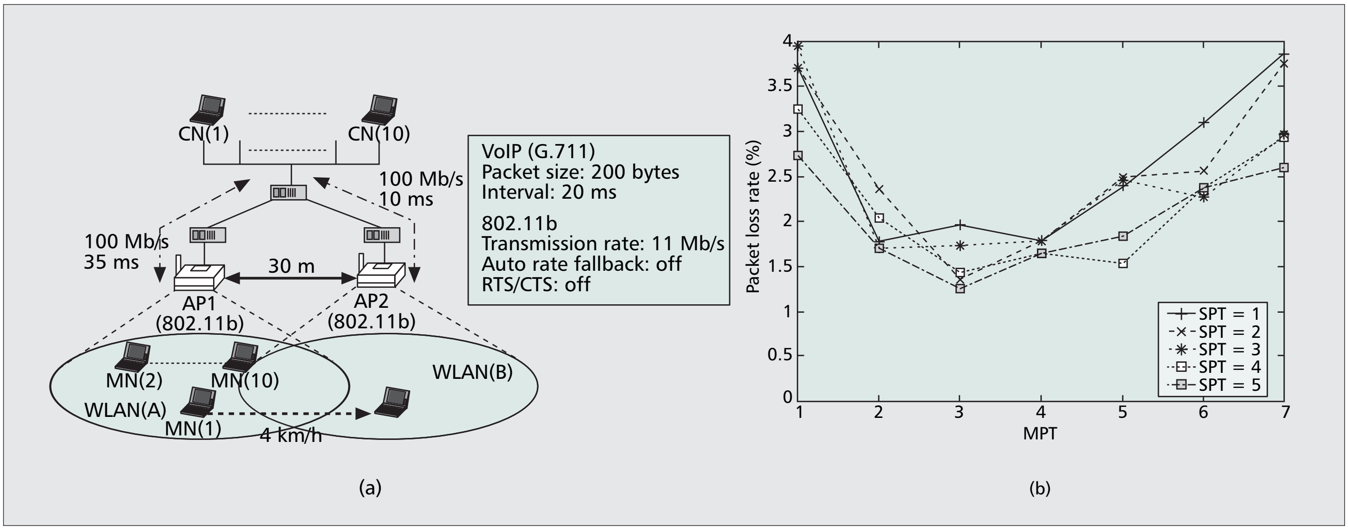
Figure 4b shows how the TCP goodput varies with the value of Ret_Thr, which is set from one to four. The MN begins to execute handover at around 43 and 45 s, when Ret_Thr is set to one. The MN starts the handover process before degradation of TCP performance actually occurs. Then, although the MN first selects WLAN(B) as the optimal WLAN, the optimal WLAN moves back and forth between the two WLANs several times until the condition of WLAN(B) becomes stable. The goodput performance degrades during this unstable period (I in Fig. 4b). In contrast, the MN begins to execute handover after 50 s, when Ret_Thr is set to three or four. As a result, the goodput decreases to quite a low value due to the long latency of the handover decision (II in Fig. 4b).

From these results, the MN can promptly detect deterioration of a wireless link condition when the value of Ret_Thr is too small, so that multiple handovers occur even though the condition of WLAN(A) has not yet worsened. On the contrary, when Ret_Thr is set to a relatively large value, although handovers occur infrequently, the TCP goodput drops drastically due to the long latency of the handover decision. Accordingly, we can see that Ret_Thr strongly affects the goodput performance during handover, so that Ret_Thr should be determined carefully.

In contrast, when Ret_Thr is set to two, the proposed scheme can promptly and reliably detect deterioration of the wireless link condition and can appropriately select the optimal WLAN. Therefore, our proposed scheme can maintain excellent TCP goodput performance even during handover.

VoIP Communication

Figure 5a illustrates the simulation model for VoIP. We explain only the difference from Fig. 4a. Each of the ten MNs in WLAN(A) executes VoIP communication with their ten CNs, after which MN1 moves from WLAN(A) to WLAN(B), while the other MNs do not move. All MNs and CNs send 200-byte voice packets, encoded by the G.711 codec, to each other at



■ **Figure 5.** *Simulation a) model; b) result for VoIP communication.*

20-ms intervals. The retry limit is seven because the VoIP packet size is smaller than the RTS threshold.

Figure 5b shows the average lost packet rate during handover. To maintain VoIP quality, the lost packet rate should be maintained at or below three percent [12]. From Fig. 5b, we can see that the lost packet rate depends on both MPT and SPT. In particular, MPT should not be set too small or too large. When MPT is set to one, the HM can sensitively and frequently switch to multi-path transmission in response to only one retransmission of a packet. The delay to the CN through WLAN(B) is smaller than that through WLAN(A), and there is no contention in WLAN(B). In such a case, a packet sent from WLAN(B) immediately after the start of multi-path transmission can arrive at the CN earlier than packets sent from WLAN(A) just before the start of multi-path transmission. Consequently, these late packets, sent from WLAN(A), are regarded as lost packets at the CN. Therefore, packet loss can frequently occur when the MPT is small. On the other hand, as WLAN(A) is shared by ten MNs and one AP, frames are very likely to wait in the interface buffer due to their contentions. Then, if we set MPT to seven, the MN continues to try to transmit the same data frame until its transmission fails seven times, and several frames are queued in the interface buffer during the waiting period. After that, when the HM switches to multi-path transmission, a packet sent through WLAN(B) arrives at the CN earlier than packets queued on IF1(WLAN(A)), thereby causing bursty lost packets. From these results, we can see that the occurrence of packet losses is drastically influenced by setting MPT and SPT. In [10], with MPT of three and SPT of two, an MN can move between WLANs with low packet loss rate and low additional network load (0.004 percent). Our proposed scheme can therefore achieve the communication quality required by VoIP during handover, while limiting the amount of redundant traffic due to multi-path transmission to an acceptable level.

Conclusion and Future Work

To achieve seamless handover irrespective of any kind of applications, in this article, we proposed a service oriented mobility management scheme. We first proposed the number of frame retransmissions as a new handover decision criterion to promptly and reliably detect changes in wireless link condition. We then showed that the number of frame retransmissions has the potential to serve as an optimal handover decision criterion for both non real-time and real-time applications, through experimental results. Next, we described the main concepts of our proposed mobility management architecture, in which MAC layer informs an HM on the transport layer of the number of frame retransmissions (cross-layer approach). In addition, to eliminate communication interruption during handover, we employed a multi-homing approach. In this architecture, the HM handles all handover processes, that is, handover initiation and selection of an optimal WLAN, based on the number of frame retransmissions. From simulation results, we showed that our proposed scheme can achieve seamless handover without degradation of FTP and VoIP application quality.

The implementation issue is of practical importance. We have implemented the prototype system for VoIP communication on Linux Kernel and presented a demonstration at MobiHoc 2006 [13]. Therefore, the feasibility of our proposed scheme was confirmed. In the next step, we plan to implement the prototype system for FTP communication and evaluate the resultant performance in real systems.

Finally, to successfully send packets over a wireless radio link, a sender in any wireless access network retransmits data frames when a data or ACK frame is lost. That is, the characteristics of frame retransmissions described in this article can be applied to other wireless access networks such as 802.11n and WiMAX. Therefore, a mobility management employing the number of frame retransmissions in any other wireless access network remains as a matter for further discussion.

Acknowledgments

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 Internal Affairs and Communications, Japan.

References

[1] IEEE 802.11, 1999 ed; http://standards.ieee.org/getieee802/download/802.11-1999.pdf
[2] C. Perkins, "IP Mobility Support for IPv4," IETF RFC 3344, Aug. 2002.
[3] M. Riegel and M. Tuexen, eds., "Mobile SCTP," IETF Internet draft, draft-riegel-tuexen-mobile-sctp-07.txt, Oct. 2006.
[4] R. Koodli, "Fast Handovers for Mobile IPv6," IETF RFC 4068, July 2005.
[5] H. Soliman *et al.*, "Hierarchical Mobile IPv6 Mobility Management (HMIPv6)," IETF RFC 4140, Aug. 2005.
[6] K. Muthukrishnan *et al.*, "WLAN Location Sharing through a Privacy Observant Architecture," *Proc. COMSWARE '06*, Jan. 2006, pp. 1–10.
[7] K. Tsukamoto *et al.*, "Impact of Layer 2 Behavior on TCP Performance in WLAN," *Proc. IEEE VTC '05 Fall*, CD-ROM, Sept. 2005.
[8] K. Tsukamoto *et al.*, "Experimental Evaluation of Decision Criteria for WLAN Handover: Signal Strength and Frame Retransmission," *Proc. UCS '06*, LNCS 4239, Oct. 2006, pp. 239–53.
[9] S. Kashihara *et al.*, "A Simple Heuristic for Handover Decisions in WLANs," IETF Internet draft, draft-shigeru-simple-heuristic-wlan-handover-02.txt, Mar. 2006.
[10] S. Kashihara and Y. Oie, "Handover Management Based on the Number of Retries for VoIP on WLANs," *Proc. IEEE VTC '05 Spring*, CD-ROM, May 2007.
[11] K. Tsukamoto *et al.*, "Handover Management based on the Number of Frame Retransmissions for TCP over WLANs," *Proc. CCNC '07*, Jan. 2007.
[12] A. Schmitter *et al.*, "Analysis of Network Conformity with Voice over IP Specifications," *Proc. ISSC '03*, July 2003, pp. 82–86.
[13] S. Kashihara *et al.*, "Handover Management Based on the Number of Frame Retransmissions for VoWLANs," *MobiHoc '06 Demo Session*, May 2006.

Biographies

SHIGERU KASHIHARA [M] (shigeru@is.naist.jp) received his D.E. degree in information science from Nara Institute of Science and Technology (NAIST), Japan in 2003. He is currently an assistant professor in the Graduate School of Information Science, NAIST. His research interests include ubiquitous computing and networks. From January 2004 to February 2005 he was a postdoctoral researcher at Kyushu Institute of Technology (KIT), Japan. He is a member of the IEICE.

KAZUYA TSUKAMOTO [M] (kazuya@infonet.cse.kyutech.ac.jp) 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 IEICE.

YUJI OIE [M] (oie@cse.kyutech.ac.jp) received B.E., M.E., and D.E. degrees from Kyoto University, Japan, in 1978, 1980, and 1987, respectively. Since April 1997 he has been a professor in the Department of Computer Science and Electronics, Faculty of Computer Science and Systems Engineering, Kyushu Institute of Technology (KIT), Japan. His research interests include performance evaluation of computer communication networks, high-speed networks, and queuing systems. 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. He is a fellow of the IEICE and IPSJ.