Adaptive Handover Method with Application-Awareness for Multimedia Streaming Service in Wireless LAN

Jae-Wook Nah, Student Member, IEEE, Sung-Man Chun, Song Wang, Student Member
Jong-Tae Park, Senior Member, IEEE

School of Electrical Engineering and Computer Science, Kyungpook National University, 1370 Sankyuk-Dong, Buk-gu, Daegu 702-701, Korea
jwnah@ee.knu.ac.kr

Abstract—In this paper, we present an adaptive handover method with application-awareness for multimedia streaming service in wireless local area networks. Traditionally, the handover method in wireless local area network did not take into account the characteristics of the application service. Thus, the severe service disruption may occur during the movement of mobile node. In order to prevent the degradation of the applications service, the scanning period and the number of channels per scanning for handover are dynamically adapted to the type of applications and wireless networking environment. The performance of the proposed adaptive handover mechanism has been evaluated by simulation. The simulation results show that the proposed mechanism the proposed mechanism can prevent service quality degradation and provide session connectivity.

Index Terms—Adaptive handover, Application-aware handover in WLAN, Cross-layer optimization

I. INTRODUCTION

THE IEEE 802.11 wireless local area network (WLAN) [1] has widely deployed throughout the world due to its low cost and high bandwidth. While the elastic data traffic such as Web and e-mail currently constitutes the bulk of Internet traffic carried over the WLAN, the real-time multimedia service such as VoIP, video/audio streaming, video conferencing, and IPTV is becoming more common due to the recent development of IEEE 802.11 n standard [2]. However, the limited coverage of IEEE 802.11 WLAN incurs the service disruption or quality degradation while the users are moving around between the access points of the WLAN. This is because the small service range of the access points (APs) of WLAN makes the mobile host to do frequent handover between different cells, and the handover latency of the IEEE 802.11 standard is about several hundred milliseconds. This latency is too long for the real-time multimedia applications such as VoIP and multimedia streaming services.

The aforementioned 802.11 handover latency problem has led to substantial research interest in providing fast handover for real-time multimedia traffic in WLAN. The previous related includes proactive scan [3], smooth scan [4], and the IEEE is also developing international standard which includes the recent IEEE 802.11 k [5] and r [6] standards. Most of these research works make use of make-before-break principle in which the neighboring AP information is obtained in advance before the handover occurs. However, the focus of these previous works has been mainly on reducing the layer 2 handover delay, without considering the additional power consumption overhead due to the scanning operation. However, since the mobile device has limited power, the energy-efficiency in mobile handover is very important issue [7]. Especially, the power consumption is key issues when considering voice service over 802.11-based system [8].

In this article, we present power-efficient layer 2 handover mechanism in wireless LAN, which can provide fast handover with guaranteed QoS for real-time multimedia service. To this end, the scanning period and the number of channels per scanning for handover are dynamically adapted to the type of applications and wireless networking environment, so that both the degradation of application service quality and signaling overhead are minimized simultaneously. To our knowledge, almost no published work exists on dynamically adapting the layer 2 handover procedure to the application type for guaranteeing the quality of service (QoS) of the diverse real-time multimedia applications.

The salient feature of the proposed scheme is that the fast handover is accomplished without degradation of the...
multimedia application service during handover procedure in such a way that it minimizes both packet loss and signaling power consumption overhead. The model and parameters have been designed to specify the QoS constraints of the multimedia application. Extensive simulations have been performed to illustrate the efficiency of the proposed approach.

The rest of this paper is organized as follows. In Section 2, we introduce related work. Section 3 describes the handover algorithm proposed by the authors, and Section 4 presents the simulation result. And finally, in Section 5, we conclude the paper.

II. RELATED WORK

In wireless LAN, the handover procedure consists of the four steps: channel discovery, channel switching, authentication, and association. Among these steps, the channel discovery is known to be a dominant factor for causing the handover latency, taking up almost 90% of the handover procedure [9]. There have been diverse efforts to reduce the handover latency of channel discovery. The research attempts can be largely classified into two categories; active scanning and passive scanning. In the active scanning mode, the mobile node sends a probe request message to the neighboring APs to receive the received radio signal strength indicator (RSSI) or signal to noise ratio (SNR) from the APs. In the passive scanning mode, the mobile node gets the handover information from the beacon signals which are periodically broadcasted by the APs. Since the broadcasting period of the passive mode is too large, about 100 milliseconds, providing the fast handover for the real-time multimedia application services are usually based on the active scanning method [9]. The end-to-end delay for VoIP service should be less than 400 milliseconds [10]. In IEEE 802.11 standard, there are 11 (or 13) channels and the time to scan a channel is between 20 milliseconds and 60 milliseconds [9]. So, the total channel access time is about 200 milliseconds, which seems to be too large for the delay-sensitive multimedia application like VoIP, which may result in severe service disruption. In passive scan, there appeared a fast handover mechanism called "Sync Scan", which achieve fast handover by synchronizing the scanning time of all APs [11]. However, the AP synchronization is easy to be found.

In active scanning, the mobile node finds out in advance the neighboring AP information during movement in order to reduce the wireless channel access time. The representative active scanning method is that of IEEE 802.11k standard, which makes use of the neighbor report and selective scanning method. The neighbor report contains the accessible AP information for an AP. In the selective scanning method, the channels in which APs are accessible are only scanned, and the optimal AP which provides the best performance with respect to the signal strength (and loading condition) is selected, so that the scanning time can be reduced. However, since this requires the modification of the APs, it may be difficult to widely implement this technique to the existing wireless LAN environment. Furthermore, this technique may not be suitable for the indoor and/or metropolitan area in which APs are densely installed.

Recently, there appeared other fast layer 2 handover techniques which do not require the modification of the existing IEEE 802.11 standard. These include proactive scanning [3], smooth scanning [4]. Both of these two methods employ the pre-scanning technique in which the neighboring AP information is gathered in advance during movement, and the APs are rapidly switched when the handover condition is satisfied. These methods however may incur the unnecessary large power consumption for handover preparation, and also non-negligible packet loss. Moreover, these methods do not reflect the QoS constraint of the streaming multimedia application during handover, which is described subsequently in more detail. In summary, most previous work on layer 2 handover operation do not reflect the varying QoS constraint of the multimedia application and network conditions, so that the quality of the application service during handover preparation and execution can be severely degraded, and the excessive additional power can be consumed for the handover operation.

Cross layer design is a new paradigm in network architecture design than takes into account the dependencies and interactions among layers, and supports optimization across layer boundaries [12][13]. Most of previous work on cross layer design for multimedia streaming service has mainly focused on optimizing the QoS parameters of applications and/or MAC protocol parameter for data transmission. For example, in [14] a cross-layer scheduling framework is proposed with adaptive modulation and coding, and in [15][16] application layer adaptation mechanisms are combined with lower-layer data transmission mechanism such as error correction protocol for low-delay wireless video streaming. Within our knowledge, there have been no previous attempts to cross layer design in which the QoS parameters of application layer are utilized for the layer 2 signaling operation for guaranteed QoS of the multimedia application.

III. SYSTEM AND MODEL FOR APPLICATION-AWARE HANDBOVER

A. Background of system for Application-Aware Handover

Fig. 1 shows the variations of RSSI, Handover Prepare Threshold, Tp, and Handover Execute Threshold, Ts, as the mobile nodes moves from the current AP to the next AP. In this paper, the access point (AP) to which the mobile node currently is attached is called the current AP, and the access point to which the mobile node will be attached after handover is called the next AP. In the proposed application-aware algorithm, as the mobile node moves, if the RSSI value measured from the current AP becomes less than Ts and that from the next AP is beyond Tp, the mobile node rapidly changes the point of attachment from the current AP to the next AP. The more detailed algorithm is described subsequently.

In the active scanning, the mobile node periodically
transmits the probe message to the near-by APs in order to get the
neighboring AP information. The response message form the AP includes the SSID and MAC address of the AP, signal
strength, and other information related to the authentication.
The mobile node also gets the RSSI value form the current AP. Sometimes signal to noise ration (SNR) values is used instead of RSSI. SNR (Signal to Noise Ratio) without scanning.

In the proposed algorithm, if the RSSI value from the current AP is beyond $T_p$, the mobile node tries to find the APs which are located near to itself. If the signal strength from the current AP becomes very weak, it can then change rapidly the point of attachment to one of the neighboring AP which provides the best condition. In other words, the mobile node uses the make-before-break principle which is often used in other restorable network technology.

**B. Application-Aware Handover Model for Real-Time Multimedia Streaming Service**

As described before, there have been no previous research attempts on layer 2 handover mechanism, which adjust scanning intervals to guarantee the QoS of multimedia applications. Actually, the service quality of the streaming service is in most cases acceptable to users although the layer 2 connection is disconnected for a few seconds due to the temporary channel switching of pre-scanning phase. Therefore, the service quality of the multimedia streaming application service may not be degraded even though the layer 2 is intermittently disconnected due to the periodic pre-scanning operations. This is because the application buffer can store enough data to sustain the continuous delivery of the multimedia traffic for the duration of discontinuity of the layer 2.

![Fig. 1. Variations of RSSI as the mobile node moves](image)

**Definition 1:** The interrupt allowance time (AIAT) of an application service is defined as the time during which the streaming application can be played out without service degradation although the layer connection is disrupted due to scanning operation of the handover procedure.

**Definition 2:** $Q(t)$ is defined as the amount of pre-fetched data in the application buffer at time $t$.

In this section, we present the service model for application-aware handover. Fig. 2 depicts the buffering structures at mobile node for multimedia streaming application. The mobile node starts to play out the streaming data after some amount of data is pre-fetched in the buffer as shown in Fig. 2. After playing out the service, the data is continuously get into the buffer from the network, with rate of $F(t)$, so the service can be provided without disruption. Here, $F(t)$ is called Fill Rate, and it indicates the amount of data which arrives at the application buffer during unit time. In Fig. 2, $D$ is the amount of data which leaves from the buffer during time unit. $F(t)$ is termed as input rate, and $D$ is called output rate. It is noted that $F(t)$ varies along with time, being dependent of the network conditions, delay and bandwidth, but $D$ is fixed and independent of time.

**Definition 2:** $Q(t)$ is defined as the amount of pre-fetched data in the application buffer at time $t$.

In Table 1, the maximum application interrupt times for various types of real-time multimedia applications are described. The store video or audio applications may take 2 to 5 seconds of application interrupt allowance time, which are usually RTSP streaming protocol. For VoIP using G. 711 codec, the application interrupt allowance time is assumed to take 50 milliseconds. For real-time video conferencing service, it is also assumed to take 50 milliseconds.

As shown in Fig. 2, the multimedia application using streaming protocol is played out after the amount of data filled up at the application buffer reaches some threshold value $Q$. Therefore, even though the connections at the layer 2 is disrupted, the application can be continuously played out without service disruption until the pre-fetched data, i.e., $Q$ bytes of data, in the application buffer is drained out. Therefore, the maximum of application interrupt allowance time for a multimedia application using streaming protocol is $Q/D$ sec. Let us denote this $Q/D$ time as $T$.

**Lemma 1:** In wireless LAN, the maximum value of AIAT of an application at a mobile node is $T$.  

![Fig. 2. Application buffer structure for multimedia streaming service](image)

**TABLE I**

<table>
<thead>
<tr>
<th>Application type</th>
<th>Protocol</th>
<th>AIAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stored video/audio multimedia streaming</td>
<td>RTSP</td>
<td>2sec-5sec</td>
</tr>
<tr>
<td>VoIP</td>
<td>RTP, G.711</td>
<td>50ms</td>
</tr>
<tr>
<td>Video conferencing</td>
<td>RTP, H.323</td>
<td>50ms</td>
</tr>
</tbody>
</table>
Proof: For a given time $t$, there exists $Q(t)$ data at the application buffer. Since the streaming application can be at most played out without service degradation during $Q(t)/D$, $T$ is the maximum value of $A_{1\text{AIT}}$.

Lemma 2: In wireless LAN, the power consumption for handover procedure is proportional to the number of scan operation.

Theorem 1: As a mobile node moves in a wireless LAN, the handover procedure with minimal power consumption while sustaining the service quality of an application is performing one scan operation during $T$ sec.

Proof: By Lemma 1, an application at a mobile node can at most play out the service without service degradation during $T$ sec, regardless to the number of scan operations of the layer 2 handover procedure. According to Lemma 2, since the power consumption for handover procedure is proportional to the number of scan operation, performing one scan operation during $T$ sec results in the minimal power consumption for handover procedure while sustaining the service quality of the application.

Corollary 1: As a mobile node moves in a wireless LAN, the handover procedure for getting the best neighboring AP information with minimal power consumption while sustaining the service quality of an application is performing one scan operation during $T$ sec.

Lemma 3: Accessing one wireless channel during each scan operation results in minimum packet loss for the layer 2 handover procedure.

Proof: Since the packet loss due to the layer 2 handover procedure is proportional to the time to access a channel of the wireless LAN, it is also proportional to the number of wireless channels which is accessed during one scan operation. Therefore accessing one wireless channel during each scanning operation results in minimum packet loss.

Up to now, we have done some theoretical analysis for performing optimal handover procedure with respect to power consumption, service quality, and packet loss. However, in reality, $Q(t)$, i.e., the input rate of the application buffer, depends on the network conditions. So, it is desirable to make the scan period to be adaptively changed in accordance with the network conditions, in order to minimize both the power consumption and packet loss while preserving the service quality during handover. In a dense networking environment such as indoors and large metropolitan area, the small scan granularity is desirable. This is because the sufficient AP information can usually be obtained even though a few channels are only accessed during a scan operation in dense networking environments.

In Fig. 3, we show the application-aware handover algorithm which can guarantee the service quality with minimal power consumption. The operation of application-aware handover algorithm in Fig. 3 is as follows. It consists of three phases; Searching of neighboring AP information, Handover preparation, and handover execution. These phases are determined by two signal strength threshold values $T_p$ and $T_s$, where $T_p$ is called the handover preparation threshold, and $T_s$ is called the handover execution threshold. First, during initialization, the algorithm decides the values of $T$, $Q$, and $G$ where $T$, $Q$, and $G$ are the maximum of application interrupt allowance time, application data to play out, and scan granularity, respectively. These values are determined by the type of application.

During the phase of searching the neighboring AP information, if the moving average value of RSSI from the current AP is greater than $T_p$, the mobile node first scans the $G$ number of channels, getting the neighboring AP information, and stores them into Neighbor_AP_List file. Next, it waits for another $T$ seconds, and start scanning operation again for another $G$ number of channels, if the amount of pre-fetched data is larger than or equal to $Q$ bytes. These steps are repeated until all the channels are scanned. After all the channels are scanned, it starts scanning operation again from the first set of $G$ channels. However, for each scanning period, if the amount of data filled up at the application buffer is less than $Q$, the scanning period is prolonged until the buffer is filled up sufficiently enough to preserve the service quality of the application. In this way, during the searching phase, the service quality of an application is preserved with minimal power consumption due to handover operation.

Procedure adaptive_handover($T$, $Q$, $G$, $T_p$, $T_s$)

Begin
Step (1) Identify type of application;
  Determine $T$ and $Q$;
  Determine $G$;

Step (2) Wait until the application buffer filled up to $Q$;

Step (3) Scan $G$ number of channels and store their RSSI values in Neighbor_AP_List;

Step (4) Check the Current AP’s RSSI, IF the RSSI is bigger than $T_p$ goto Step(2) ELSE goto Step(4-1);
  Step(4-1) Select Candidate_AP_List from Neighbor_AP_List, IF no candidate AP goto Step(3) ELSE goto Step(4-2);
  Step(4-2) update RSSI values of APs in Candidate_AP_List; Adjust $G$;
  Step(4-3) Check the Current AP’s RSSI, IF the RSSI is bigger than the $T_s$, goto Step(3) ELSE goto Step (5);

Step (5) Check the current communication application CASE: VoIP, Select the best RSSI AP;
  CASE: Multimedia conference, Scan the Best AP;
  CASE: Stored video/audio, Selected scanning;

Step (6) Authentication and association with the best AP;
End

As the moving average value of RSSI becomes less than $T_p$ due to the movement of the mobile node, the mobile node gets into the handover preparation phase. During the handover preparation phase, higher priority is put on finding optimal next
AP and performing fast handover. So, the mobile node first selects candidates for next AP from the Neighbor_AP_List, and stores them into Candidate_AP_List. The APs with good AP/Channel load, and whose RSSIs are greater than Ts are selected for candidate APs. AP/Channel load conditions mean the channel load and traffic load which are defined in IEEE 802.11 k standard. The scan granularity G is adjusted to achieve fast handover with minimal packet loss. If the RSSI from the current AP is greater than Ts, the mobile node performs, with reduced scanning period, the scan operation for the APs in Candidate_AP_List, and updates the candidate AP information. In this way, both the scan period and scan granularity are getting smaller, so that the mobile node can switch to the optimal next AP with minimal handover latency.

IV. SIMULATION AND PERFORMANCE EVALUATION

A. Simulation Configuration

In this section, we have evaluated the performance of context-aware mobility management by simulation. In particular, we have compared the performance of the application-aware fast handover with that of existing layer 2 handover method based on the IEEE 802.11 standard. Since the existing simulator such as NS2 does not provide mechanism for measuring the power consumption, we have written a simulator with C++ language, which runs under Window XP operating system. Additional benefit resulting from writing a simulator is that we can simulate the performance of the cross-layer optimization mechanism to model the application type, and the pre-scanning and partial scanning methods at layer 2. The simulator gets the scan granularity, scan cycle (T-T/N), application service Q(t) and Q value, and power consumption as input values of the application-aware handover algorithm in Fig. 3, and investigates the tradeoff between scanning delay and power consumption for guaranteed QoS of various type of medical application service.

For simulation, we have placed the four APs in the 500m x 500m area. APs are placed on the following coordinates: (50, 50), (50, 350), (350, 50), and (250, 350). It is assumed that the mobile node moves in the diagonal direction from (0, 0) to (500, 500). The data transmission rate is assumed to be a maximum of 54 mbps of the IEEE 802.11g standard. The model of calculating the radio signal strength as the mobile node is in movement is obtained by using the formula in [17], which is determined by the distance between a mobile node and an AP. And the movement of the MN follows below function Y=aX.

B. Simulation Result

As described earlier, in WLAN, the handover latency occurs at three operational procedures: scanning, authentication and association. Since the scanning takes most time of the handover latency, the pre-scanning and partial scanning technique is employed in the application-aware handover algorithm to achieve the optimality in both handover latency and power consumption.

Fig. 3. Handover latency of IEEE 802.11 standard and application-aware handover

Fig. 5. Power consumption of IEEE 802.11 standard and AIAT 50ms
made large for maintaining the service quality, since the AIT is power consumption also increases, and vice versa. However, if about a few seconds.

On the contrary, the handover latency of those obtained by the proposed application-aware mechanism exhibits the varying performance in accordance with the type of the application. For example, the stored video service with the value of a few seconds of AIAT generally has the larger value of handover latency than those of EEG and ECG. In summary, the application-aware demonstrate better performance in handover latency than the IEEE standard mechanism. Another fact is that the handover latency is also dependent on the type of the medical application, so that if the characteristics of application reside within the scope of some AIAT value, the QoS-guaranteed service can be achieved.

Fig. 4 shows the average handover latency for each application service. As shown in Fig. 4, in the case of the IEEE 802.11 standards handover had hundreds of ms latency irrespective of the type of application, but the application-aware handover had different handover latency depending on the type of application. This is because it dynamically performs the scanning operation, depending on the application types and buffer state of applications. C. Evaluation of Power consumption

The power consumption generated by using the partial scanning is generally larger than that of the full scan operation, for the case that the entire channels are scanned. This is because each scan operation, whether it is partial or not, may need some overhead for the initiation of scan operation. In the algorithm of the application-aware mobility management, the partial scan interval is dynamically adjusted to preserve the QoS constraint of an application. As the partial scan operation increases, the power consumption also increases, and vice versa. However, if the frequency of partial scanning operation decreases below some threshold value, the handover latency abruptly increases, which may result in severe degradation of the quality of application service. For VoIP, since the AIAT is very small, the interval of the partial scanning operation can be made very small to guarantee the service quality. For the stored video application, the interval of the partial scanning operation can be made large for maintaining the service quality, since the AIT is about a few seconds.

In Fig. 5, the upper-most graph shows the power consumption generated by the full scan operation, employing the IEEE standard handover mechanism. It shows that the power consumption for the full scan operation is in the range of about 0.07 mW and 0.1 mW. Power consumption has some relation with the elapsed time for the scan operation, which is determined by minimum channel waiting time, the maximum channel waiting time, and the total number of channels, and the channel switching time.

The lower graph in Fig. 5 shows the power consumption for various applications which have the AIAT value of 50 ms, in which the application-aware handover algorithm is applied. As described before, the application-aware handover algorithm takes advantage of both pre-scan and partial scan mechanisms in order to guarantee the QoS of an application. For the application with small AIAT, the partial scan is more frequently executed to prevent the degradation of the QoS of an application service. This may incur more frequent switching between scan mode and data transmission mode at WLAN interface card, so that the power consumption increases due to this frequent switching operation. For the value of 50 ms AIAT, the range of consumed power is between 0.23 mW and 0.33 mW, as shown in Fig. 9. In comparison with that of IEEE standard scanning method for the full scan operation, the power consumption is almost three-fold. However, it gives an advantage that the quality of the application service is not degraded.

Fig. 6 compares the average power consumption in the IEEE standard handover and application-aware handover. As shown in Fig. 6, the power consumption of the IEEE standard handover mechanism is the least. This is because it performs the handover operation using a full scanning, regardless of the type of an application. On the other hand, the power consumption which is generated by using the application-aware handover varies in accordance with the type of application. The application which consumes the largest power is the application with the smallest value of AIAT. This is because the smaller the value of AIAT is, the bigger the power consumption due to more frequent execution of partial scans operation. Fig. 6 also shows the other case in which the larger value of AIAT results in the smaller value of the power consumption. However, the larger AIAT is, the larger handover latency is. Therefore, there exists a trade-off among the values of AIAT, handover latency, and power consumption. By adjusting these values, the scanning frequency, scanning period, and the number of channels per a partial scan operation can be obtained to achieve the optimality in handover.

V. CONCLUSION

In this article, we present adaptive fast handover algorithm with application-aware for multimedia service in wireless LAN. And we also consider about the application and power consumption. This The most benefits of our algorithm are that it provides seamless handover for VoIP, multimedia conference
service, and stored media service with considering both QoS and power consumption. The performance of the simulation shows that the proposed mechanism is useful for the wireless local area network.

REFERENCES


