This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE INFOCOM 2010 proceedings
This paper was presented as part of the main Technical Program at IEEE INFOCOM 2010.

HaND: Fast Handoff with Null Dwell Time for IEEE 802.11 Networks

Xi Chen and Daji Qiao
Iowa State University, Ames, IA 50011
{leon6827, daji}@iastate.edu

***Abstract*—How to reduce the handoff delay and how to make appropriate handoff decisions are two fundamental challenges in designing an effective handoff scheme for 802.11 networks to provide seamless and satisfactory data roaming services to mobile users. In this paper, we propose a unique fast handoff scheme called HaND (Handoff with Null Dwell time). HaND adopts a novel zero-channel-dwell-time architecture which leverages on the communication backbone between APs to relay the information about wireless channels, and allows the AP (rather than the station) to make appropriate handoff decisions aiming at providing fair service satisfaction to all stations. HaND is a software-only solution and compatible with the 802.11 standard without modifying the 802.11 protocol or introducing new wireless frames. We have implemented it in the Madwifi device driver and demonstrated its effectiveness via experiments.**

I. Introduction

IEEE 802.11 is one of the dominant technologies for broadband wireless networking. Portable 802.11 enabled devices, such as smart phones, netbooks and personal multimedia devices, are becoming increasingly popular. As many new applications for 802.11 enabled devices emerge which require higher quality of services and better mobility support, there is an increasing demand for improving the performance of 802.11 networks. Deployment of multiple APs (Access Points) is an effective way to improve the capacity and performance of 802.11 networks. Nowadays, more and more large-scale 802.11 networks with multiple APs have been deployed in public venues and even city-wide in order to provide broadband wireless networking services to a larger area. One of the important goals of deploying such networks is to support seamless data roaming when users are mobile. For example, a user making VoIP (Voice over IP) calls from an 802.11 enabled smart phone should not experience service disruption or quality degradation to the calls when the user moves around in the network. In practice, this is a challenging task due mainly to the long delay incurred in the 802.11 handoff procedure, during which the data services are disrupted.

One of the core components in the 802.11 handoff procedure is to collect the information about other channels and nearby APs operating on these channels. In a typical 802.11 handoff procedure, a station switches to a new channel, sends a probe frame, and then dwells on the channel for a certain amount of time waiting for responses from the APs (if any) operating on the channel. Channel dwell time has been recognized as the most significant contributor to the overall handoff delay. In general, it is desirable to minimize channel dwell time as much as possible to reduce the handoff delay and minimize the disruption to data roaming services. On the other hand, the station needs to dwell on the scanned channel long enough to wait for returning response frames. If the APs on the channel are busy serving other stations, or if there are many APs operating on the channel, the station may not be able to collect all the responses when channel dwell time is set too small. So there is a tradeoff. In practice, it is difficult to set a proper channel dwell time for an 802.11 station since the context information about the scanned channel, such as the number of APs on the channel and the loads of the APs on the channel, is unavailable to the station. Various fast handoff schemes have been proposed in the past. However, they either use simple heuristics to set a small channel dwell time, which may not perform well under certain circumstances, or require modification or upgrading of the already deployed AP infrastructure or the 802.11 device firmware.

In this paper, we address the above challenge from a different angle. We propose a unique fast handoff scheme, called HaND (Handoff with Null Dwell time), which does *not* require a station to dwell on the probed channels. Specifically, HaND has the following salient features:

- HaND is based on a novel *zero-channel-dwell-time* architecture. Different from existing handoff schemes, a HaND station switches back to its current channel immediately after sending out a probe frame on the new channel. Instead, the current AP of the station is responsible for collecting the responses from nearby APs and forwarding them to the station over its current channel. These information will then be used by the station to adjust its channel scanning strategy (e.g., scan interval).
- The key innovation of HaND is to leverage on the wired[1] communication backbone between APs to relay the information about wireless channels as well as exchange other related information.
- Different from existing schemes, in HaND, handoff decisions are *not* made by the station itself *but* its currently-associated AP. This way, more sophisticated handoff heuristics may be developed since the AP exchanges more related information with nearby APs periodically. HaND adopts a novel *satisfaction-based-fairness* heuristic, which aim to provide fair service satisfaction to all stations and hence avoids possible performance anomaly caused by the transmission rate diversity among them.
- More importantly, HaND is a software-only solution and compatible with the 802.11 standard without modifying the 802.11 protocol or introducing new wireless frames. Information dissemination from AP to station is via modified Probe Response management frames. This is practically feasible and can be implemented with commodity 802.11 devices since the 802.11 management

The research reported in this paper was supported in part by the Information Infrastructure Institute (iCube) of Iowa State University and the National Science Foundation under Grants CNS 0716744 and CNS 0831874.

[1]It is also possible to have a wireless communication backbone between APs, e.g., a WiMAX network.

978-1-4244-5837-0/10/$26.00 ©2010 IEEE

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE INFOCOM 2010 proceedings
This paper was presented as part of the main Technical Program at IEEE INFOCOM 2010.

frames usually are generated in the device driver, which is different from 802.11 control frames (e.g., RTS, CTS and ACK) whose generation process typically is hard-coded in the device firmware. We have implemented HaND in the Madwifi device driver.

The rest of the paper is organized as follows. The 802.11 handoff procedure and related work are discussed briefly in Section II. A few observations about the handoff procedure from experiments are presented in Section III. Section IV gives an overview of the proposed HaND scheme and Section V describes its design and implementation details. Experiment-based performance evaluation results are presented in Section VI and the paper concludes in Section VII.

II. Preliminaries and Related Work

A. IEEE 802.11 Handoff Procedure

IEEE 802.11 [1] describes the general handoff procedure whose actual implementation is vendor specific and usually proprietary. Fig. 1 shows the general 802.11 handoff procedure which has the following three phases: *scan phase*, *authentication phase* and *re-association phase*.

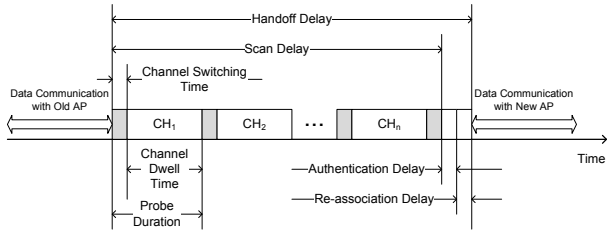


Fig. 1. The IEEE 802.11 handoff procedure.

As shown in the figure, the overall handoff delay consists of the following components: *channel switching time*, *channel dwell time*, *authentication delay* and *re-association delay*. Channel switching time typically is a small value dependent on the 802.11 card. For example, it is around 4.8 ms for cards with Atheros chipsets, and 2.9 ms for cards with Intel chipsets. During the authentication phase, the station sends an Authentication Request frame to prove its identity to the AP, which can accept or reject the request based on its policy. When using *Open System Authentication*, authentication delay usually is a few milliseconds. The re-association phase involves the exchange of Re-association Request and Response frames, as well as the transfer of station's context information between the old AP and the new AP if IAPP (Inter-Access Point Protocol) is used. Without IAPP involved in the re-association phase, re-association delay usually is a few milliseconds.

While channel switching time, authentication delay and re-association delay are all in the order of a few milliseconds, channel dwell time could be as large as 100 milliseconds, depending on the channel scan technique. With passive scan, the station passively waits for the Beacon frames, if any. Since Beacon frames usually are broadcast by the AP every 100 ms, channel dwell time should be at least 100 ms to guarantee reception of Beacon frames. On the other hand, with active scan, the station sends a Probe Request frame on the channel and then dwells on the channel for a certain amount of time waiting for Probe Response frames from APs operating on the channel, if any. In this case, channel dwell time (for each channel) usually is set to around 40 ms. During the scan phase, the station probes several channels and the number of channels to probe depends on the 802.11 device. For example, an 802.11g card has 11 channels to scan. So the total channel dwell time during the scan phase could be as large as a few seconds. Clearly, channel dwell time contributes to a major part of the handoff delay.

B. Related Work

Various fast handoff schemes have been proposed in the past. Most of them have focused on reducing channel dwell time. In [2], the authors showed that the scan phase is the most significant contributor to the overall handoff delay and the variations in channel dwell time account for the large variations in the handoff delay. The schemes proposed in [3], [4] try to reduce the number of channels to probe during the scan phase. For example, [4] adopts a neighbor graph approach and the station only probes the channels that have active APs operating on. D-Scan [5], Proactive-Scan [6], [4], and [7] try to reduce channel dwell time by tuning the values of `MinChannelTime` and `MaxChannelTime` in the case of active scan. SyncScan [8] reduces the probe duration by having all APs synchronize their Beacon frames in the case of passive scan. D-Scan, Proactive-Scan and the scheme in [9] interleave the long scan phase to reduce the packet delay. Different from the above schemes which all try to reduce the scan delay, the scheme in [10] aims to reduce the re-association delay.

During the handoff procedure, another important issue is for an 802.11 station to decide whether to re-associate with a new AP. One of the commonly-used metrics is the signal strength of an AP. For example, hysteresis-based approaches in [5], [6], [9] direct a station to re-associate with the new AP that has a stronger signal strength (by a certain threshold) than its current AP. [3] compares several re-association metrics that are based on signal strength statistics. [11] combines the signal strength value with the large-time-scale performance measurement. In [12], another metric is proposed which is called available capacity. It is the product of the percentage of free air time and the expected transmission rate.

III. Observations from Experiments

In this section, we present a few interesting observations from experiments which help us understand the tradeoffs involved in the handoff procedure and design our proposed scheme. Experiments are conducted with Dell Latitude laptops equipped with D-link WNA-1330 802.11b/g and Wistron CB9-GP-EXT 802.11a/b/g cards. Each laptop is loaded with the Madwifi device driver v0.9.4 [13] to collect experimental data.

A. Effects of Channel Dwell Time and AP Load

In general, it is desirable to minimize channel dwell time as much as possible to reduce the handoff delay and minimize the disruption to data services. On the other hand, the station needs to dwell on the scanned channel long enough to wait for returning Probe Response frames. If the APs on the channel are busy serving other stations, or if there are many APs on the channel, the station may not be able to collect all the Probe Responses when channel dwell time is set too small. From the experimental results plotted in Fig. 2, it can be seen clearly that the percentage of Probe Response frames received during channel dwell time drops significantly as channel dwell time reduces or as the AP load goes up. So there is a tradeoff. In

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE INFOCOM 2010 proceedings
This paper was presented as part of the main Technical Program at IEEE INFOCOM 2010.

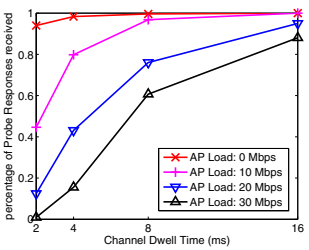


Fig. 2. Effects of channel dwell time and AP load on the percentage of Probe Response frames received.

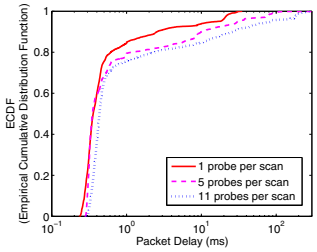
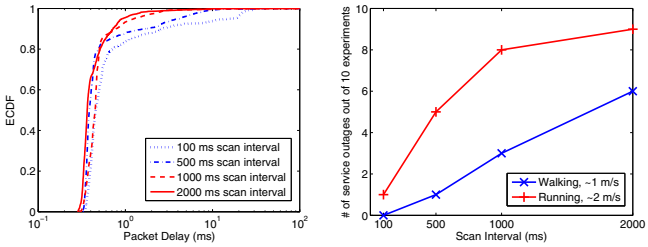


Fig. 3. Effects of probe interleaving on packet delay. Scan intensity is 1 channel per 100 ms. Channel dwell time is set to 20 ms.



(a) packet delay vs. scan intensity (b) service outage vs. scan intensity

Fig. 4. Effects of scan intensity. We fix the number of channels to probe per scan phase to one, and vary the scan intensity level by varying the scan interval. Channel dwell time is set to 20 ms.

practice, it is difficult to set a proper channel dwell time for an 802.11 station since the context information about the scanned channel, such as the number of APs on the channel and the loads of the APs on the channel, is unavailable to the station.

B. Effects of Probe Interleaving

We define *scan intensity* as the number of channels to probe per unit time. To maintain the same scan intensity level, an 802.11 station has two options: (i) probe all channels during one single scan phase with a long scan interval; or (ii) probe fewer channels during one scan phase but with a shorter scan interval. We call the latter one the *probe interleaving* option. Since data transmissions are disrupted during the scan phase when the station switches to other channels, we expect a better packet delay performance when channel probing is more interleaved, which is confirmed by the experimental results plotted in Fig. 3. As shown in the figure, in the case of 1 probe per scan phase (with a scan interval of 100 ms), the maximum packet delay is around 30 ms, while in the case of 11 probes per scan phase (with a scan interval of 1100 ms), the maximum packet delay is around 295 ms, which is unacceptable for real-time data transmissions. This observation is also consistent with the observations from [6], [9].

C. Effects of Scan Intensity

We now study the effects of scan intensity on the handoff performance by varying the scan interval from 100 ms to 2000 ms with a fixed number (i.e., one) of channels to probe per scan phase. The first experiment is performed between an AP and a static station. Fig. 4(a) plots the relation between scan intensity and packet delay. As shown in the figure, packet delay decreases as the scan intensity level decreases (i.e., as the scan interval increases). In the second experiment, we set up two APs with overlapping coverage and have a mobile station moving from one AP to another at different speeds (walking and running). The station hands off to the new AP when the signal strength of the new AP is 5 dB higher than that of its current AP. We define *service outage* as the event when the station loses the connection to the current AP but has not associated with the new AP yet. Fig. 4(b) plots the relation between scan intensity and number of service outages. As shown in the figure, as the scan intensity level decreases, the probability for an 802.11 station to experience a service outage increases. So there is a tradeoff and it is a challenging task to set a proper scan intensity level in practice.

D. Potential Performance Anomaly Caused by Handoff

As discussed in Section II-B, another important issue during the handoff procedure is to decide whether an 802.11 station should re-associate with a new AP, and most existing schemes make the decision based on the comparison of the signal strengths of the APs. We conduct a simple experiment to study the performance of such simple *strongest-signal-first* heuristic.

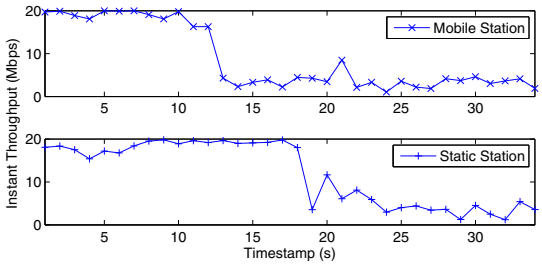


Fig. 5. Performance anomaly caused by handoff.

In the experiment, a static station associates with AP_1 and transmits at a high rate of 54 Mbps. A mobile station moves from AP_2 towards AP_1 and re-associates with AP_1 at around the 18-second mark when the signal strength of AP_1 is higher than that of AP_2. However, at that moment, the mobile station is still at quite a distance away from AP_1 and hence communicates with the AP at a low rate. Fig. 5 plots the instant throughputs of the two stations, where *instant throughput* is defined as the throughput measured for the one-second period ending at the corresponding time instance. We can see that the throughput of the static station drops sharply right after the mobile station joins the network. This, in fact, is the well-known *performance anomaly* in 802.11 networks [14], which is caused by the transmission rate diversity among all stations associating with the same AP. The high-rate station is "slowed down" by the low-rate station because the 802.11 protocol is designed to allow all contending stations to access the channel with equal probability. Various schemes have been proposed to deal with this issue. Unfortunately, during the conventional 802.11 handoff procedure, since the context information about the APs such as the transmission rates of their associated stations is unavailable to the mobile station, such performance anomaly is difficult to avoid.

IV. Overview of the Proposed Scheme

In the following, we give an overview of the proposed fast handoff scheme, called HaND (Handoff with Null Dwell time), with emphasis on how HaND deals with the issues discussed in the previous section.

1) HaND is based on a unique zero-channel-dwell-time architecture: As discussed in Sections II and III-A, channel dwell time contributes to a major part of the handoff delay and it is difficult to set a proper channel dwell time in practice

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE INFOCOM 2010 proceedings
This paper was presented as part of the main Technical Program at IEEE INFOCOM 2010.

since the context information about the scanned channel is unavailable to the station. Different from existing solutions which use various heuristics to reduce channel dwell time and handoff delay, HaND addresses this challenge from a different angle by adopting a novel zero-channel-dwell-time architecture. Instead of having the station dwell on a new channel to collect the responses from the APs on the channel (if any), a HaND station switches back to its current channel immediately after sending a probe frame on the new channel. The current AP of the station collects the responses from the APs on the new channel via the wired communication backbone between them and forward them to the station over the current channel. With this architecture, (i) the difficulty of setting a proper channel dwell time is avoided; (ii) zero channel dwell time minimizes the effects of channel probing on packet delay; and (iii) the coordination between APs helps to make more appropriate handoff decisions for the station.

2) HaND interleaves the channel probes: As discussed in Section III-B, probe interleaving helps decrease the packet delay. This technique is also used in HaND.

3) HaND adjusts the scan intensity adaptively: As discussed in Section III-C, there is a tradeoff when setting scan intensity. It should not be set too high (which may increase the packet delay) or too low (which may increase the service outage probability). Based on this observation, we propose scan intensity adaptation in HaND, where a station adjusts its scan intensity dynamically to the RSSI (Received Signal Strength Indicator) value of its current AP.

4) HaND uses a new satisfaction-based-fairness heuristic to make handoff decisions: Unlike most existing schemes in which handoff decisions are made by the station, in HaND, handoff decisions are made by the station's currently-associated AP. HaND adopts a new handoff heuristic, called satisfaction-based fairness, which emphasizes fair service satisfaction among all stations and hence avoids the performance anomaly described in Section III-D.

5) HaND is a practical fast handoff scheme: HaND is a software-only solution and does not require any modifications to the hardware of the already deployed AP infrastructure or 802.11 devices. It is compatible with the 802.11 standard without modifying the 802.11 protocol or introducing new wireless frames. This makes HaND practical and implementable with commodity 802.11 devices.

V. Design and Implementation of HaND

In this section, we first introduce the overall structure of the proposed HaND scheme, and then describe the station and AP behaviors in detail.

A. Overall Structure of HaND

The overall structure of HaND is shown in Fig 6. Suppose that the station currently associates with AP_{curr} which operates on channel CH_{curr}, and another access point AP_{new} operates on channel CH_{new}. HaND works in the following steps:

(i): The station switches to channel CH_{new} and broadcasts a Probe Request frame, and then switches back to its working channel, i.e., CH_{curr}, immediately without dwelling on the new channel CH_{new}.

(ii): Upon receiving the Probe Request frame from the station, AP_{new} immediately replies a Probe Response frame via its wireless interface `ath0`. Note that since the station has already switched back to channel CH_{curr}, it will not be able to receive this Probe Response frame. However, the presence of this step is important since it guarantees the backward compatibility of the proposed HaND scheme with other 802.11 stations that follow the default handoff procedure. More details will be discussed in Section V-B.

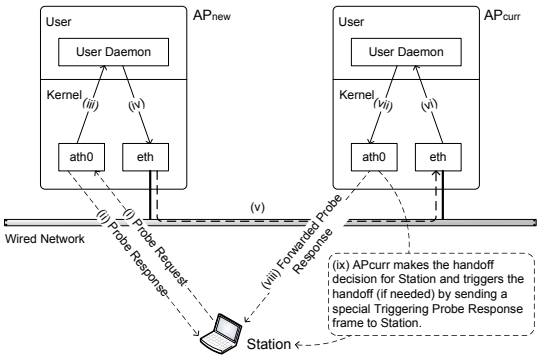


Fig. 6. Overall structure of the proposed HaND scheme.

(iii): In addition, AP_{new} passes the information about its own channel (i.e., CH_{new}), the RSSI value of the received Probe Request frame, and a copy of the generated Probe Response frame to a user daemon.

(iv,v,vi): The user daemon at AP_{new} in turn forwards these information to a user daemon at AP_{curr} via the wired interface `eth` and the wired communication backbone between APs.

(vii,viii): Upon receiving the forwarded information, AP_{curr} updates a local database that stores all the information forwarded from neighboring APs, generates a special *Forwarded Probe Response* frame, and broadcasts it via its wireless interface `ath0` to notify the station of the information about AP_{new}. This way, the station gets to collect neighboring APs' information without wasting time to dwell on other channels. Details will be discussed in Section V-C3.

(ix): A unique characteristic of HaND is that the AP_{curr} makes the decision for the station on whether to re-associate with AP_{new}, instead of the station itself making the decision. This becomes possible with HaND because AP_{curr} coordinates with neighboring APs and collects their information via the wired communication backbone. The actual handoff is triggered by a special *Triggering Probe Response* frame sent by AP_{curr} to the station. Details will be discussed in Section V-C4.

B. HaND Station Behaviors

Depending on whether it currently associates with an AP, a HaND station behaves in two different ways. If it currently does not associate with any AP, it performs a regular scan (in comparison to the adaptive scan discussed below) immediately. During the regular scan, the station probes all channels and channel dwell time for each channel is 20 ms. After collecting all the information about nearby APs on all channels, it associates with the AP that has the highest RSSI value. On the other hand, if a HaND station currently associates with an AP (denoted as AP_{curr}), its behaviors become more complicated, which we describe from the following three aspects: *when to trigger a scan*, *how to scan* and *which channel to scan*.

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE INFOCOM 2010 proceedings
This paper was presented as part of the main Technical Program at IEEE INFOCOM 2010.

1) When to Scan: The station maintains a moving average of the RSSI values collected from downlink traffics from AP_{curr} to the station such as data transmissions and Beacon broadcasts, and another moving average of uplink RSSI values based on the information carried in the Forwarded Probe Response frames. To deal with potential link asymmetry between the station and AP_{curr}, we use the minimum of these two moving averages (denoted as $\text{RSSI}_{\text{AP}_{\text{curr}}}$) to reflect the link quality between them. When $\text{RSSI}_{\text{AP}_{\text{curr}}}$ is below a certain threshold (35 dB[2] in our implementation), the station starts scanning other channels. Neighboring APs' information such as the RSSI values will be reported to the station by AP_{curr} via Forwarded Probe Request frames. This procedure will be described in detail in Section V-C3. Reported information will then be used to decide *how to scan* and *which channel to scan*.

2) How to Scan: An important parameter to characterize a station's scan behaviors is the *scan interval* (denoted as I_s). In HaND, the value of I_s varies with $\text{RSSI}_{\text{AP}_{\text{curr}}}$. In general, when the link quality with AP_{curr} gets worse ($\text{RSSI}_{\text{AP}_{\text{curr}}}\downarrow$), the station should increase the scanning frequency ($I_s\downarrow$) to search for potential backup APs to re-associate with. In our implementation, we set three values for I_s: 450 ms if $\text{RSSI}_{\text{AP}_{\text{curr}}} \in$ [25,35) dB, 300 ms if $\text{RSSI}_{\text{AP}_{\text{curr}}} \in$ [15,25) dB and 150 ms if $\text{RSSI}_{\text{AP}_{\text{curr}}}$ is below 15 dB.

Since data services are disrupted during the scan phase, HaND does the following to minimize the effects: (i) only one of the channels is probed every I_s time; and (ii) the station switches to the channel to send a Probe Request frame and then switches back to its current working channel immediately with zero channel dwell time. This achieves the purpose because:

- Channel switching time typically is very small, e.g., around 4.8 ms for cards with Atheros chipsets, based on our experiments, and around 2.9 ms for cards with Intel chipsets, according to [6];
- Transmission time of the Probe Request frame is negligible since it has a maximum size of 72 octets [1] which can be transmitted in less than 1 ms even at the lowest rate of 1 Mbps.

3) Which Channel to Scan: During each scan phase, the station selects a channel to probe on a probabilistic basis. In general, channels with operating APs that have a good link quality to the station are scanned with a higher probability. It works as follows. Let m denote the total number of channels. Then, based on the information collected from downlink traffics from AP_{curr} as well as from Forwarded Probe Response frames, the station maintains an AP RSSI vector:

$$\overrightarrow{\text{RSSI}_{\text{AP}}} = \{\text{RSSI}_{\text{AP}_{1,1}}, \cdots, \text{RSSI}_{\text{AP}_{2,1}}, \cdots\cdots, \text{RSSI}_{\text{AP}_{m,1}}, \cdots\},$$

where $\text{RSSI}_{\text{AP}_{i,j}}$ is the moving average of the RSSI values for the j-th AP operating on channel CH_i, where $1 \leqslant i \leqslant m$. Then, as shown in Fig. 7, the station maps the AP RSSI vector to a channel RSSI vector:

$$\overrightarrow{\text{RSSI}_{\text{CH}}} = \{\text{RSSI}_{\text{CH}_1}, \text{RSSI}_{\text{CH}_2}, \cdots, \text{RSSI}_{\text{CH}_m}\},$$

where

$$\text{RSSI}_{\text{CH}_i} = \max_j \text{RSSI}_{\text{AP}_{i,j}}, \quad 1 \leqslant i \leqslant m.$$

[2]In Madwifi, the reported RSSI value is indeed the SNR (Signal-to-Noise Ratio) value and hence the unit is dB [15].

Note that if no AP operates on a channel, the corresponding RSSI_{CH} is set to minus infinity. Then, the channels are sorted into two groups: G_{H} and G_{L}, where

$$G_{\text{H}} = \left\{\text{all CH}_i \text{ with } \text{RSSI}_{\text{CH}_i} \geqslant \min\{\text{TH}_{\text{RSSI}}, \text{RSSI}_{\text{AP}_{\text{curr}}}\}\right\};$$

$$G_{\text{L}} = \left\{\text{all CH}_i \text{ with } \text{RSSI}_{\text{CH}_i} < \min\{\text{TH}_{\text{RSSI}}, \text{RSSI}_{\text{AP}_{\text{curr}}}\}\right\}.$$

The threshold TH_{RSSI} is a design parameter and it is set to 15 dB in HaND. When the station chooses a channel to probe, it first chooses between G_{H} (with a probability of P_{H}) and G_{L} (with a probability of P_{L}). In our implementation, P_{H} and P_{L} are set to 0.9 and 0.1, respectively. Then, one of the channels from the chosen group is randomly selected (with a uniform distribution) to be probed.

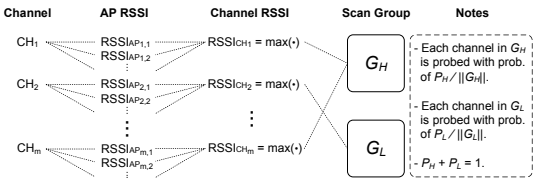


Fig. 7. HaND station selects a channel to probe on a probabilistic basis.

C. HaND AP Behaviors

The main innovations of the proposed HaND scheme reside on the AP side. In this section, we describe the behaviors of a HaND AP from the following five aspects.

1) Probe Request Overhearing Avoidance: From the experiments, we notice that an AP may overhear the Probe Request frames transmitted on other channels different from the one it operates on, due to imperfect signal filter implementation in practical 802.11 devices. Clearly, the AP should not respond to these requests since the RSSI measurement is skewed and does not reflect the actual link quality between the AP and the station that sends the request.

To filter out the overheard Probe Request frames, we propose to include in each Probe Request frame the index of the channel that it is intended to probe. Then, upon receiving a Probe Request frame, the AP checks the channel index. If it matches the AP's own operating channel, the request is accepted; otherwise it is discarded. This proposal is practically feasible and can be implemented with commodity 802.11 devices since the 802.11 management frames (e.g., Probe Request, Probe Response, etc.) usually are generated in the device driver, which is different from the 802.11 control frames (e.g., RTS, CTS and ACK) whose generation process typically is hard-coded in the device firmware. In the HaND implementation, we modify the content of *SSID Information Element (IE)* in the Probe Request frame for this purpose.

2) Information Exchange over the Wired Network: The key idea of the proposed HaND scheme is to leverage on the wired communication backbone between APs to relay the context information about wireless channels.

As shown in Fig. 6, upon receiving a Probe Request frame, AP_{new} takes two actions. Firstly, it replies a Probe Response frame immediately via its wireless interface. This is to make sure that the HaND AP is backward-compatible with regular 802.11 stations that run the default handoff procedure. Secondly, it passes the following information to a user daemon: the index of the channel it operates on, the RSSI value of the

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE INFOCOM 2010 proceedings
This paper was presented as part of the main Technical Program at IEEE INFOCOM 2010.

received Probe Request frame, and a copy of the generated Probe Response frame. Then, the user daemon updates a local table where each entry of the table corresponds to an 802.11 station in the network and it contains:

- MAC_{AP}: the MAC address of each AP it has probed;
- RSSI_{AP}: the moving average of the RSSI values of the Probe Request frames received by each AP it has probed (including the current AP);
- Contents of the most recent Probe Response frame generated for the station by each AP it has probed.

Any time when a table entry is updated, it will be sent to other APs. This is the first piece of the context information exchanged over the wired network. Another piece of context information exchanged between APs includes their own IP addresses as well as transmission rates and MAC addresses of their associated stations. This information is exchanged periodically and used by an AP to make the handoff decisions for its associated stations. Details will be discussed in Section V-C4.

3) Forwarded Probe Response Frames: After receiving the updated table entry from AP_{new}, AP_{curr} updates its local table accordingly and sends a *Forwarded Probe Response* frame to the corresponding station. Forwarded Probe Response frame is a special Probe Response frame with modified frame fields. To convey the information about AP_{new}, the following fields are modified in the Forwarded Probe Response frame: *source address* (set to the MAC address of AP_{new}), *the Length field of the DS Parameter Set IE* and *the reserved bits in the Capability Information fixed field*. Note that these fields are not used or checked by the legacy 802.11 devices; as a result, our proposed scheme will not cause any confusion or compatibility issue to the legacy 802.11 devices.

According to the IEEE 802.11 standard, a Probe Response frame contains several IEs and fixed fields, including *DS Parameter Set IE* and *Capability Information fixed field*. In the current 802.11 implementation, the eight-bit `Length` field of the DS Parameter Set IE has a constant value of one and is not used or checked in Madiwfi. In HaND, we use this field to carry the moving average of the Probe Request RSSI values that are measured by AP_{new}. On the other hand, the Capability Information fixed field currently has five reserved bits, as shown in Fig. 8. We use bits B9, B12, B14 and B15 (called `CHAN_BITS`) to carry the index of the channel AP_{new} operates on. The differences between regular Probe Response frames and Forwarded Response Frames and Triggering Response Frames (which will be discussed in Section V-C5) are summarized in Table I.

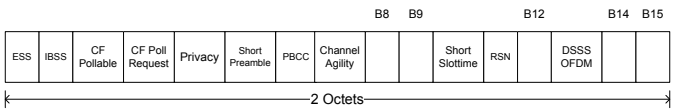


Fig. 8. Format of the Capability Information fixed field. There currently are five reserved bits: B8, B9, B12, B14 and B15. In HaND, we use B9, B12, B14 and B15 (called `CHAN_BITS`) to carry the channel index and use B8 (called `TRIG_BIT`) to specify a Triggering Probe Response frame.

4) Smart Handoff Decision with the Satisfaction-based-Fairness Heuristic: The popular *strongest-signal-first* heuristic, which ranks the candidate APs for a station (including AP_{curr} and possibly multiple neighboring APs) according to the RSSI value or its variants (such as transmission rate), has been shown to be inefficient for making handoff decisions as it may cause unbalanced loads between APs. Recent works such as [12] try to consider the AP loads in order to make proper handoff decisions. However, since none of these approaches considers the APs' detailed context information such as the transmission rates of their associated stations, they may result in performance anomaly because of the potential transmission rate diversity between the newly-associated station and the already-associated stations, as discussed in Section III-D.

TABLE I
DIFFERENCES BETWEEN REGULAR, FORWARDED AND TRIGGERING PROBE RESPONSE FRAMES

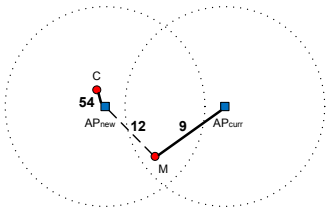
Frame Type	`Length` field in DS Parameter Set	`CHAN_BITS` in Capability Information fixed field	`TRIG_BIT` in Capability Information fixed field
Regular Probe Resp. (Default)	1	all zeros	0
Forwarded Probe Resp.	RSSI measure	channel index	0
Triggering Probe Resp.	RSSI measure	channel index	1

To deal with this issue, we propose a new heuristic for making handoff decisions, called *satisfaction-based fairness*. The key idea is to emphasize max-min fair bandwidth satisfactions among all stations. *Bandwidth satisfaction level* is not an absolute bandwidth measurement (in Mbps) but a percentage value (no unit). A mobile station may have the options of staying associated with AP_{curr} or re-associating with one of the new APs. Its handoff decision may affect its own as well as other stations' bandwidth allocations. For a certain handoff option, the bandwidth satisfaction level of a station is defined as the ratio of its expected bandwidth allocation to its maximum attainable bandwidth allocation among all possible handoff options, based on a worst-case-scenario assumption that all stations are transmitting continuously to saturate the network. In HaND, we use a simple model specified in [16], which ignores the transmission overheads such as contention window and backoff, to estimate stations' bandwidth allocations as follows. Suppose that there are n stations currently associating with an AP with the transmission rate of R_1, R_2, $\cdots$, R_n, respectively. Then, the bandwidth allocated to each station is $1/\left(\sum_{i=1}^{n} 1/R_i\right)$.

Now let's take a look at the example scenario shown in Fig. 9(a). AP_{curr} and AP_{new} are two APs operating on non-interfering channels, C is a static station and M is a mobile station moving from right to left. Each line in the figure represents a possible association and the number near the line represents the transmission rate (in Mbps) of the corresponding wireless link. Currently, C associates with AP_{new} and M associates with AP_{curr}. As shown in Fig. 9(b), when using the strongest-signal-first heuristic, M will re-associate with AP_{new}, resulting in a system throughput of 19.6 Mbps, which is only 31 percent of the maximum possible 63 Mbps. This performance anomaly occurs because the high-rate station C is "slowed down" by the newly-associated low-rate station M.

In comparison, using the proposed satisfaction-based-fairness heuristic, the maximum attainable bandwidth allocation for C and M is 54 Mbps ($= \frac{1}{(1/54)}$, which is achieved when M associates with AP_{curr}) and 9.8 Mbps ($= \frac{1}{(1/54)+(1/12)}$, which is achieved when M associates with AP_{new}), respectively. As a result, as shown in Fig. 9(c), M will stay associated with AP_{curr} as this will result in a higher minimum bandwidth satisfaction level among C and M, which also results in a higher overall system throughput.

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE INFOCOM 2010 proceedings
This paper was presented as part of the main Technical Program at IEEE INFOCOM 2010.



(a) network topology

AP	B_{total}	$\{R_C, R_M\}$
AP_{new} (∗)	19.6	{54, ***12***}
AP_{curr}	63	{54, ***9***}

(b) strongest-signal-first

AP	B_{total}	$\{S_C, S_M\}$
AP_{new}	19.6	{***0.18***, 1.00}
AP_{curr} (∗)	63	{1.00, ***0.92***}

(c) satisfaction-based-fairness

Fig. 9. An example to illustrate how the proposed satisfaction-based-fairness heuristic deals with the performance anomaly that exists with the strongest-signal-first heuristic. For each heuristic, candidate APs for M to associate with and the corresponding transmission rate (denoted as R) or bandwidth satisfaction level (denoted as S) are compared. The association decisions are marked with an asterisk. B_{total} is the overall system throughput.

To implement HaND with the proposed satisfaction-based-fairness heuristic, two pieces of information are needed: (i) the transmission rates of all candidate APs' associated stations; and (ii) the potential transmission rate of the station with each candidate AP. While the former one is exchanged periodically between APs, as discussed in Section V-C2, the latter one is not directly available. Instead, AP_{curr} only has the information about the RSSI value of the station to each candidate AP. Although the potential transmission rate depends on not only the received signal strength, but also a variety of other factors such as rate adaptation scheme and channel dynamics, results in [12], [17] show that a simple mapping between RSSI values and transmission rates can yield a satisfactory performance. In HaND, we use the mapping in Table II, which is obtained from our experiments and consistent with [17].

TABLE II
MAPPING BETWEEN RSSI VALUES AND TRANSMISSION RATES

RSSI Range (dB)[3]	0 ≀ 5	5 ≀ 8	8 ≀ 12	12 ≀ 15	15 ≀ 18	18 ≀ 22	22 ≀ 27	27 ≀ 33	33 ≀ 60
PHY Rate (Mbps)	1	2	5.5	12	18	24	36	48	54

[3] Please refer to Footnote 2.

5) Triggering Probe Response Frames: Once AP_{curr} decides that the station should re-associate with a new AP, it triggers the handoff procedure by sending a special *Triggering Probe Response* frame to the station. As shown in Table I, Triggering Probe Response frame has the same format as Forwarded Response Frame except that the B8 bit (called `TRIG_BIT`) of the Capability Information fixed field is set to one. Upon receiving it, the station starts the re-association process.

VI. EXPERIMENTAL STUDY

We have implemented HaND in Madwifi [13]. In this section, we evaluate its effectiveness using experimental results.

A. Experimental Setup

All the experiments are conducted on the third floor of our department building. The (logical) network topology, the hardware configurations, and the trajectory of the mobile station are shown in Fig. 10. We use off-the-shelf hardware instead of sophisticated equipments to conduct experiments as this makes our experimental results comparable to what users of commodity 802.11 devices may expect in realistic scenarios.

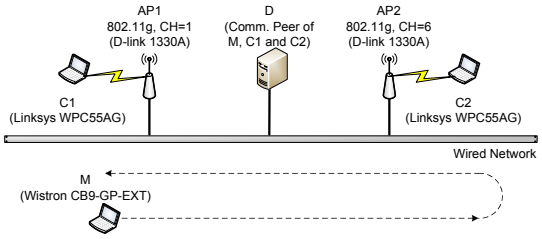


Fig. 10. Experimental setup.

As shown in the figure, both APs operate in the 802.11g mode but on different channels (Channel 1 and Channel 6) and the wired communication backbone between them is a 100 Mbps Ethernet. Two static stations C_1 and C_2 are very close to and associate with AP_1 and AP_2, respectively. A mobile station M moves and hands off between two APs and its trajectory is shown as the dashed curve. A desktop computer D that is connected to the Ethernet serves as the communication peer for M, C_1 and C_2. We use Iperf [18] as the UDP packet generator to measure the throughput performance; CBR (Constant Bit Rate) traffic is generated with the packet size of 1470 octets. We use ICMP Echo Request packets generated by M to measure the delay performance in terms of RTT (Round Trip Time) between M and D. The SampleRate [19] rate adaptation scheme is used in all stations.

We conduct the experiments in the following scenarios. In the first set of experimental scenarios, the traffic loads on both APs are equal, while in the second set, AP_1 has heavier loads than AP_2. The results for each scenario are averaged over five experimental runs. In order to minimize potential unexpected performance variation caused by people's movement and interference from other 802.11 and Bluetooth devices, all experiments are conducted at nighttime or weekends.

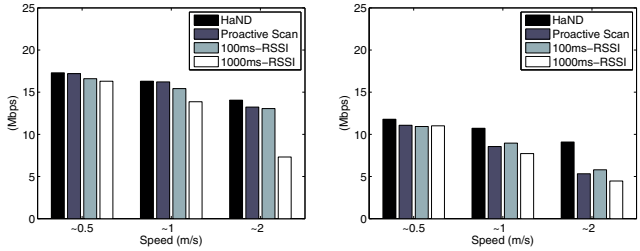
B. Scenario I: Balanced AP Loads

In this scenario, two APs have the same traffic load, i.e., C_1 and C_2 generate the UDP traffic at the same application-layer rate, which varies from zero (i.e., C_1 and C_2 have no data communications with their associated APs) to 20 Mbps (i.e., both APs are heavily loaded).

We compare the performance of the proposed HaND scheme against the following schemes: (i) the *Proactive Scan* scheme proposed in [6]; and (ii) two naive handoff schemes with different scan intensity levels: *100ms-RSSI* which probes one channel per scan phase and has a scan interval of 100 ms, and *1000ms-RSSI* which probes one channel per scan phase and has a scan interval of 1000 ms. Channel dwell time is set to 5 ms in these schemes. Moreover, in these schemes, the mobile station makes the handoff decisions by ranking the candidate APs according to the RSSI value or its variants (such as transmission rate). In comparison, channel dwell time is zero in HaND and the handoff decisions are made by the mobile station's currently-associated AP.

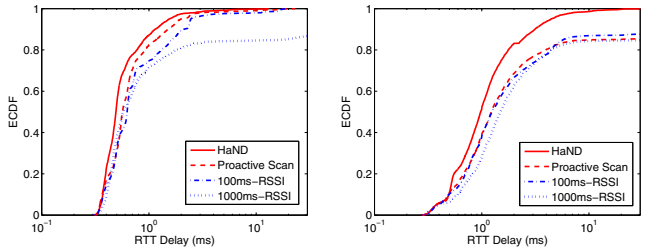
1) Throughput Performance: We first compare the throughput performance of M when different testing schemes are used. M generates UDP traffic at a CBR rate of 20 Mbps and communicates with D via its associated AP. We vary the speed of M from around 0.5 m/s (slow walking) to 1 m/s (normal walking) to 2 m/s (running). Experimental results are plotted in Fig. 11. We have the following observations.

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE INFOCOM 2010 proceedings
This paper was presented as part of the main Technical Program at IEEE INFOCOM 2010.



(a) zero AP load (b) 20 Mbps AP load

Fig. 11. Throughput performance of M under balanced AP loads.



(a) zero AP load (b) 20 Mbps AP load

Fig. 12. Delay performance of M under balanced AP loads.

In general, all testing schemes yield lower throughput when (i) APs are more heavily loaded; or (ii) the speed of M goes up. The former one is because of the contention nature of the IEEE 802.11 protocol and M has to contend with static stations to receive the services from the AP, while the latter one is due mainly to the ineffectiveness of the SampleRate rate adaptation scheme in the presence of high fluctuation of channel conditions in mobile environments.

HaND outperforms all other schemes in all experiments, particularly when the AP load is high and/or M moves fast. As shown in Fig. 11(b), in the running situation under 20 Mbps AP load, the throughput of M with the proposed HaND scheme is almost twice that with other schemes. We suspect that the poor performance of Proactive Scan, 100ms-RSSI and 1000ms-RSSI may be due to the service outage experienced by M with these schemes when the AP load is high and/or M moves fast. In order to have a better understanding of this interesting phenomenon, we conduct more experiments to study the delay performances of testing schemes.

2) Delay Performance: We conduct the following experiment with M moving at the running speed (i.e., around 2 m/s). Instead of using the UDP data traffic, M generates ICMP Echo Request packets (at an interval of 20 ms), which provide straightforward RTT (Round Trip Time) measurements.

Fig. 12(a) compares the ECDFs (Empirical Cumulative Distribution Functions) of RTTs with different testing schemes when there is no loads on the APs. Both HaND and Proactive Scan perform better than others due to the dynamic adjustment of the scan interval to the link quality between M and its currently-associated AP. HaND yields better performance than Proactive Scan because of its unique zero-channel-dwell-time architecture, which reduces the service disruption period further. In comparison, 100ms-RSSI enters the scan phase at a fixed interval of every 100 ms and hence disrupts the data services more often. 1000ms-RSSI performs the worst among all testing schemes. This is because it enters the scan phase every 1000 ms. Therefore, when M moves fast, it may be slow in collecting the information about nearby APs. As a result, M may have already lost the connection to its current AP before re-associating with a new one, i.e., service outage may occur. As shown in the figure, about 14.9 percent of the packets experience infinite RTT, meaning that they are transmitted during the service outage. This explains the low throughput performance of 1000ms-RSSI in the running situations.

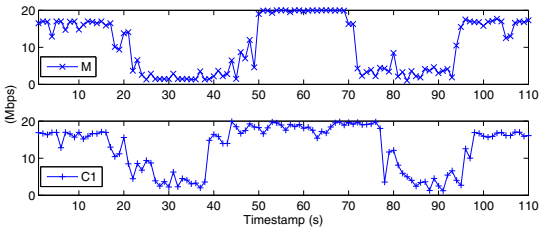
In Fig. 12(b), the AP load is increased to 20 Mbps and the service outage phenomenon can be observed more clearly from the experiments. When the AP is heavily loaded, in addition to 1000ms-RSSI, Proactive Scan and 100ms-RSSI also experience service outage as well. The reason is as follows. Service outage may have two different causes: (i) it may be caused by a long channel scan interval, as explained in the previous paragraph for 1000ms-RSSI; or (ii) it also may occur if the Probe Responses are not received by the station. In our experiment, a short channel dwell time of 5 ms may not be long enough for M to collect the returning Probe Responses from the AP since the AP is busy serving other stations with a load of 20 Mbps. In comparison, HaND continues to perform well (with zero service outage) as it does not rely on dwelling on a new channel to collect Probe Response frames; instead, it leverages on the wired communication backbone between APs to relay the information about the new AP through the station's currently-associated AP. This is one of the key advantages of the proposed HaND scheme.

C. Scenario II: Unbalanced AP Loads

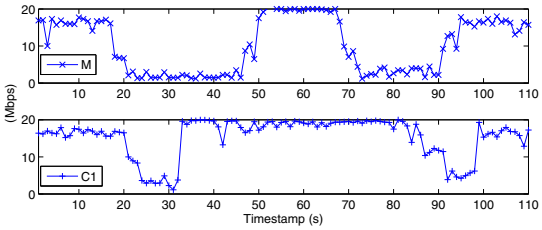
In this section, we study the performance of the proposed satisfaction-based-fairness heuristic used by a mobile station's current AP to make the handoff decisions. In the experiment, two APs are set up with unbalanced loads as follows: C_1 generates the UDP traffic at a CBR rate of 20 Mbps while C_2 remains idle all the time. We vary the speed of M from around 0.5 m/s to 2 m/s. We compare the performance of HaND with a variant of HaND called HaND-RSSI, which differs from HaND in the handoff triggering heuristic. In HaND-RSSI, the station's currently-associated AP makes the handoff decisions based on the RSSI values of the candidate APs.

Fig. 13 plots the snapshots of the instant throughputs of M and C_1 as well as the overall system instant throughput for HaND and HaND-RSSI, respectively, when M moves at around 0.5 m/s. Results with HaND-RSSI are shown in Fig. 13(a). Initially, M is close to and associates with AP_1. As it moves away from AP_1, its transmission rate decreases as dictated by the adopted SampleRate rate adaptation scheme. As a result, the instant throughput of M starts dropping. Meanwhile, the instant throughput of C_1 also starts dropping since the the IEEE 802.11 protocol is designed to provide equal access to the shared channel among stations associating with the same AP. At around the 40-second mark, the signal strength of AP_2 becomes higher than that of AP_1. Hence, M triggers a handoff and re-associates with AP_2. Now M and C_1 associates with different APs without contending with each other. As a result, the instant throughput of both stations increases immediately. This can be seen clearly from the figure. M continues to move and at around the 78-second mark, M triggers another handoff and re-associates with AP_1, resulting in another drop of the instant throughput of C_1. As M moves closer to AP_1, the link quality between them gets

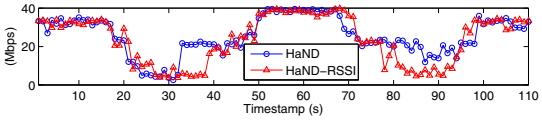
This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE INFOCOM 2010 proceedings
This paper was presented as part of the main Technical Program at IEEE INFOCOM 2010.



(a) Snapshot of instant throughout of HaND-RSSI



(b) Snapshot of instant throughout of HaND



(c) System instant throughout (Mbps)

Fig. 13. Snapshots of instant throughput under unbalanced AP loads. The speed of the mobile station M is around 0.5 m/s.

better. At around the 98-second mark, both M and C_1 transmit at high rates and receive high-throughput data services.

In comparison, the instant throughputs of HaND are shown in Fig. 13(b). Since M's currently-associated AP makes the handoff decisions based on the bandwidth satisfaction level, we observe that when M moves away from AP_1, a handoff is triggered at an earlier time instance than HaND-RSSI at approximately the 32-second mark. This is because at that moment, AP_1 realizes that M has become a low-rate station and it is a better option to handoff M to AP_2 to prevent M from "slowing down" the high-rate station C_1. As a result, between 32 and 40 seconds, HaND yields a higher system throughput than HaND-RSSI, which is shown in Fig. 13(c). Similarly, when M moves back towards AP_1, HaND triggers a handoff later than HaND-RSSI, resulting a higher system throughput between 78 and 86 seconds.

We also compare the system throughput performance of HaND and HaND-RSSI at different speeds of M in Table III. We observe that with slow or normal mobility (at 0.5 or 1 m/s), HaND outperforms HaND-RSSI by about 17%. However, it is interesting to see that, with high mobility (at 2 m/s), the performances of HaND and HaND-RSSI are similar. This is probably due to the following reasons: (i) the time period during which HaND outperforms HaND-RSSI shrinks as the moving speed increases; and (ii) the default SampleRate rate adaptation scheme is unable to track the dynamic channel variation well in mobile environments. The latter issue has been recognized by several works on rate adaptation. We expect that a larger performance gain may be achieved if the station adopts a more responsive rate adaptation scheme to work with the proposed HaND scheme.

VII. Conclusions

From experiments, we find that channel dwell time contributes to a major part of the handoff delay and it is difficult to set a proper channel dwell time for an 802.11 station since the context information about the scanned channel, such as the number of APs on the channel and the loads of the APs, is unavailable to the station. To address this challenge, we propose a practical fast handoff scheme, called HaND, which adopts a novel zero-channel-dwell-time architecture to reduce the handoff delay. Moreover, in HaND, the handoff decisions are made by a station's currently-associated AP (rather than the station itself) based on a new satisfaction-based-fairness heuristic. We have implemented HaND in Madwifi and experimental results show that HaND outperforms other testing schemes in terms of both throughput and delay.

TABLE III
Comparison of System Throughput (in Mbps) under Unbalanced AP Loads

Speed of M	∼0.5 m/s	∼1 m/s	∼2 m/s
HaND	27.49	25.86	18.66
HaND-RSSI	23.48	22.51	18.31

References

[1] IEEE 802.11, *Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*. Standard, IEEE, Aug. 1999.
[2] A. Mishra, M. Shin, and W. Arbaugh, "An empirical analysis of the ieee 802.11 mac layer handoff process," in *ACM Computer Communications Review*, vol. 33, no. 2, 2003.
[3] V. Mhatre and K. Papagiannaki, "Using smart triggers for improved user performance in 802.11 wireless networks," in *ACM Mobisys'06*, 2006.
[4] M. Shin, A. Mishra, and W. Arbaugh, "Improving the latency of 802.11 hand-offs using neighbor graphs," in *ACM Mobisys'04*, 2004.
[5] J. Teng, C. Xu, W. Jia, and D. Xuan, "D-scan: Enabling fast and smooth handoffs in ap-dense 802.11 wireless networks," in *IEEE InfoCom Miniconference'09*, 2009.
[6] H. Wu, K. Tan, Y. Zhang, and Q. Zhang, "Proactive scan: Fast handoff with smart triggers for 802.11 wireless lan," in *IEEE InfoCom'07*, 2007.
[7] H. Velayos and G. Karlsson, "Techniques to reduce the ieee 802.11b handoff time," in *IEEE ICC'04*, 2004.
[8] I. Ramani and S. Savage, "Syncscan: Practical fast handoff for 802.11 infrastructure networks," in *IEEE InfoCom'05*, 2005.
[9] Y. Liao and L. Gao, "Practical schemes for smooth mac layer handoff in 802.11 wireless networks," in *IEEE WoWMoM'06*, 2007.
[10] A. Mishra, M. Shin, and W. Arbaugh, "Context caching using neighbor graphs for fast handoffs in a wireless network," in *IEEE InfoCom*, 2004.
[11] A. Giannoulis, M. Fiore, and E. Knightly, "Supporting vehicular mobility in urban multi-hop wireless netwroks," in *ACM MobiSys'08*, 2008.
[12] R. Murty, J. Padhye, R. Chandra, A. Wolman, and B. Zill, "Designing high performance enterprise wi-fi networks," in *USENIX NSDI'08*, 2008.
[13] Multiband Atheros Driver for Wifi, http://www.madwifi.org/.
[14] M. Heusse, F. Rousseau, G. Berger-Sabbatel, and A. Duda, "Performance anomaly of 802.11b," in *IEEE InfoCom'03*, 2003.
[15] RSSI in Mafwifi, http://madwifi-project.org/wiki/UserDocs/RSSI.
[16] Y. Bejerano, S. J. Han, and L. E. Li, "Fairness and load balancing in wireless lans using association control," in *ACM MobiCom'04*, 2004.
[17] J. Zhang, K. Tan, J. Zhao, H. Wu, and Y. Zhang, "A practical snr-guided rate adaptation," in *IEEE InfoCom'08*, 2008.
[18] Iperf, http://dast.nlanr.net/projects/Iperf.
[19] J. Bicket, "Bit-rate selection in wireless networks," Master's thesis, MIT, 2005.