Pre-Scanning and Dynamic Caching for Fast Handoff at MAC Layer in IEEE 802.11 Wireless LANs

Noor Mustafa, Waqar Mahmood, Ahsan A. Chaudhry, M. Ibrahim
NUST Institute of Information Technology
NUST University
Rawalpindi, Pakistan
{55noor, drwaqar, ahsanch, 55ibrahim}@niit.edu.pk

***Abstract*—In recent years, wireless LANs have been widely deployed in public places because of its low cost and high-speed connectivity. To cover larger areas, many APs are required which necessitate the use of handoff between different APs. Therefore handoff latency has become essential issue in wireless LANs. The delay caused by handoff process affects the quality of real-time applications like VoIP. So, in order to resolve this problem, an improved fast handoff mechanism is required that can minimize the overall handoff latency enough to run voice and video applications seamlessly. Handoff consists of three phases including scanning, reauthentication, and reassociation. Scanning is the most time consuming phase among all three. We propose a fast handoff scheme in which channels are scanned prior to the time of handoff. This novel approach almost eliminates scanning delay by using a pre-scanning algorithm and it takes only few milli seconds by using a dynamic caching mechanism.**

Keywords- IEEE 802.11, WLAN, Handoff, Selective Scanning

I. INTRODUCTION

IEEE 802.11-based wireless local area networks (WLANs) have immense growth in the wireless domain in recent years because of its low cost and high-speed connectivity. Wireless LANs have been deployed in homes, campuses, small businesses and enterprises. So, the use of multimedia programs has been increased in wireless networks. But the delay caused by handoff process is significant enough to affect the quality of especially multimedia applications. Therefore, the handoff management between Access Points (APs) is one of the critical issues in wireless LAN.

WLAN AP can cover 100 meters to 400 meters. So, many APs are needed for covering large areas. A handoff occurs when a mobile station moves beyond the radio range of one AP and enters another AP's region. During the handoff, 802.11 management frames are exchanged between the mobile node (MN) and AP, and the MN's certain context information is also exchanged between participating APs. Consequently, handoff latency occurs during which the mobile node experiences the disruption in exchanging data of time critical applications.

The complete MAC Layer (L2) handoff process can be divided into three distinct logical phases: probing (scanning), reauthentication, and reassociation. The scanning phase is observed as the most significant contributor to over all handoff latency [4]. Therefore, we have focused on solution for reducing scanning delay in our proposed handoff process.

The remainder of this paper is organized as follows. Section 2 outlines work done in this particular area, followed by a brief description of how we tackle the problem. Section 3 gives a brief introduction to IEEE 802.11. Section 4 describes the original handoff process performed in WLANs. In section 5, recommend techniques be discussed to resolve the problem and describe two approaches from [1] and their consequences. In section 6, we propose the pre-scanning fast handoff process. Section 7 concludes the paper and gives some future directions.

II. RELATED WORK

Many individuals have worked a lot in this regard by reducing delay in different phases of the overall handoff process.

An algorithm to reduce L2 handoff delay is suggested in [3]. Using a caching mechanism on the AP side, which is based on the IAPP protocol in order to exchange the client context information between neighboring APs, reduces the reassociation delay. But IAPP creates an additional delay.

In [1], an approach reduces the total handoff latency by reducing the scanning time. This is achieved by using a selective scanning algorithm and a caching mechanism. Selective scanning takes 129 ms average delay and cache mechanism does not always guarantee the cache hit. However, we have used the selective scanning algorithm in our proposed solution also. In [6] and [7] work is done on the IEEE 802.1x authentication delay, which reduces reauthentication delay by using the Frequent Handoff Region (FHR) selection algorithm.

Our work follows a novel approach and reduces the total handoff latency by obviating the scanning phase. This can be achieved by performing probe process at background (before handoff is invoked). This approach uses pre-scanning algorithm with selective channel mask and a dynamic caching mechanism, which can significantly reduce the scanning delay to zero. This scanning process and caching data structure is maintained at the client side and no modifications are required in the existing network infrastructure or the IEEE 802.11 standard. But some changes might be done in the client side wireless card driver.

III. OVERVIEW OF 802.11

A. IEEE 802.11 standards

IEEE 802.11 standards for wireless LAN function on physical and MAC layers. Standard 802.11, released in 1997, operates at 2.4 GHz ISM frequency band and supports data rate of 1 to 2 mbps. Where as 802.11b, 802.11g, and 802.11a provide higher transmission rate, as shown in Table I.

IEEE 802.11g standard is backward compatible with the 802.11b standard. Our discussion is focused on the IEEE 802.11b standard, even though, the proposed architecture is valid for 802.11g with minor changes. 802.11b uses 11 of 14 possible channels that are distributed over the range from 2.402 GHz to 2.483 GHz, each channel being 22 MHz wide [2], as shown in Fig. 1. Out of these 11 channels, only three channels 1, 6 and 11 do not overlap. So, in a well-configured wireless network, most of the APs operate on non-overlapping channels. Two adjacent APs do not have same channel frequency, in order to avoid co-channel interference.

B. Entities

1) Station (STA): A station is a device that has 802.11 implemented in PHY and MAC layers. Typically, the 802.11 functions are implemented in the hardware and software of a network interface card (NIC). A station could be a laptop PC, handheld device, or an AP.

2) Access Point: AP is a station that connects wireless terminals to the wired network and other mobile nodes in its surroundings.

3) Mobile Node (MN): A mobile node is a device that often changes its location and point of attachment. An MN is also a STA in WLAN context. We use MN in this discussion to distinguish between AP and MN.

4) Basic Service Set (BSS): BSS is the basic building block of an 802.11 wireless LAN. The BSS consists of an AP and a group of any number of mobile nodes.

5) Extended Service Set (ESS): More than one BSS (many APs connected through distribution system DS) can be configured as an ESS.

6) Service Set Identifier (SSID): SSID is unique label that distinguishes one wireless LAN from another. All APs and MNs of a specific WLAN must use the same SSID.

C. Infrastructure mode and ad-hoc Mode

IEEE 802.11 architecture comes in two operating modes: ad-hoc and infrastructure. In *ad hoc mode*, two or more stations discover each other and establish a peer-to-peer relationship without any AP or pre-existing network. In *infrastructure mode*, an AP provides network connectivity to its associated stations to form a BSS. Here, AP acts as a bridge between MNs and wired LANs. In this paper, we concentrate on infrastructure mode.

TABLE I. SUMMARY OF 802.11 STANDARDS

Standard	ISM band	Data rate	Channels
802.11b	2.4 GHz	11 Mbps	14 (3)
802.11a	5 GHz	54 Mbps	32 (8)
802.11g	2.4 GHz	54 Mbps	14 (3)

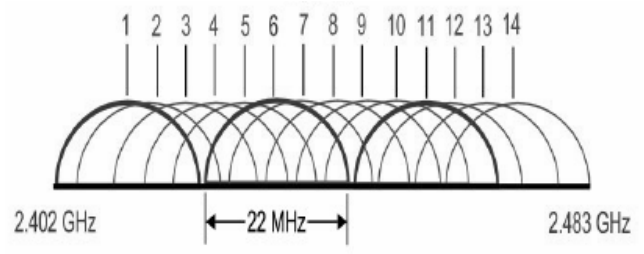

Figure 1. Channels of IEEE 802.11b

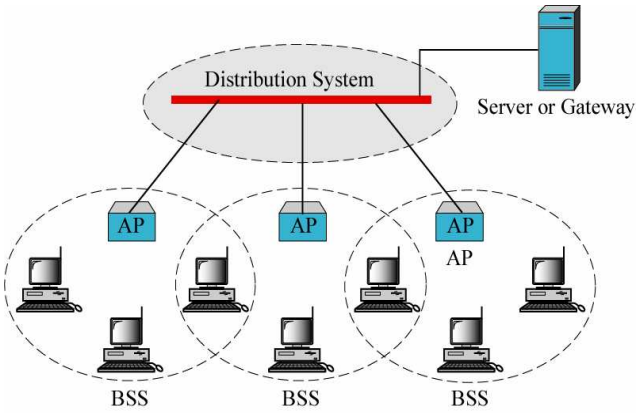

Figure 2. IEEE 802.11 Architecture

D. IEEE 802.11 Management Frames

The IEEE 802.11 management frames enable mobile stations to establish and maintain communications. These management frames constitute probing frames, authentication frames and association/reassociation frames. Some probing frames are briefly defined here, as explained in [5].

1) Beacon frame: APs periodically send beacon frames to announce their presence and relay information, such as timestamp, SSID and other parameters relating the AP's capabilities, to MNs that are within the range.

2) Probe request frame: A MN sends probe request frame to determine which APs are within range.

3) Probe response frame: After receiving a probe request frame, APs reply with a probe response containing capability information, supported data rates.

Some other management frames used in handoff process are authentication request frame, authentication response frame, (re)association request frame, (re)association response frame, and disassociation frame.

IV. MAC LAYER HANDOFF

When a MN switches its association from one AP to another, this is called *'handoff'*. This involves a sequence of messages being exchanged between the MN and the participating APs. As a result, MN's state information is transfer from the old AP to the new AP, consisting of authentication, authorization and accounting information. The Inter Access Point Protocol (IAPP) that is currently under draft in IEEE 802.11f supports the communication between APs [1].

A. Steps during Handoff

The handoff process can be divided into two logical steps: discovery and reauthentication [4].

1) Discovery (Scanning phase): When the MN is moving away from its currently connected AP, the signal strength and the signal-to-noise ratio of the signal from that AP might degrade than a certain threshold level as shown in Fig. 3. This causes the MN to loose connectivity and initiate a handoff somewhere (at T1, T2 or T3 points). Now, the MN needs to find other potential APs (in range) to associate to. The MAC layer scanning function accomplishes this. After scanning, the station can create a list of APs prioritized by the received signal strength.

There are two kinds of scanning methods defined in the standard: *passive* and *active*. In passive scanning, the MN listens to the wireless medium for beacon frames on each channel. Using the information obtained from beacon frames, the MN can elect to join an AP. In active scanning, the MN broadcasts probe request packets on each channel and receives probe responses from the APs. The basic procedure of the active scan mode includes the following steps as explained in [1]:

a) MN uses the normal channel access procedure (CSMA/CA) to gain control of wireless medium.

b) Broadcast a probe request frame.

c) Start a probe timer.

d) Listen for probe responses.

e) If no response has been received by minChannelTime, scan next channel.

f) If one or more responses are received by minChannelTime, stop accepting probe responses at maxChannelTime and process all received responses.

g) Move to next channel and repeat the above steps.

After scanning all channels, information received from probe responses is processed so that the MN can select an AP with good signal quality to join next.

2) Reauthentication: The MN attempts to reauthenticate to an AP according to the priority list. This process typically involves authentication and reassoiciation to the new AP as well as transfer of the MN's credentials from the old AP to the new AP. Authentication is a process by which the AP either accepts or rejects the identity of the MN. After successful authentication, the MN need to reassociate with the new AP, then AP will respond back to the MN, containing an acceptance or rejection notice.

Fig. 4 shows the sequence of messages expected during the handoff. The handoff process starts with the first probe request message and ends with a reassociation response message from an AP. The entire handoff latency can be divided into three delays:

a) Probe messages: The MN starts sending out probe requests and then processes received probe responses. The time involved in this process is called probe delay. The actual number of messages during this process may vary from 3 to 11.

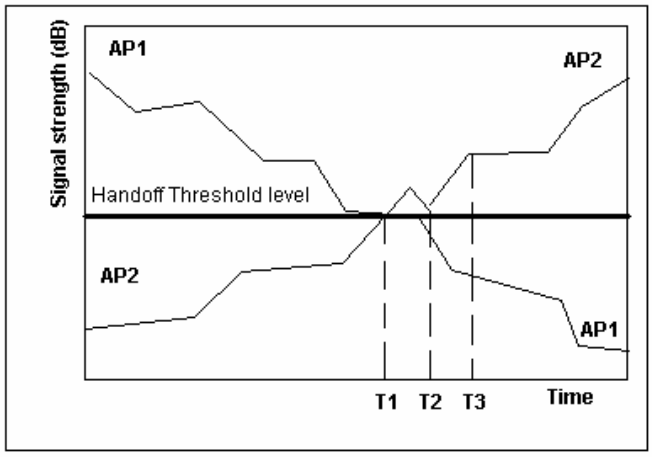

Figure 3. AP's signal strength and handoff threshold

b) Authentication messages: Once the MN decides to join an AP, authentication messages are exchanged between MN and the selected AP. The time consumed by this process is called authentication delay.

c) Reassociation messages: After a successful authentication, the MN sends a reassociation request and expects a reassociation response back from the AP. This is called reassociation delay.

V. Reducing Handoff Delay

Handoff delay at MAC layer is composed of three delays; probe delay, authentication delay, and association delay. As experimented in [4], the value of the handoff delay varies from 56.74ms to 396.76ms and the probe delay accounts for more than 90% of the overall handoff delay. Hence, reducing probe delay can significantly shorten the handoff process. Probing is performed to find an AP with good channel quality. A MN simply scans all possible channels one by one during scanning stage [2].

Total probe time = 'Number of channels scanned' X 'Max. waiting time for each channel' + 'Processing time' (1)

We can address this problem from three points:

1) If the number of channels (to be scanned) is reduced, enough portion of probe time can be saved.

2) If the MN knows all possible APs to be scanned from its previous scanning experience, it can stop waiting before the waiting time is due. Accordingly, we can shorten the waiting time.

3) If a MN knows which AP to switch to, there will be no need of scanning.

In [1], two approaches are used to achieve above points. Our proposed fast handoff procedure also uses these approaches at some places. These two approaches are selective scanning algorithm and caching mechanism. In the following sections, we briefly define these approaches first and then their consequences.

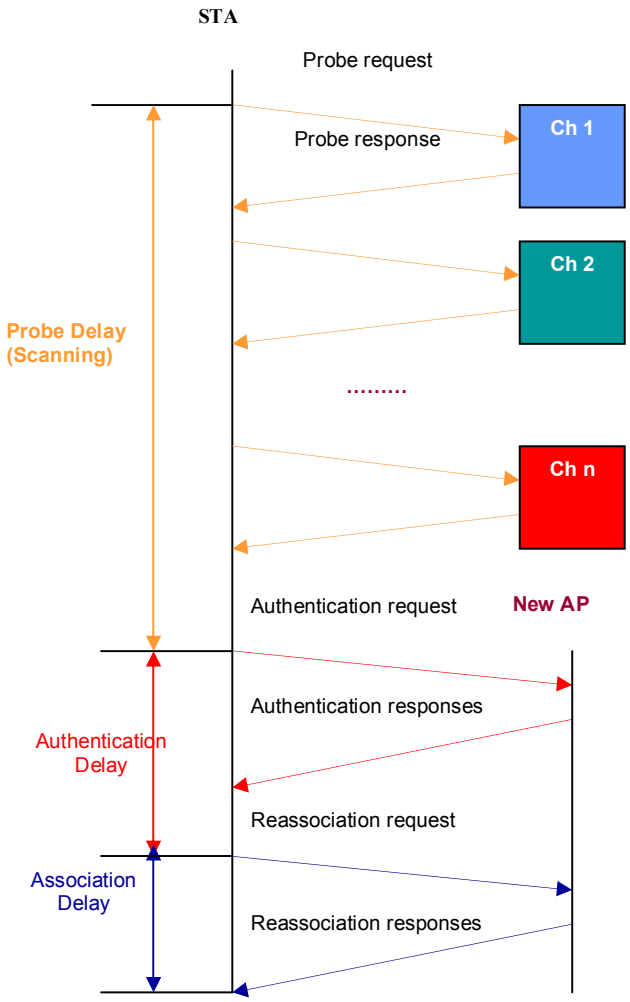

Figure 4. Handoff process using active scanning

A. *Selective scanning*

In selective scanning, only those channels are scanned, which had responded the probe request in previous handoff. In this algorithm, when a MN scans APs, a channel mask is built, shown in Fig. 5. Each bit in channel mask corresponds to possible channel. The channel mask is set by turning on the bits based on received probe responses. The bits for non-overlapping channels are also set because these are more likely to be used by APs [1].

When the driver is first loaded, MN set all the bits and does a full scan. This algorithm is performed in the following way:

1) MN does active scanning for the channels whose bits are set in the channel mask.

2) If no APs are discovered with current channel mask, the channel mask is inverted and a new scan is done. If still no APs are discovered, a full scan is performed.

3) If APs are detected, compute the new channel mask based on the received AP responses.

4) Select the best AP with stronger signal strength and send association request to connect to it.

5) After successful association with new channel, reset the corresponding bit to '0' in the channel mask.

In [1], experiments show that the average handoff delay is between 100ms to 150ms.

1	0	0	0	1	1	0	0	0	0	1	0	0	0
1	2	3	4	5	6	7	8	9	10	11	12	13	14

Figure 4: Selective scanning mask

B. *Caching*

Using cache can skip the scanning phase and go directly into authentication and association phases [1]. This mechanism maintains a table, which stores two adjacent APs for each AP while roaming the network. When a handoff is initiated, MN checks for the APs in the current AP's adjacent list. If an entry is found (cache hit), the MN send association request to connect this new AP. On successful association, the handoff procedure is completed. When the MN fails to connect to the first entry in cache, the second entry is tried and if association with the second entry fails as well (cache miss), then selective scanning algorithm is used.

Table II is taken from [1], and it shows that the average handoff delay is 343ms for original handoff, 129ms for selective scanning, and 3 ms for caching. With selective scanning the handoff latency improved considerably, with an average reduction of 40%. For seamless VoIP, it is recommended that overall latency does not exceed 50 ms [4], but selective scanning cannot reduce handoff delay to 50ms. However, using the cache, the handoff latency time drops to a few ms, making it possible to have seamless VoIP. But caching does not guarantee constant handoff latency because cache carries the results from MN's roaming history but not the current results (the network infrastructure can be changed with time). If there occurs cache-miss, the handoff would be over 100ms. Now the problems are still there.

1) How to reduce/eliminate the scanning delay, and
2) How to guarantee that there is always a cache hit.

One possible solution is to perform probe at background. Performing probe before the handoff and dynamically caching the current results for presently connected AP could solve the problems mentioned above.

TABLE II. HANDOFF DELAY IN (MS) OF 802.11B IN LINK LAYER

Experiment	1	2	3	4	5	6	7	8	9	10	Avg
Original handoff	457	236	434	317	566	321	241	364	216	274	343
Selective scanning	140	101	141	141	139	143	94	142	101	141	129
Caching	2	2	4	3	4	2	2	2	2	2	3

VI. PROPOSED FAST HANDOFF PROCEDURE

In this section, we propose a fast handoff scheme, which comprises two mechanisms; pre-scanning process with selective channel mask and dynamic caching. Before describing the complete handoff process we would look at these two mechanisms.

A. Pre-Scanning with selective channel mask

This process performs scanning at the background and obtains the best neighbors APs before the handoff is invoked. When signal strength gets lowered than a pre-defined threshold but still higher than the handoff threshold, MN triggers probing in following manner, also shown in the Fig. 6.

1. Store the association information of currently connected AP.
2. Switch to a channel from selective channel mask (described in previous section) and perform active scanning.
3. If any AP replies the probe requests, keep the results.
4. Switch back to the association saved earlier.
5. After a certain time, pick another channel by the selective mask and probe again and so on.
6. After scanning all the channels in selective mask, if the probe responses from adjacent APs are received, then compute new channel mask.
7. If no AP has replied, then flip the channel mask and repeat 1-6 steps.
8. Select the best APs (up to five APs maximum) from the stored results and put them into cache.

The average delay of one probe is about 20ms, so, such short amount of time cannot affect the quality of running multimedia application. Because 50ms delay time can be endured to run such application without interruptions. Consequently, the probe phase is completed well before handoff. Thus, we do not need scanning at the time of handoff process, which completely eliminates the probing delay.

B. Dynamic caching

When the handoff occurs, MN disassociates with the AP to which it was connected before. By using this caching mechanism, MN does not need to perform scanning, but it directly picks an AP from the cache and issues an association request to selected AP, as shown in Fig. 7. Dynamic caching contains a list of APs provided by the background probe. All the entries in that list are neighbor APs of disassociated AP. Here, in this cache, we give the provision for only five best AP entries, although it can be extended up to eight entries also. Whenever the handoff is invoked, MN uses the dynamic caching in the following way.

1. First check for availability of the entries in adjacent list in the cache.
2. If entry is not found (cache miss), the MN performs a scan using the selective scanning algorithm (described in previous section). Cache miss occurs when no required entry is found in the cache list.
3. If an entry is found (cache hit), MN issues a request to associate to this new AP. On success, the handoff procedure is complete. Cache hit occurs when required entry is found in the cache list.
4. When the MN fails to connect to the first entry in cache, the second entry is tried and so on, if association with all the entries in the cache fails, then the selective scanning algorithm is used.

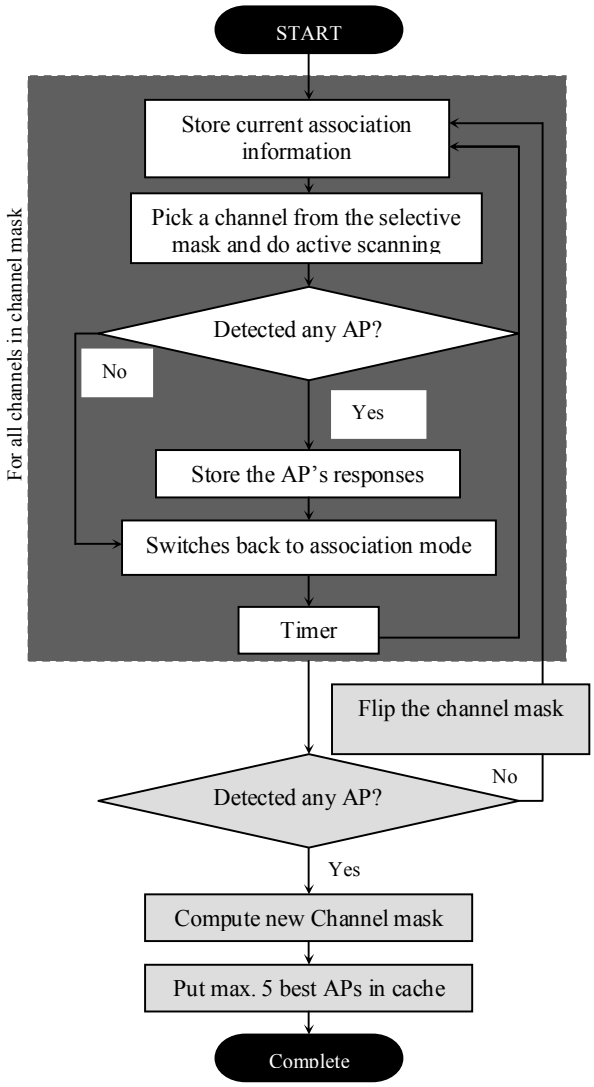

Figure 6. Pre-scanning with selective channel mask

This caching is named as dynamic because each time the MN connects to new APs, the cache will be updated with new results, for every new associated AP. Thus, cache always holds the current and accurate AP entries for MN to switch to, and because of that reason, there is hardly any chance of cache miss occurrence.

Reference [3] described that, "Usually, using cache, it takes less than 5 ms to associate to the new AP. But, when the MN fails to associate to the new AP, the firmware waits for a long time, up to 15ms. To reduce this time-to-failure, a timer

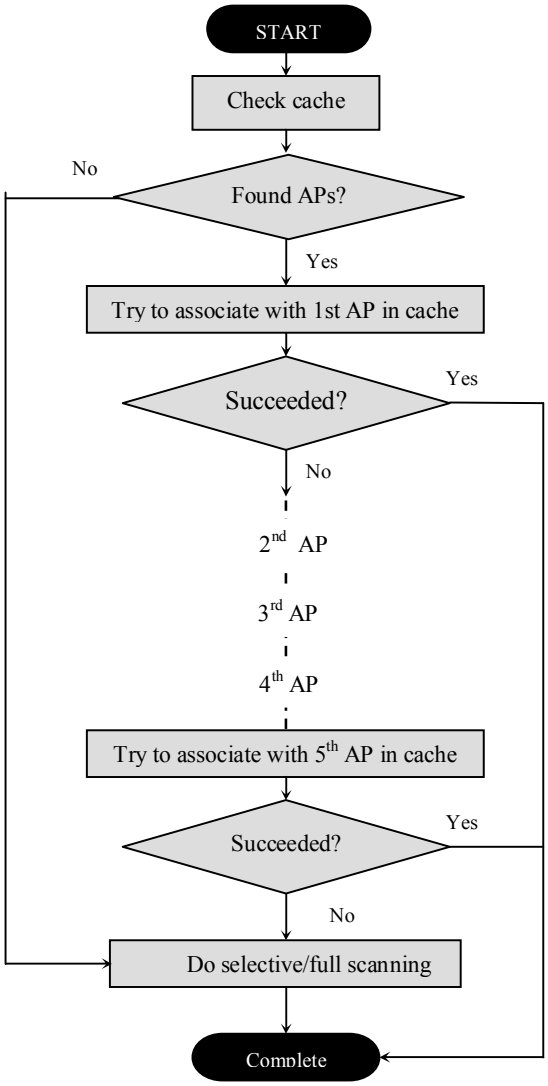

Figure 7. Dynamic caching mechanism

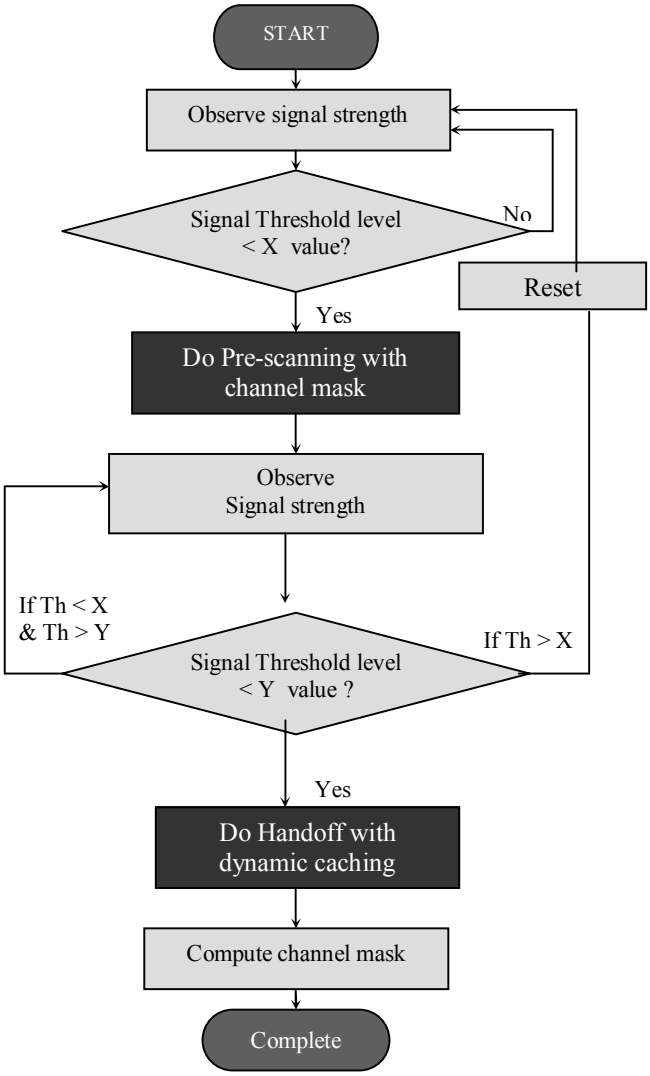

Figure 8. Proposed fast handoff process

is used. The timer expires after 6 ms, and the MN will then try to associate to the next entry in cache and so on".

Thus, in our dynamic caching, we can try multiple entries without affecting the quality of real time application, because such applications can digest up to 50ms delay. We suppose a worst case, e.g. if the fist four cache entries miss and the fifth one hits, the total handoff delay is only 29 ms, which is less than 50ms. Hence, the total handoff delay is still resulting in a significant improvement compared to either the original handoff time or the fast handoff algorithm time, as in [1].

C. Complete Proposed fast handoff process

In this section, we describe our proposed fast handoff process, which uses both the above-mentioned mechanisms, as shown in Fig. 8. In this process, 'X' is a signal strength threshold level where pre-scanning would be initiated. Where as, 'Y' is the signal threshold level where handoff process would be initiated. The signal strength varies as MN moves closer or away from the AP. This fast handoff process keeps watching the variations in MN's receiving signal strength throughout the process. If certain condition becomes true, MN will go for right action at right time. The 'X' condition always comes before 'Y' condition, as shown in Fig. 9. The duration between these two conditions is enough for pre-scanning process. The description of this overall proposed fast handoff process is given below.

1) When the signal strength goes lower than a pre-scanning threshold 'X' but still higher than handoff threshold 'Y', the MN triggers background probe (pre-scanning) with selective channel mask.

2) As the result of step 1, we will get a list of current AP's adjacent APs in the cache.

3) After the scanning is done, MN keeps watching over signal strength to decide whether to:

a) Stay observing the signal strength (if signal strength lies between X and Y) or

b) Initiate handoff (if signal strength is less than Y) or

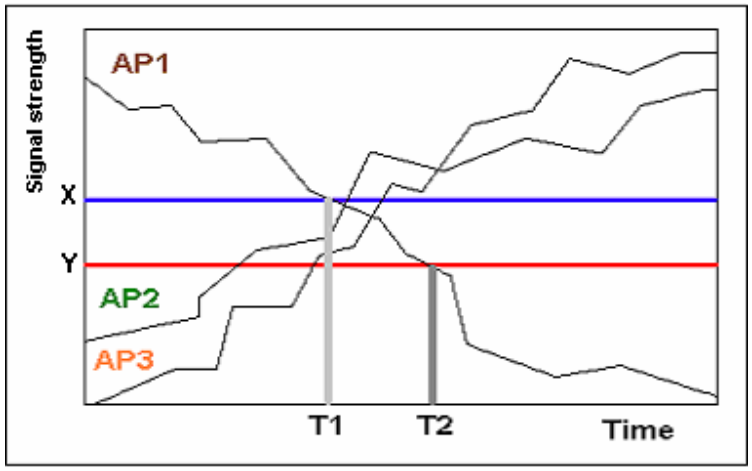

Figure 9. Signal thresholds for pre-scanning and handoff process

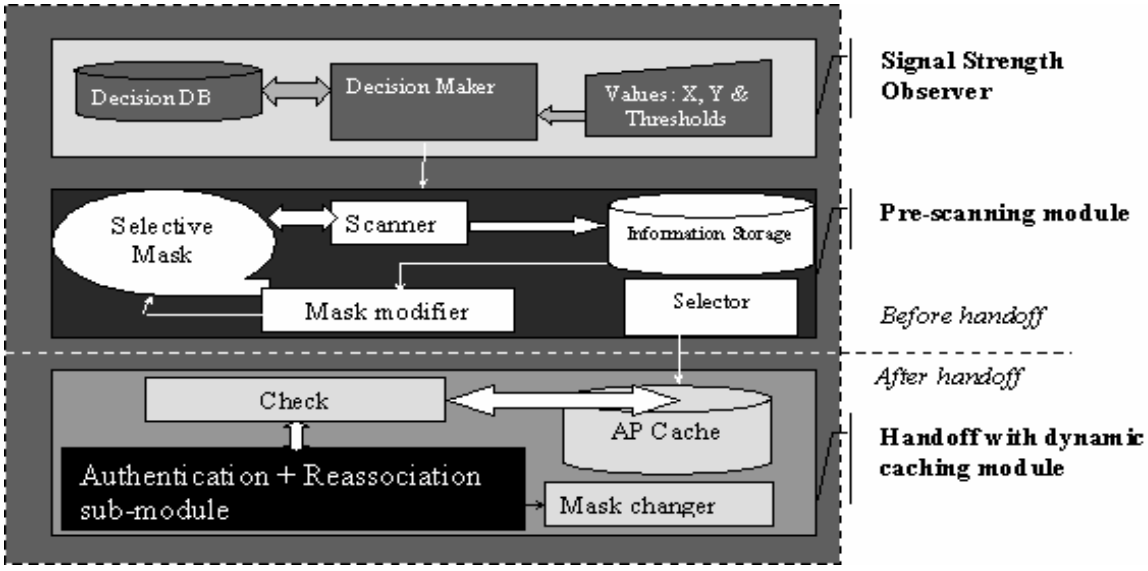

Figure 10. Architectural design for the proposed process

c) Reset the results and go back to step 1 (if signal strength becomes greater than X)

4) If the handoff process is invoked, the MN goes to perform handoff through dynamic caching.

5) After the successful association with new AP, two bits into the channel mask are reset;

a) '0' for the new connected AP

b) '1' for the old connected AP.

The reasons for changing the bits are; i) two adjacent APs do not run on the same channel (this avoids co-channel interference) and ii) the old AP is neighbor of the new AP, so, this channel should be scanned when next handoff is invoked.

By predicting the adjacent APs prior to the handoff time, we can always eliminate the most time consuming phase, the probe delay in the handoff process. Due to current-ness and correctness of neighbor AP entries in the cache, MN always end ups with cache hit. That causes handoff delay not interrupt the voice or video sessions. Now, the expected handoff delay time is up to 5ms or even less. Hence, multimedia application can run seamlessly on wireless LANs.

D. Architectural design for proposed fast handoff procedure

Here, we propose an architectural design, which consists of three main modules; signal strength observer, pre-scanning module, and handoff with dynamic caching module, as shown in Fig. 10. This architecture is actually a summarized form of the flowcharts shown above. This design will be very useful when this handoff process is going to be implemented.

1) *Signal strength observer:* This module runs at MN's side and observes its corresponding receiving signal strength. It contains the decision database and values of signal thresholds, through which this module will decide to perform pre-scanning or handoff.

2) *Pre-scanning module:* This module performs probe for APs in the background and stores the probe responses and other related information. This also keeps updating the channel mask. The selection of best APs is also done here.

3) *Handoff with dynamic caching:* In this module, after the handoff is initiated, availability of adjacent AP entries in cache is checked. After selecting an AP, authentication request and reassociation request are issued to connect that selected AP. On successful completion of handoff, little changes are also done in the channel mask.

VII. Conclusion and Future Work

This paper presented MAC layer fast handoff mechanism in wireless LAN. The approach, pre-scanning fast handoff mechanism, is proposed to shorten handoff delay in order to support the real time applications. This approach is comprised of pre-scanning with selective channel mask and dynamic caching mechanisms. The first approach performs scanning at background and gets the adjacent AP's information before the handoff is initiated. Thus, scanning time is significantly minimized and probe delay becomes negligible. In dynamic caching, the entries are dynamically updated with the change of signal strength, and therefore, it always carries the current and correct entries of adjacent APs. As the result, dynamic caching does guarantee cache-hit every time and there is hardly a chance of cache-miss to occur. This fast handoff mechanism can reduce the handoff delay to 5 ms or even less and can support voice and video application seamlessly.

There are many techniques to reduce handoff latency. However, due to time constrain, no simulation and implementation have been done at this stage. The future work has three parts. First, setting up experiment environment; second, implementing and testing ideas proposed in this paper; third, a simulation might be useful to test the scalability of the system.

References

[1] S. Shin, A. S. Rawat, H. Schulzrinne, "Reducing MAC layer handoff latency in IEEE 802.11 wireless LANs", ACM MobiWac'04, Oct 2004, Philadelphia, Pennsylvania.

[2] Simon Xin Cheng, "The intra-domain WLAN handoff for real-time applications and its implementation suggestion", School of computing science, Simon Fraser University, 2004, unpublished.

[3] M. S. A. Mishra, W. Arbaugh, "Context caching using neighbor graphs for fast handoffs in a wireless network", Technical report, University of Maryland, February 2004.

[4] M. S. A. Mishra, W. Arbaugh, "An empirical analysis of the IEEE 802.11 MAC layer handoff process", ACM SIGCOMM Computer Communication Review, 33(2): 93{102, April 2003.}

[5] J. Geier, "Understanding 802.11 frame types", Technical report, Jupitermedia Corporation, August 2002.

[6] S. Park, Y. Choi, "Fast inter-AP handoff using predictive-authentication scheme in a public wireless LAN", Networks2002 (Joint ICN 2002 and ICWLHN 2002), August 2002.

[7] S. Park, Y. Choi, "Pre-authenticated fast handoff in a public wireless LAN based on IEEE 802.1x mode", Singapore, October 2002, IFIP TC6 Personal Wireless Communications.

[8] "IEEE Std. 802.11, wireless LAN medium access control (MAC) and Physical (PHY) specifications", High Speed Physical Layer Extension in the 2.4 GHz Band, 1999.

[9] Ishwar Ramani, Stefan Savage,"SyncScan: Practical Fast Handoff for 802.11 Infrastructure Networks", Proceedings of the IEEE InfoCom'05, Apr 2005.

[10] IEEE Computer Society LAN MAN standards Committee, "IEEE. Recommended Practice for Multi-Vendor Access Point Interoperability via an Inter-Access Point Protocol Across Distribution Systems Supporting IEEE 802.11 Operation," Jan 2002, IEEE Draft 802.1f/D3.