Techniques to reduce the IEEE 802.11b handoff time

Héctor Velayos Gunnar Karlsson
KTH, Royal Institute of Technology
P.O. Box Electrum 229
SE-16440 Kista, Sweden
{hvelayos,gk}@imit.kth.se

***Abstract*—We propose and evaluate via simulations techniques to minimize the IEEE 802.11b handoff time. We describe the handoff procedure and divide it into three phases. Our main contribution is a set of techniques to reduce the two longer phases, detection and search**

I. INTRODUCTION

Wireless LANs based on the IEEE 802.11b standard are the predominant option for wireless access to the Internet. The performance of the cells permits the use of real time services, such as voice over IP, when admission control is added and the MAC scheduler is modified [1]. However, experimental measurements in our testbed, which are summarized in Table 1 and described later, indicate that current implementations of link-layer handoff do not meet the needs of real time traffic. In this paper, we propose and evaluate via simulations techniques to minimize the IEEE 802.11b handoff time.

Table 1: Link-layer handoff time for different IEEE 802.11b cards

	D-Link 520	Spectrum24	ZoomAir	Orinoco
Detection	1630 ms	1292 ms	902 ms	1016 ms
Search	288 ms	98 ms	263 ms	87 ms
Execution	2 ms	3 ms	2 ms	1 ms
Total	**1920 ms**	**1393 ms**	**1167 ms**	**1104 ms**

II. HANDOFF PROCEDURE AND MEASUREMENTS

Link-layer handoff is the change of the access point (AP) to which a station is connected. In the case of IEEE 802.11b wireless LANs it implies an interruption of data frame transmission. The duration of this interruption is called handoff time. Although buffering and routing update lead to different handoff time for uplink and downlink traffic, several authors have proposed solutions to make them equal [2] [3]. Thus, we assume that downlink and uplink handoff times are the same.

We propose to analyze the handoff process by splitting it into three sequential phases: detection, search and execution. The detection phase is the discovery of the need for the handoff. The search phase covers the acquisition of the information needed to perform the handoff. Finally, the handoff is performed during the execution phase.

The duration of each phase was measured in our testbed in which stations performed link-layer handoffs. Four commercial IEEE 802.11b cards with different chipsets were selected to measure their handoff time as an average of 10 repetitions. During the tests, the only traffic in the cells was a flow of VoIP packets generated by the station.

We noted in preliminary measurements that commercial wireless LAN cards start the search phase when the strength of the received signal degrades below a certain threshold. Since we were interested in measuring the performance of the handoff including the link-layer detection (i.e. without support from the physical layer), the handoff was forced by abruptly switching off the radio transmitter of the AP to which the station was connected. Handoff measurements using physical layer information have already been reported by Mishra et al. [4].

Our handoff measurements are summarized in Table 1. From them we can draw the following conclusions. First, different stations showed different performance, but none matched the delay requirements of real time applications. Second, detection is the longest phase, while execution could be neglected. And third, detection and search times widely vary among different models. The length differences in detection and search could be explained by analyzing the frames captured during the handoffs. The number of failed frames is the main factor in controlling the duration of the detection phase and varies with each card model. Regarding the search phase, the duration's variance is due to the different number of probe requests sent per channel and more significantly due to the time to wait for probe responses.

The main conclusion from our measurements is that detection and search phases are the main contributors to handoff time. Therefore, we suggest how to reduce them in the following sections.

III. REDUCING THE DETECTION PHASE

Stations have to detect the lack of radio connectivity based on weak received signal reported by the physical layer or failed frame transmissions. QoS concerned stations implement the former method because no frames are lost. This method assumes that there is a better AP in range as soon as the received signal gets weak. In contrast, the latter method produces less handoff events because the handoff is not triggered by temporary radio fading but only when transmission is actually interrupted.

Our study focuses on the optimization of the second method. Its main difficulty is to determine the reason for frame failure among collision, radio signal fading, or the station being out of range. We have observed in our measurements that stations firstly assume collision and retransmit several times. If transmission remains unsuccessful, then radio fading is assumed and the link is probed by sending probe requests. Only after several unanswered requests, the station declares the out of range status and starts the search phase. As Table 1 indicates, this type of detection procedure tends to be long, so we suggest a different approach: stations must start the search phase as soon as collision can be excluded as reason for failure. If the actual reason was a temporary signal fading, the selected access point after the search would likely be the current one and the handoff will not be executed. Thus, a key factor in our detection algorithm is the number of collision that a frame can suffer before it is transmitted. Let C be the random variable representing the number of collisions per successfully transmitted frame. Its cumulative distribution function (CDF) is given by:

$$\Pr\text{ob}(C \le k) = \sum_{i=0}^{k} (1-p)p^i = 1 - p^{k+1} \qquad (1)$$

Where *p* is the probability, seen by the station, that its transmitted frame collides. This probability depends on the number of stations competing for the medium, and it can be calculated with the non-linear system reported by Bianchi in [5] for saturated conditions (i.e. all stations always have a frame ready to transmit) that is the worst case for collisions. The CDF of the number of collisions per transmitted frame is plotted in Figure 1. This figure shows that three consecutive collisions is a rare event, even in saturation. Therefore, our link-layer detection algorithm can be formulated as follows: if a frame and its two consecutive retransmissions fail, the station can discard collision as the cause of failure and start the search phase. There is no need to explicitly probe the link. In the same conditions we used during our measurements, this time would be around 3 ms, which is approximately 300 times shorter than the fastest measured detection phase.

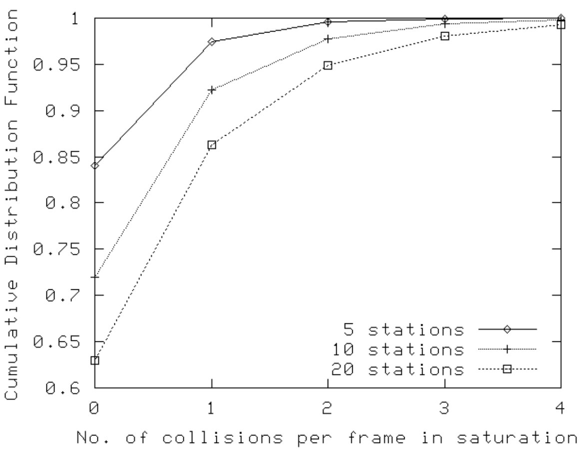

Figure 1: No. of collisions per transmitted frame in saturation

IV. REDUCING THE SEARCH PHASE

The search phase includes the actions performed by the station to find all APs in range. The standard specifies two methods to scan a channel, active and passive scanning. In passive scanning, stations listen to each channel for the beacon frames. When faster scanning is needed, stations must perform active scanning. It means that stations broadcast a probe-request frame and wait for probe responses. The time to wait for responses depends on the channel activity after the probe transmission. If the channel is idle during *MinChannelTime*, the scanning is finished and the channel is declared empty. If there is any traffic during this time, the station must wait *MaxChannelTime*. MaxChannelTime should be large enough as to allow the APs to compete for the medium and send the probe response. Both MaxChannelTime and MinChannelTime are measured in steps of 1024 microseconds called *Time Unit* (TU).

Despite that MinChannelTime and MaxChannelTime control the duration of the scanning, the IEEE standard does not specify their values. We calculate them below to minimize the search phase. Firstly, we compute MinChannelTime that is the maximum time an AP would need to answer given that the AP and channel were idle. If propagation time and probe response generation time are neglected, the IEEE 802.11b medium access function establishes that the maximum response time is:

$$MinChannelTime = DIFS + (aCWmin \times aSlotTime) \qquad (2)$$

Where *DIFS* is the Distributed InterFrame Space, *aCWmin* is the maximum number of slots in the minimum contention window, and *aSlotTime* is the length of a slot. Table 2 contains these values for the IEEE 802.11b standard. Inserting them in (2), we obtain 670 μs. Since MinChannelTime must be expressed in *Time Unit*, we can conclude that MinChannelTime should be one TU (i.e. 1024 μs).

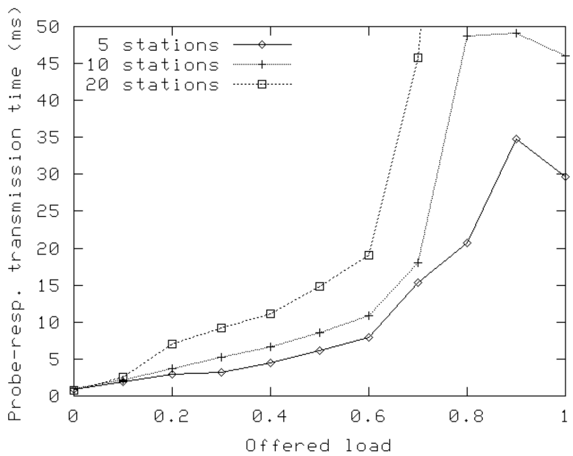

Figure 2: Probe response transmission time (ms)

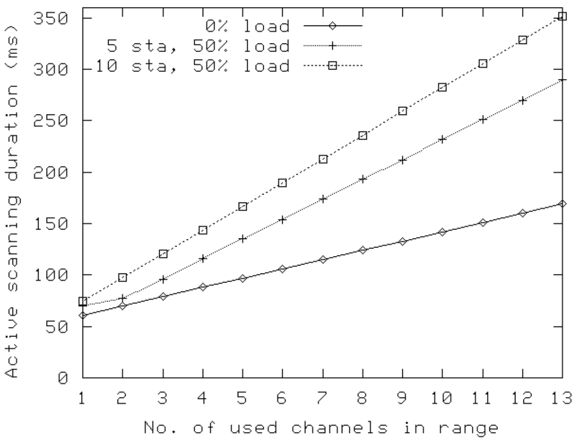

Figure 3: Total search time (ms)

Table 2: Physical characteristics for IEEE 802.11b standard

	IEEE 802.11b
aSlotTime	20 µs
aCWmin	31 slots
DIFS	50 µs

The calculation of MaxChannelTime is more complex. It is the maximum time to wait for a probe response when the channel is being used. In order to find an upper bound for MaxChannelTime, we have run simulations with NS-2 to measure the time to transmit the probe response. Figure 2 presents the results of our simulations.

Our simulations indicate that the transmission time of a probe response depends on the offered load and number of stations. They also show that MaxChannelTime is not bounded as long as the number of stations can increase. We suggest then to set a value for MaxChannelTime that would prevent overloaded access points to answer in time. Since 10 stations transmitting per cell seems to be an adequate number to achieve a good cell throughput [5], Figure 2 indicates that 10 TU (10.24 ms) would be a reasonable choice for MaxChannelTime.

Finally, the total search time *s* that includes the time to scan all available channels can be calculated as:

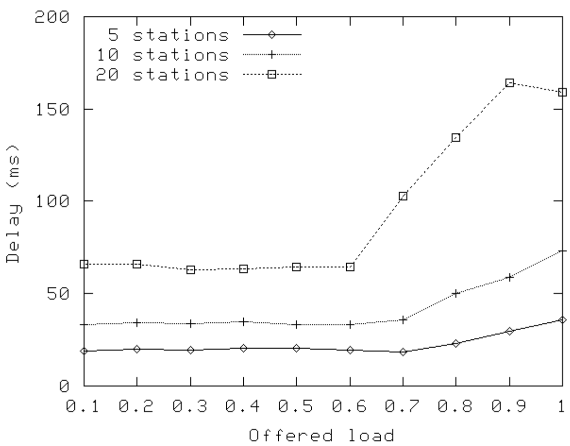

Figure 4: Delay versus load

$$s = uT_u + eT_e \quad (3)$$

Where u is the number of channels with traffic and T_u is the time needed to scan a used channel. Respectively, e is the number of empty channels and T_e is the time to scan an empty channel. We can now determine T_u and T_e. When a channel is scanned, the probe request is sent to the broadcast address, so its reception will not be acknowledged. Therefore, at least two probe requests must be sent to overcome a possible collision. Let T_d be the transmission delay to send each probe, then we can calculate T_u and T_e as:

$$\begin{aligned} T_u &= 2\,T_d + MaxChannelTime \\ Te &= 2\,T_d + MinChannelTime \end{aligned} \quad (4)$$

Total search time can be calculated with (3) and (4), as well as the transmission delay. Figure 3 shows the total search time versus number of used channels in range for different load conditions. To plot it, we obtained T_d from our delay simulations reported in Figure 4. In Figure 3, we included a no-load case that is comparable with our measurements conditions reported in Table 1. This case shows that the search time can be reduced to 70 ms when handing over between two APs, which is 20% faster than the shortest search phase measured.

V. CONCLUSIONS

We have measured, analyzed and suggested how to reduce the link-layer handoff time in IEEE 802.11b networks. The handoff process was split into three sequential phases: detection, search and execution. We have shown that the link-layer detection phase can be reduced to three consecutive non-acknowledged frames, which is approximately 300 times shorter than the fastest measured detection phase. We have also shown that using active scanning with its timers MinChannelTime and MaxChannelTime set to 1 ms and 10.24 ms respectively reduces the search phase by 20% compared to the shortest measured one.

REFERENCES

[1] M. Barry, A. T. Campbell, A. Veres, "Distributed control algorithms for service differentiation in wireless packet networks", Proc. IEEE INFOCOM 2001, Alaska.

[2] Amre El-Hoiydi, "Implementation options for the distribution system in the 802.11 Wireless LAN Infrastructure Network", Proc. IEEE ICC2000, New Orleans.

[3] Anne H. Ren, Gerald Q. Maguire Jr., "An adaptive real-time IAPP protocol for supporting multimedia communications in wireless LAN systems", Proc. of ICCC99, Japan, 1999.

[4] Arunesh Mishra, Minho Shin, Willian Arbaugh, "An empirical analysis of the IEEE 802.11 MAC layer handoff process", University of Maryland Technical Report, UMIACS-TR-2002-75, 2002.

[5] Giuseppe Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function", IEEE JSAC, 2000.