Context-Aware Mobility Management with Energy Efficiency for Multimedia Streaming Service in Wireless LAN

Jong-Tae Park, Jae-Wook Nah, Song Wang and Seong-Man Chun
School of Electrical Eng. and Computer Science
Kyungpook National University
Daegu, 702-701 Korea
jtpark@ee.knu.ac.kr, jwnah@ee.knu.ac.kr, swang@ee.knu.ac.kr, and smchun@ee.knu.ac.kr

Abstract— **Providing handover with guaranteed QoS for real-time multimedia service such as VoIP, video/audio streaming, and video conferencing in wireless LAN is a challenging task. In this paper, we present a context-aware mobility management 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 application context, so that both the degradation of application service quality and energy consumption are minimized simultaneously. The performance of the proposed mechanism is verified by the simulation. The simulation results show that the proposed mechanism is useful for the personalized network which is comprised of wireless LANs.**

Keywords- Context-aware mobility management, Fast handover in WLAN, Cross-layer optimization

I. Introduction

The IEEE 802.11 wireless local area network (WLAN) 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. 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 [1], smooth scan [2], and the IEEE is also developing international standard which includes the recent IEEE 802.11 k and r standards [3]. 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 [4]. Especially, the energy or power consumption is key issues when considering voice service over 802.11-based system [5].

In this article, we present context-aware mobility management 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 context of the multimedia service. The context includes the he type of applications and wireless networking environment.

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 context. 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 model and algorithm of the mobility management 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

978-1-4244-2309-5/09/$25.00 ©2009 IEEE

handover latency, taking up almost 90 % time of the handover procedure [6]. 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 ration (SNR) from the APs. In the passive scanning mode, the mobile node gets the handover information from the becon 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 [6]. The end-to-end delay for VoIP service should be less than 200 milliseconds [7]. 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 [6]. 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 [8]. 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.

Cross layer design is a new paradigm in network architecture design which takes into account the dependencies and interactions among layers, and supports optimization across layer boundaries [9, 10]. 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. 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 AP scanning operation in order to guarantee the QoS of the multimedia application in WLAN systems.

III. Model and Algorithm for Context-aware Mobility Management

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. The pre-scanning operation to find out the neighboring AP information can be performed without any degradation of the service quality of the multimedia application, if the pre-scanning interval is properly adjusted.

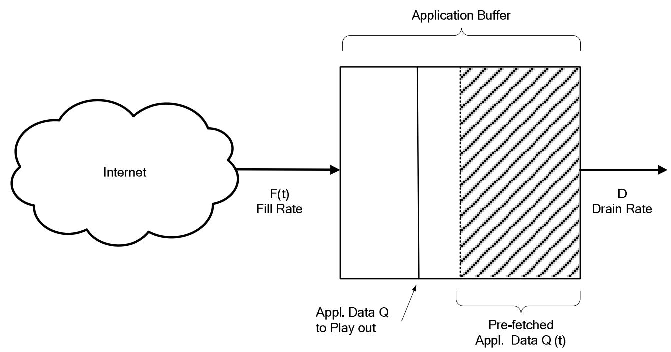

Figure 1. Application buffer structure for multimedia streaming service

In this section, we present the service model for context - aware mobility management. "Fig. 1" shows 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. 1". 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. 1", 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 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 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 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.

TABLE I. QOS REQUIREMENT OF VARIOUS APPLICATION CONTEXT

Application type	Protocol	AIAT
Stored video/audio multimedia streaming	RTSP	2sec-5sec
VoIP	RTP, G.711	50ms
Video conferencing	RTP,H.323	50ms

As shown in "Fig. 1", 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.

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 AIAT. □

In wireless LAN, the energy 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 energy 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. 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. □

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.

Theorem 2: 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 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, access of 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. 2", we show the context-aware mobility management algorithm which can guarantee the service quality with minimal energy consumption. The operation of context-aware mobility management algorithm in "Fig. 2" 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 Tp and Ts, where Tp is called the handover preparation threshold, and Ts 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 of RSSI value from the current AP is greater than Tp, the mobile node first scans the G

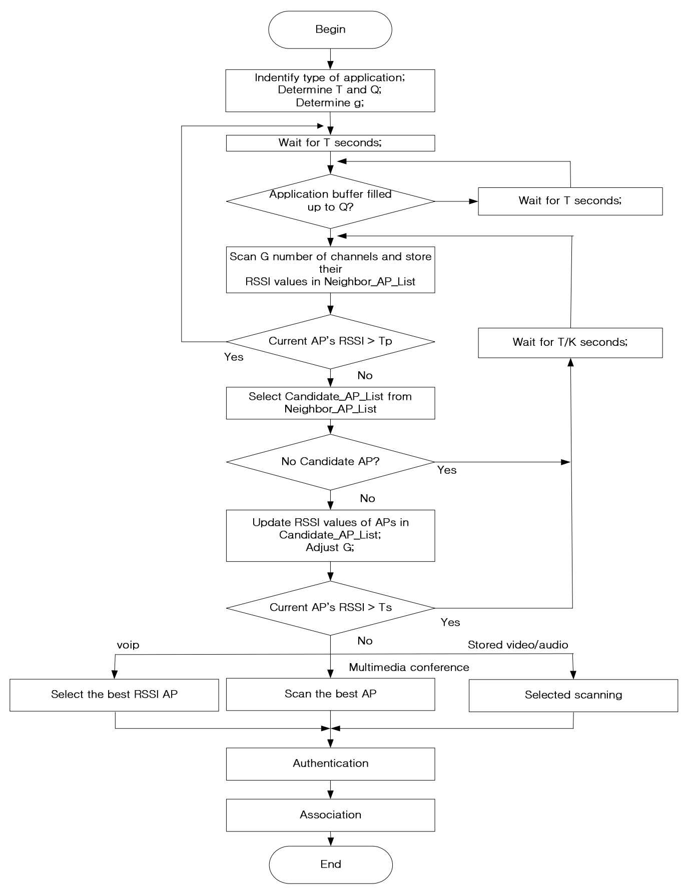

Figure 2. Context-aware handover algorithm with guaranteed service quality and minimal power consumption

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.

As the moving average value of RSSI becomes less than Tp 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 context-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 energy 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 energy consumption as input values of the context-aware handover algorithm in "Fig. 2", 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×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.

In Table 2, we describe the simulation parameters used for the simulation. In Table 2, the values of minimum channel waiting time and maximum channel waiting time for the active scan operation are set to 20 ms and 40. The channel switching time is assumed to be 5 ms.

B. Simulation Result

We supposed that the user used general internet service. The services are FTP, music streaming service and network diagnosis utility. Network diagnosis utility is popular because many user wants to know their network condition.

"Figs. 3, 4 and 5" show the latencies of the FTP, music streaming application and real-time network monitoring application traffics due to handover operations. The latencies

TABLE II. Simulation Parameter

Parameter	Setup Value
Area	500 m × 500 m
Number of AP	4
Coverage of an AP	50 m
Movement Speed of Mobile Node	1.5 m/s
Minimum Channel Waiting Time	20 ms
Maximum Channel Waiting Time	40 ms
Channel Switching Time	5 ms

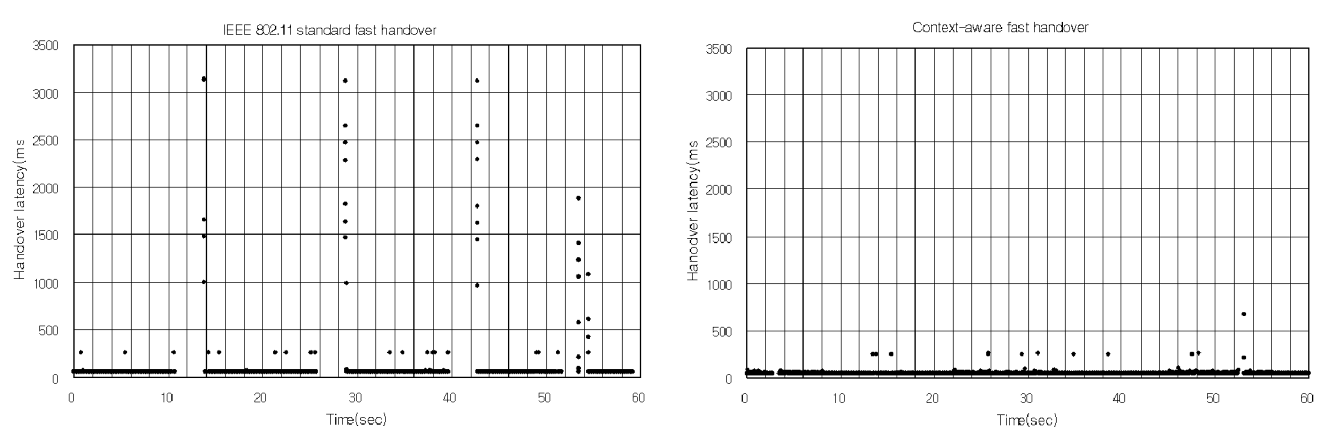

Figure 3. Handover lantency of FTP application contex

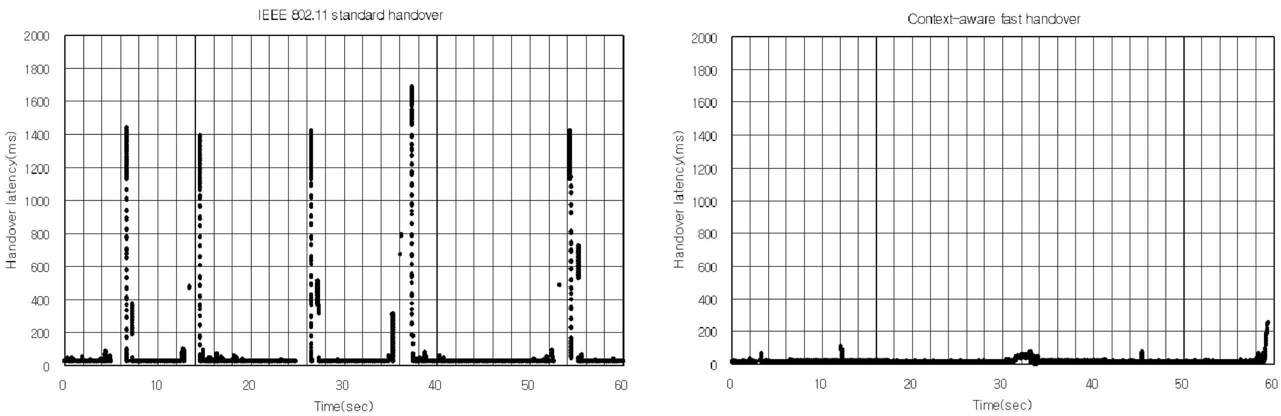

Figure 4. Handover lantency of music streaming application context

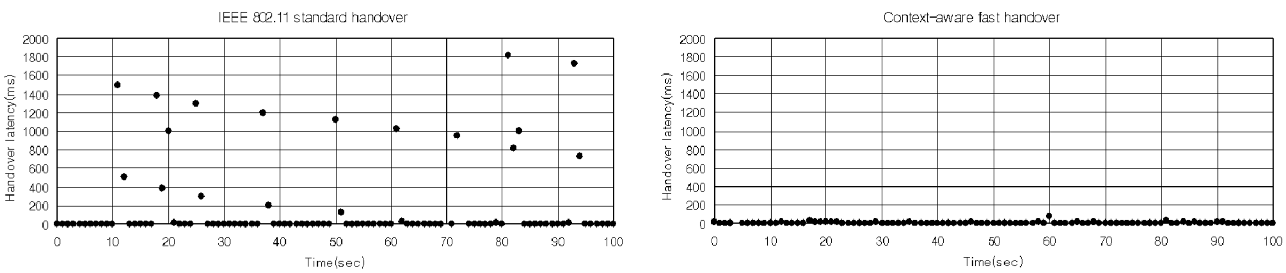

Figure 5. Handover latency of real-time network diagnosis application context

resulting from the context-aware handover are compared with those resulting from the standard IEEE 802.11 handover procedure.

"Fig. 3" shows the handover latency concerning FTP traffic. The delay time is greater than 3000 ms by the IEEE 802.11 standard handover procedure. This is due to the re-transmission mechanism of TCP protocol operation of FTP application for the lost packet recovery after the handover occurs. On the contrary, for the case of the context-aware handover mechanism, the delay time is less than 50ms in most simulation experiments. This is because the active scanning latency in IEEE 802.11 standard is eliminated in the case of context-aware handover mechanism.

"Fig. 5" shows handover latency concerning audio streaming application in which the user listen to the music player playing on the WEB. Generally, audio streaming services used UDP protocol. In UDP traffic, there is no re-transmission mechanism. So, the latency of the IEEE 802.11 standard handover decreases to about 1600 ms, which is less than that of the TCP traffic. However, the disruption of the application service may not be prevented because it still has the handover latency of average hundreds ms. However, the context-aware handover could prevent the deterioration of the application service quality because most of the handover latency is less than 50 ms. This result is similar to the TCP traffic.

Finally, we have tested the performance of network diagnosis utility for detecting his network condition. "Fig. 6" shows the handover latency concerning network diagnosis utility. Network diagnosis utility uses ICMP protocol for network condition testing. In "Fig. 6", context-aware fast handover method has good performance in network diagnosis. When transmitting the FTP and music streaming traffic, the IEEE 802.11 standard handover has considerably large latency ranging from hundreds of ms to 1800 ms. However context-aware handover has about 50 ms latency, so that the disruption of the real-time application service can be prevented.

V. Conclusion

In this article, we present context-aware mobility management scheme which takes into account the context of application type and energy consumption in wireless LAN. In the proposed scheme, the fast handover is accomplished without degradation of the multimedia application session with minimal energy consumption. The model and parameters have been designed to specify the QoS constraints of the multimedia application. The performance of the proposed mechanism is verified by the simulation. The simulation results show that the proposed mechanism is useful for the personalized network which is comprised of wireless LANs.

References

[1] H. Tao, et. al, "Proactive Scan: Fast Handoff With Smart Triggers for 802.11 Wireless LAN." INFOCOM2007, May 2007 Pages 749-759.

[2] Y. Liao and L. Cao, "Practical schemes for smooth MAC layer handoff in 802.11 wireless networks." World of wireless, Mobile and Multimedia networks,2006. 26-29 June 2006 Page(s) 10ppt

[3] Draft Standard for Information Technology-Telecommunications and information exchange between systems-Local and metropolitan area

networks-Specific requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications Amendment : Radio Resource Measurement of Wireless LANs (Amendment to IEEE Std 802.11-2007)

[4] A. Loow, "A Handoff Method for Use with Mobile Devices Based on Active Poling with IEEE 802.11b," M.S. thesis, Dept. Comput. Sci. and Elec. Eng., Lulea. Univ., Sweden, 2007

[5] Y. Chen, N. Smavatkul, and S. Emeott, "Power management for VoIP over IEEE 802.11 WLAN," IEEE WCNC 2004, Atlanta, March 21-25, 2004

[6] A. Mishra, M. Shin, and W. Arbaugh. "An empirical analysis of the IEEE 802.11 MAC layer handoff process," SIGCOMM Comput. Commun.Review, 3(2):93-102, 2003.

[7] F. James, Kurose, and W. R. Keit, "Computer Networking – A Top-Down Approach Featuring the Internet." Addison Wesley Longman, 2007.

[8] I. Ramani and S. Savage, "SyncScan: practical fast handoff for 802.11 infrastructure networks." Proceeding of IEEE INFOCOM 2005, pp 675 – 684.

[9] S. Shakkottai, T. Rappaport, and P. Karlson, "Cross-Layer Design for Wireless Networks," IEEE Commun.Mag., vol. 41, no. 10, Oct. 2003, pp. 74–80.

[10] G. Carneiro, I. Porto, etc., "Cross-Layer Design in 4G Wireless Terminals," IEEE Wireless Communications, April 2004.