Effective AP Selection and Load Balancing in IEEE 802.11 Wireless LANs

Jyh-Cheng Chen[1], Tuan-Che Chen[1], Tao Zhang[2], and Eric van den Berg[2]
[1]National Tsing Hua University, Hsinchu, Taiwan
[2] Telcordia Technologies, Piscataway, NJ 08854, USA

***Abstract*— IEEE 802.11 wireless LANs have been widely deployed. How to efficiently balance the traffic loads among Access Points (APs), which will lead to improved network utilization and higher quality of user experience, has become an important issue. A key challenge is how to select an AP to use during the handoff process in a way that will achieve overall load balance in the network. The conventional approaches typically use Signal-to-Noise Ratio (SNR) as the criterion without considering the load of each AP. In this paper, we propose two effective new AP selection algorithms. These algorithms estimate the AP traffic loads by observing and estimating the IEEE 802.11 frame delays and use the results to determine which AP to use. The algorithms can be implemented in software on mobile stations alone. We will show that better performance can be achieved when the APs provide assistance in delay measurements; however, the improvement is not significant. There is no need to exchange information among APs. The proposed algorithms are fully compatible with the existing standards and systems and they are easy to implement. We will present extensive simulation results to show that the proposed algorithms can significantly increase overall system throughput and reduce average frame delay.**

I. Introduction

There has been an immense growth in the popularity of IEEE 802.11 wireless Local Area Networks (WLANs). Today, WLANs are commonly used in homes, offices, and public areas such as airports, cafes, and malls. However, recent studies on the operation of WLANs have shown that the dynamic network service demands over time and locations result in uneven traffic loads between Access Points (APs) [1], [2]. Such uneven traffic loads can occur frequently in popular areas (which are called hotspots) and result in inefficient utilization of network resources, lower network performance, and lower quality of user experience. A key to achieving better load balance, and hence increasing network resource utilization and the Quality-of-Service (QoS) for each user, is the ability for each mobile to select an appropriate AP when multiple APs are available.

When to initiate handoff and how to select an AP are not specified in the current IEEE 802.11 standards or the IEEE 802.21 standards that are being developed to support media-independent handoff. Conventional approaches for AP selection and handoff usually are based on the measurement of *Signal-to-Noise Ratio (SNR)* which supports limited link-layer handoff functionality for wireless stations (STAs). It results in poor performance for end users because STAs are unaware of the traffic loads of different APs [3]. Section II provides the background of IEEE 802.11 MAC and the SNR-based handoff procedure. Several studies have been conducted that used different load metrics to select an AP [4]. Some WLAN manufacturers have also incorporated load balancing into their products [5]. Beacon messages conveying APs' load are broadcast to STAs. Each STA then selects the least loaded AP. Although all of the studies could improve the performance of the SNR-based handoff in some extent, they normally require some modifications in APs or the backbone network to enable information exchange between APs. In addition, they may need a central server to coordinate the handoff process, which increases the complexity and causes some overhead. Besides, how to measure the traffic load effectively is critical.

In this paper, we propose two innovative and efficient algorithms, the *Probe Delay (PD)* algorithm and the *Mean Probe Delay (MPD)* algorithm, to balance the loads among APs without information exchange between APs and without the help from any server. With the PD algorithm, a STA simply measures the *probe delay (PD)*[1] of a single probe frame to approximate the load of an AP in order to determine the appropriate AP. To increase the accuracy of the PD algorithm, the MPD algorithm measures the delays of a few more probe frames and averages the delays as an estimation of the traffic load. The MPD algorithm could be implemented in software on each STA alone without any extra software or assistance from the APs or network servers. Compared with the PD algorithm, the MPD algorithm improves the performance at the expense of little extra handoff time. The MPD algorithm could also be *AP assisted*, in which a STA obtains the estimated delays from APs. The *AP-assisted MPD* algorithm improves the PD algorithm without incurring extra handoff time. Because a lighter load AP will be selected, all of the proposed algorithms also utilize network resources efficiently. Furthermore, unlike previously proposed methods, the proposed algorithms do not require centralized coordination and any assistance from extra servers. This makes them easy to implement. The proposed algorithms are compatible with existing WLANs as well.

The rest of the paper is organized as follows. Section II presents an overview of IEEE 802.11 WLANs. Section III provides detailed explanation of the proposed algorithms. Section IV shows the performance analysis results. Section V

This work was sponsored in part by National Science Council (NSC) under the grant numbers 95-2752-E-007-003-PAE, 95-2221-E-007-027, 95-2219-E-007-012, and 94-2219-E-009-024.

[1]Probing is defined in the 802.11 standards.

1-4244-0357-X/06/$20.00 ©2006 IEEE
This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE GLOBECOM 2006 proceedings.

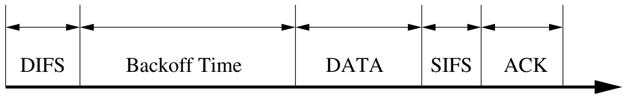

Fig. 1. IEEE 802.11 frame transmission

concludes this paper.

II. Background

The proposed algorithms were developed based on our analysis of the 802.11 Medium Access Control (MAC) protocol. Therefore, this section summarizes the aspects of the 802.11 MAC and the 802.11 handoff procedure that are relevant to our proposed designs.

A. IEEE 802.11 MAC

The MAC of 802.11 employs *Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA)*. In this paper, we mainly consider the *Distributed Coordination Function (DCF)*. When a STA has a frame to transmit, it first needs to sense the media until the media is free longer than the *DCF Inter-Frame Space (DIFS)*. After that, the STA continues to sense the media for a random backoff time which is generated by a random number generator. When listening to the media, the STA suspends the backoff counter if media is busy at any time. Otherwise the counter continues to be decremented. The frame will be transmitted until the counter reaches zero. When the destination station receives the frame without any error, it sends back an acknowledgment (ACK) after a *Short Inter-Frame Space (SIFS)* time. The whole procedure is depicted in Fig. 1. The backoff time is uniformly chosen in the range of [0, *CW*-1] slot time, where *CW* is an acronym of *Contention Window*. Whenever there is a collision, the contention window size is doubled to reduce the probability of collision. At the first transmission, *CW* is set to the value of minimum contention window size. When *CW* equals the maximum value, it will not double again. In 802.11, it appears that the frame transmission time in general would be longer if the traffic load is heavy.

B. Handoff in IEEE 802.11 Wireless LANs

Generally speaking, handoff in 802.11 link-layer is the process of changing the AP a STA associates with. When a STA moves away from its current serving AP, it will initialize handoff procedure to find a new AP. Fig. 2 shows a typical handoff process in which the AP selection is based on the measurement of SNR.

As a STA moves away from AP1 (serving AP) and approaches AP2 (a candidate AP in vicinity), the SNR of AP1 decreases while the SNR of AP2 increases. Once the SNR of AP1 drops below the predefined *search threshold*, the STA starts to discover other APs to associate with. There are two scanning techniques defined in the 802.11 standard. In *active scanning*, the STA switches to every determined channel and broadcast a *probe request* to search for new APs. If an AP receives a *probe request*, it will acknowledge the request by sending back a *probe response*, which indicates that the AP is in vicinity. In *passive scanning*, a STA listens to beacon frames which are periodically transmitted by neighboring APs. The STA learns the existence of other APs by the received beacon frames. In both active scanning and passive scanning, if the SNR of a new AP is greater than the predefined *handoff threshold*, the AP is marked as a candidate AP. Normally, the *handoff threshold* is higher than the *search threshold*. The STA then initiates the handoff process. The flow chart of active scanning is shown in Fig. 3 without considering the gray areas. Although the decision criterion for selecting a best discovered AP is not specified in the standard, currently most systems select an AP with greatest SNR as the target AP as depicted in Step $6a$ in Fig. 3.

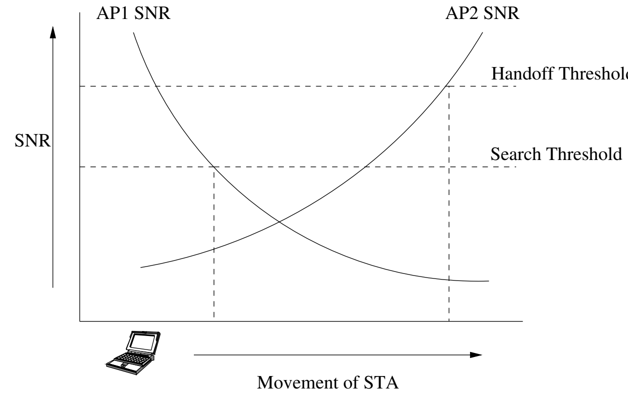

Fig. 2. Handoff based on SNR-measurement

III. Proposed PD and MPD Algorithms

An ideal AP selection algorithm should allow service to continue and should balance the loads of the overall network. As discussed earlier, conventional approaches select an AP with strongest SNR to use without considering traffic loads among APs, which often results in significantly uneven traffic loads among the APs.

Fig. 3 depicts the work flows of our proposed algorithms. The SNR is still used to discover candidate APs and to ensure the availability of the wireless links. We simply replace the SNR-based selection algorithm with our proposed algorithms. In particular, $6a$ is replaced by either $6b$, $6c$ or $6d$. Therefore, our proposed techniques are fully compatible with existing systems. Instead of selecting an AP with maximum SNR, our proposed algorithms measure the traffic load as well. In addition, our proposed techniques could be accomplished by STA alone or with little assistance from APs. There is no need to exchange information among APs. We propose two different algorithms to measure the traffic load: PD algorithm and MPD algorithm. They are discussed in Section III-A and Section III-B, respectively.

A. Probe Delay (PD) Algorithm

We first propose a simple algorithm that allows STA to estimate the traffic load of different APs by using probe frames. As shown in Fig. 1, there are five stages for a successful frame transmission. Section II-A indicates that the major variance in frame transmission time depends on the

1-4244-0357-X/06/$20.00 ©2006 IEEE

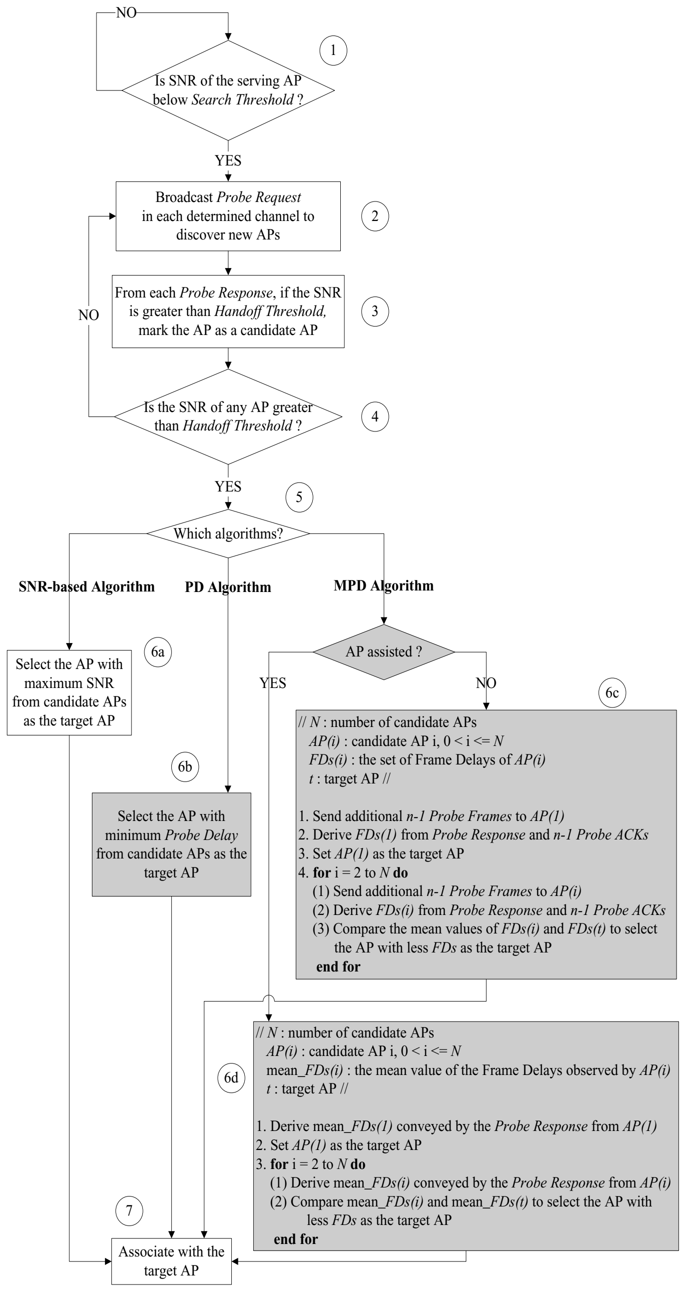

Fig. 3. Flow chart of the proposed algorithms

Backoff Time, which is greatly affected by the traffic load of the wireless media. If there are more users accessing the media simultaneously, the probability of frame collisions will be higher, which in turn increases the probability for each STA to wait for another backoff period. Because 802.11 adopts exponential backoff, the *Backoff Time* will become longer when traffic load is heavy. It results in more time required to transmit a frame. Based on this observation, we reasonably presume that the longer the transmission delay, the heavier the traffic load.

The active scanning defined in the 802.11 standard requires that both *probe request* and *probe response* follow the DCF procedure discussed in Section II-A. We take advantage of this feature by calculating the *probe delay* which is defined as the time from a STA sends out the *probe request* until the time it receives the *probe response* from a specific AP. The probe delay is then used to estimate the traffic load of the AP. The longer the probe delay, the heavier the load of the AP.

Fig. 3 shows the details of the PD algorithm. Whenever receiving a probe response (Step 3), the STA first checks the SNR. If it is greater than the *handoff threshold*, the STA marks the AP which sends the probe response as a candidate AP. The STA also records the probe delay. After the scanning of every determined channel is completed, the STA selects the AP with least *probe delay* from candidate APs as the target AP (Step 6*b*). The STA then hands off to the AP and associates with it (Step 7). By merely recording the probe delays, we can estimate the load of APs without incurring any other overhead.

B. Mean Probe Delay (MPD) Algorithm

The PD algorithm only takes one sample of the *probe delay* as the measurement of the traffic load. We assume that longer probe delay indicates heavier load. To improve the accuracy, we have also developed the MPD algorithm. In the MPD algorithm, we send additional *probe frames* to each candidate AP to better reflect the real traffic load. We then average the delays to derive the mean value. It improves the precision of the PD algorithm.

Fig. 4 illustrates how STA sends more *probe frames* in active scanning. When the STA switches to a channel N and detects that the SNR of AP 2 is greater than the *Handoff Threshold*, it sends additional *probe frames* to AP 2 after waiting for *Max Channel Time*. The *Max Channel Time* is defined in the IEEE 802.11 standard which is the time required to scan the channel. Like *probe request* and *probe response*, the transmissions of *probe frame* and *probe ack* also follow the standard DCF procedure. Because *probe frame* and *probe ack* function the same as *probe request* and *probe response*, they can use the same frame formats. We use different names just for clarity. To support this version of the MPD algorithm, only STA needs to be modified to support the ability of transmitting extra *probe frames*. Nothing else needs to be changed in the standard. The algorithm is shown in Step 6*c* in Fig. 3.

The algorithm described above could be done by the STA alone. However, it increases the handoff latency by sending extra probe frames. With a little enhancement in each AP, the frame delays can be collected by APs without incurring extra handoff delay. Each time when an AP transmits a data frame by using DCF, it can calculate the *delay* in the same way described above. The AP simply records the time from sending out the *data frame* until the time it receives the *ACK*. By recording several delays, the AP can derive the mean value and send it to the STA. The delays can be tracked each time when AP transmits data. During handoff, the STA simply requests the mean delays from candidates APs as the selection criterion. Therefore, there is no extra handoff delay. In passive scanning, the information can be carried by beacon frame which is sent periodically to STAs. Similarly, in active scanning, it can be conveyed by probe response. Therefore, it is not necessary to

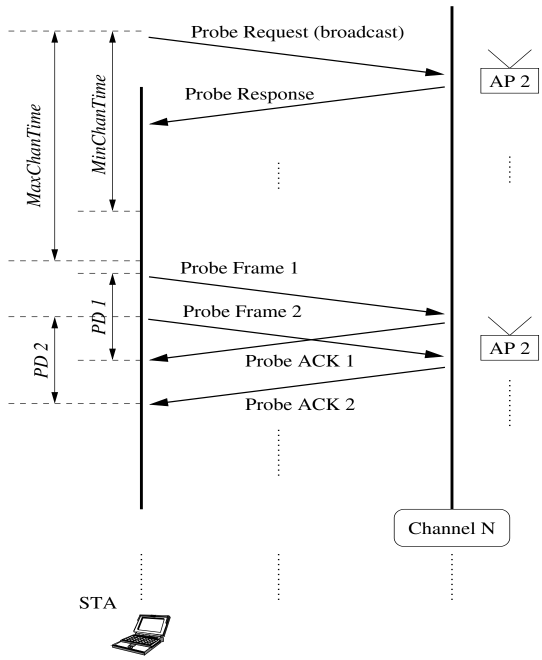

Fig. 4. Procedure for transmitting additional probe frames

send additional probe frames. We call this as *AP-assisted MPD* algorithm. The algorithm is depicted in Step $6d$ in Fig. 3, in which we only take active scanning as an example.

To use the MPD algorithm with or without the assistance from APs is a tradeoff between extra delay incurred without AP's assistance and the added complexity of implementing AP support. The AP-assisted MPD algorithm could obtain several samples to derive the mean delay without prolonging handoff process. However, the AP collects the delays only when there are transmissions. The derived mean value may not precisely reflect current traffic load. In addition, a STA cannot reduce handoff delay unless all APs support the AP-assisted MPD algorithm. On the other hand, the non-AP-assisted MPD algorithm can be done by the STA itself. A STA can perform the algorithm regardless the algorithms supported in other APs and STAs. Besides, the derived mean may better reflect instantaneous load.

It is noticeable that the MPD algorithm based on mean value comparison is easy to implement. Both Step $6c$ and Step $6d$ in Fig. 3 take only $N - 1$ iterations with some simple and quick computations, where N is the number of APs in vicinity. Usually N is less than 5. In Section IV, we will show that by measuring three more frame delays, the MPD algorithm can increase the accuracy of the estimated traffic load significantly.

IV. Performance Analysis

We have conducted extensive simulation to show that the proposed algorithms outperform the conventional SNR-based AP selection algorithm. In addition, the MPD algorithm could accurately compare the traffic load of different APs using a very small number of probe frames. Two sets of experiments have been performed. The first set of experiments aims to compare PD and MPD algorithms with the SNR-based algorithm. The second set of experiments is designed to investigate the MPD algorithm more thoroughly. The simulation was done using the Network Simulator - 2 (ns2).

A. Experiment 1

The simulation assumed an area of 1000 X 1000 square meters, in which APs were randomly located. In order to perform simulation and analysis, some assumptions need to be made:

- Each AP was assigned to a fixed channel throughout the simulation.
- The radio coverage of all APs were identical circles with a radius of 150 meters.
- For optimal channel assignment, adjacent APs with overlapping coverage did not use the same channel.
- The location of each AP was randomly chosen without violating above assumptions.
- For each channel, the number of APs in the whole topology was set as large as possible. Therefore, each area was under the coverage of at least one AP.
- The channel access in IEEE 802.11 MAC layer was implemented using DCF. The physical layer was based on 802.11b with 11 $Mbps$.
- SNR was calculated according to the received signal power.
- The number n of Probe Frames used by the MPD algorithm to measure the average delay for each AP was set to 4 in the active scanning mode. That is, three additional Probe Frames were sent to each candidate AP in the MPD algorithm. The rationale for selecting the value 4 will be discussed in the second set of experiments. For *AP-assisted* MPD algorithm, n was set to 10. It will also be discussed later.

In the beginning, some specific points were located over the entire topology. A STA started its movement at one of these points, and then randomly chose a point as its destination to move to. Once arriving at its destination, the STA stayed for a random exponentially distributed time period before moving to another point. The process was repeated until the end of the simulation. The simulation time was carefully chosen to make sure the simulation results had converged. An AP in vicinity was considered reachable for a STA if the STA was within the circular coverage of the AP. During the movement, each STA generates traffic to its serving AP according to an exponential On/Off traffic source with data rate of 256 $Kbps$ and packet size of 512 $Bytes$.

1) Throughput: Fig. 5 illustrates the overall network throughput versus the number of STAs. The environment was configured to 7 channels. There were 49 APs. Fig. 5 shows that throughput increases when the number of STAs increases. Because PD and MPD algorithms tend to choose lighter loaded APs to use, they achieved better overall system throughput than SNR-based AP selection methods. The MPD algorithm performed better than the PD algorithm because the MPD algorithm estimated the traffic load more accurately. The AP-assisted MPD algorithm achieved the highest throughput be-

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE GLOBECOM 2006 proceedings.

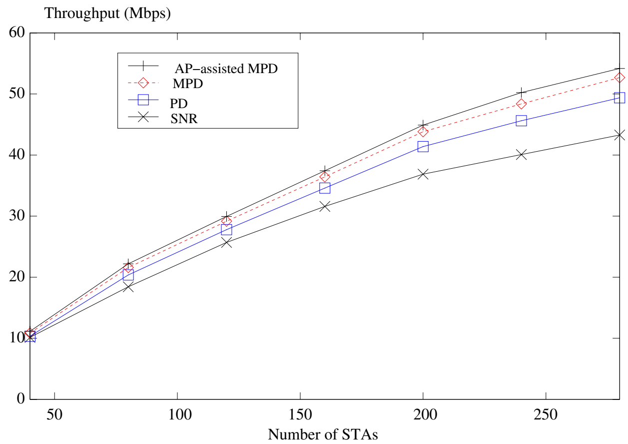

Fig. 5. Throughput vs. number of STAs

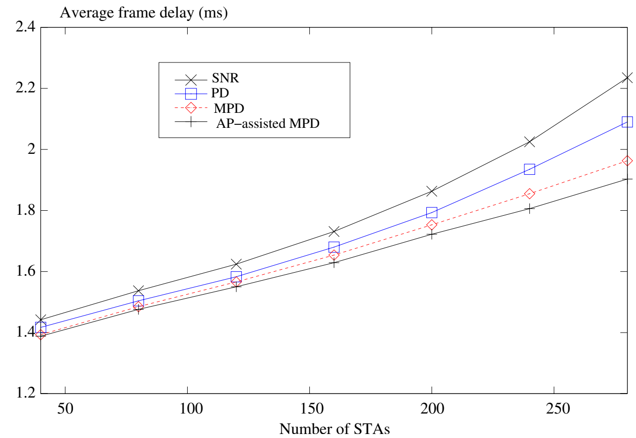

Fig. 6. Average frame delay vs. number of STAs

cause it derived the mean traffic load using even more samples. However, the throughput gain achieved by the AP-assisted MPD algorithm over the MPD algorithm is not significant.

Fig. 5 depicts that, under heavy load conditions with 280 STAs, the overall system throughput achieved by the PD algorithm, the MPD algorithm, and the AP-assisted MPD algorithm improve 14.1%, 21.7% and 25.1%, respectively, when compared with the conventional SNR-based approach.

2) Frame Delay: Fig. 6 presents the average frame delay versus the number of STAs. The frame delay refers to the time from a STA starts to contend for channel access to transit a frame until the time the AP receives the frame successfully. All the parameters were the same as for Fig. 5. Fig. 6 indicates that when the number of STAs increases, the average frame delay increases for all algorithms. For the same reason discussed for throughput, the AP-assisted MPD algorithm performs the best. Fig. 6 shows that when the number of STAs is 280, the PD algorithm, MPD algorithm and AP-assisted MPD algorithm reduce the average frame delay for 6.5%, 12.2% and 14.9%, respectively.

Fig. 7 depicts the average frame delay versus the number of channels. It shows that when the number of channels increases, the average frame delay decreases in PD, MPD and AP-assisted MPD algorithms. This is because when there are more channels available, more APs can be deployed in the 1000 X 1000 square meters area. Thus, traffic load of each AP would be reduced. The SNR-based algorithm, however, always chooses an AP with strongest SNR. Therefore, Fig. 7 shows that the average frame delay almost keeps constant. This indicates that our proposed PD, MPD, and AP-assisted MPD algorithms are especially useful when there are many candidate APs for handing off because the load will be balanced.

Please note the delays shown in Fig. 6 and Fig. 7 are for one frame only. Typically, an application will result in many frame transmissions. The delay reduction would be significant when there are many frame transmissions.

3) Overhead of the MPD Algorithm: One important concern of the MPD algorithm without AP assistance is the extra handoff delay incurred by sending extra Probe Frames.

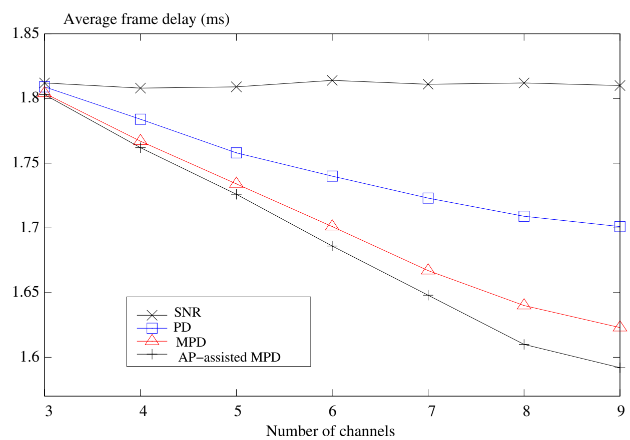

Fig. 7. Average frame delay vs. number of channels

There is no extra handoff latency incurred in the PD and AP-assisted MPD algorithms. The computation time in SNR-based algorithm, PD algorithm, and AP-assisted MPD algorithm is negligible because the main delay falls on frame transmissions. Therefore, the handoff latency of SNR-based algorithm, PD algorithm, and AP-assisted MPD algorithm are almost the same. Fig. 8 shows the extra probe latency incurred by the MPD algorithm. As in Fig. 5, there were 7 channels and 49 APs. Each STA has 3 candidate APs on average for each handoff. Other parameters used in the simulation are shown in Table I. By sending three additional *Probe Frames*, the MPD algorithm needs around 20 ms more for each handoff. The delay shown in Fig. 8 does not consider authentication. Studies have shown that 802.11i would increase the handoff delay significantly [6]. If authentication delay is considered, the additional probe latency incurred by the MPD algorithm is minimal.

B. Experiment 2

In this set of experiments, we intend to investigate the MPD algorithm more deeply. In the simulation, there were 9 APs with different levels of traffic load. Each AP held its own channel. Each time a STA sent a different number of *Probe Frames* to two of the APs. According to the MPD algorithm,

1-4244-0357-X/06/$20.00 ©2006 IEEE
This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE GLOBECOM 2006 proceedings.

TABLE I
PARAMETERS USED IN SIMULATION

Constants	Values
MaxChanTime	11*ms*
MinChanTime	7*ms*
Channel Switch Time	5*ms*

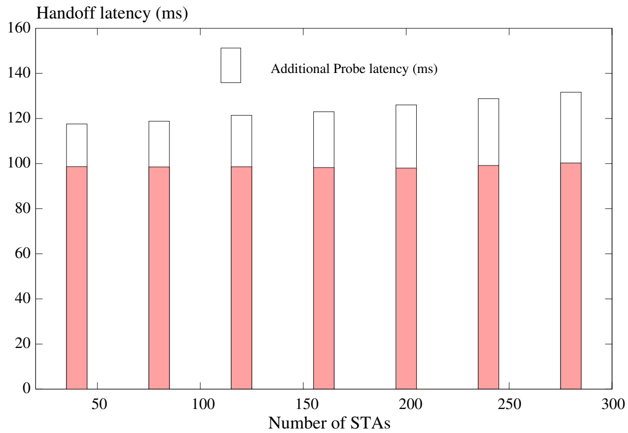

Fig. 8. Average handoff latency vs. number of STAs

the STA then compared the load of the two APs and selected the one with lower load. If the AP which really had less loaded was selected by the STA, it indicated that the STA had a *successful selection*.

Fig. 9 shows the percentage of successful selection of the MPD algorithm with different number of n, the number of samples used to estimate the traffic load of an AP. When one sample is used, it is the PD algorithm. Fig. 9 shows that the percentage of successful selection grows when n increases and/or when the difference in the loads in APs increases. An important observation obtained from these simulation results is that using three more probe frames (i.e., $n = 4$) can achieve very high percentage of successful selection. It achieves 90% of successful selection already when the difference in load is 30%. These results suggest that $n = 4$ should be a good choice for the MPD algorithm without AP support.

When AP assistance is available, more probe messages can be used to get a more accurate estimate of the AP traffic load because using more probe messages in this case will not lead to extra handoff delay. We observe from Fig. 9 that when n increases beyond 10, the percentage of successful selection will have little improvement compared with $n = 9$. This indicates that 10 is adequate enough for utilizing the MPD algorithm. Therefore, n is set to 10 in the AP-assisted MPD algorithm.

The goal of the MPD algorithm is to compare the load of several APs, then select a less loaded AP. Simulation results show that the MPD algorithm could properly select the less loaded AP to associate with by simply averaging a few samples. However, one may question that there is no theoretical validation for the proposed algorithms. We, therefore, have developed two more algorithms by using statistical techniques,

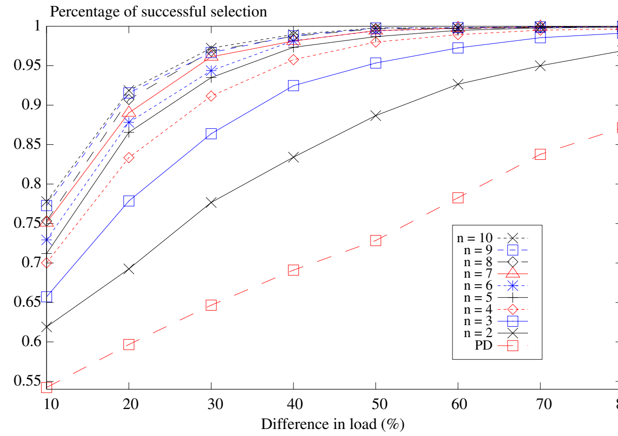

Fig. 9. Percentage of successful selection vs. difference in load (MPD algorithm)

Mann-Whitney test and Rosenbaum's test to compare the traffic load. The results show that the MPD algorithm could achieve same level of accuracy as that using Mann-Whitney test and Rosenbaum's test [4]. It validates that using mean of frame delays as selection criterion is practical.

V. CONCLUSIONS

Currently, the traffic load among different APs usually is not balanced in today's 802.11 WLANs. This paper presents the PD algorithm and MPD algorithm to select an AP effectively and balance the load. Instead of selecting an AP with maximum SNR, the proposed algorithms measure the traffic loads as the decision criterion. In addition, the proposed techniques could be accomplished by the station alone or with help from APs. There is no need to exchange information among APs. They are fully compatible with existing systems and should be easy to implement. Extensive simulation has been performed to compare the performance of the proposed PD and MPD algorithms with the conventional SNR-based approach. Results show that the proposed algorithms could increase the overall throughput and reduce average frame delay. Although the MPD algorithm would incur little extra handoff latency without the assistance from APs, the overhead should be negligible when authentication is also employed in handoff process.

REFERENCES

[1] D. Kotz and K. Essien, "Analysis of a campus-wide wireless network," in *Proc. of ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom)*, (Atlanta, GA, USA), pp. 107–118, Sept. 2002.
[2] D. Schwab and R. Bunt, "Characterising the use of a campus wireles network," in *Proc. of IEEE INFOCOM*, (Hong Kong), pp. 862–870, Mar. 2004.
[3] G. Judd and P. Steenkiste, "Fixing 802.11 access point selection," *SIGCOMM Comput. Commun. Rev.*, vol. 32, p. 31, July 2002.
[4] J.-C. Chen, T.-C. Chen, T. Zhang, and E. van den Berg, "Effective access point selection algorithms for handoffs in IEEE 802.11 wireless LANs." Technical Report, Aug. 2006.
[5] Cisco Systems Inc., "Data sheet for cisco aironet 1200 series," 2004.
[6] J.-C. Chen, M.-C. Jiang, and Y.-W. Liu, "Wireless LAN security and IEEE 802.11i," *IEEE Wireless Communications*, vol. 12, pp. 27–36, Feb. 2005.