International Journal of Parallel, Emergent and Distributed Systems
Vol. 23, No. 4, August 2008, 1–20

A new access point selection policy for multi-rate IEEE 802.11 WLANs

Murad Abusubaih, Sven Wiethoelter, James Gross, Adam Wolisz
Telecommunication Networks Group
Technical University Berlin
Email: {abusubaih}@tkn.tu-berlin.de

(In final form August 2008)

In wireless local area networks often a station can potentially associate with more than one access point. Therefore, a relevant question is which access point to select "best" from a list of candidate ones. In IEEE 802.11, the user simply associates to the access point with the strongest received signal strength. However, this may result in a significant load imbalance between several access points. Moreover, the multi-rate flexibility provided by several IEEE 802.11 variants can cause low bit rate stations to negatively affect high bit rate ones and consequently degrade the overall network throughput. This paper investigates the various aspects of "best" access point selection for IEEE 802.11 systems. In detail, we first derive a new decision metric which can be used for AP selection. Using this metric we propose two new selection mechanisms which are decentralized in the sense that the decision is performed by each station, given appropriate status information of each access point. In fact, only few bytes of status information have to be added to the beacon and probe response frames which does not impose significant overhead. We show that our mechanism improves mean quality of service of all stations and better utilizes network resources compared to the conventional one implemented today in IEEE 802.11 devices. Also, the schemes are appealing in terms of stability and provide their performance improvement even for denser or lighter network configurations.

Keywords: IEEE 802.11; Wireless LANs; Access Point Selection, Load Balancing

1 Introduction

Over the last years wireless local area networks (WLANs) [1] have become quite popular. Due to decreasing costs of the equipments (wireless access points–APs–and wireless network cards) and fixed broadband connections (digital subscriber lines), WLANs have become the preferred technology of access in homes, offices, and hot-spot areas (like airports and meeting rooms). Although originally several standards for WLAN have been competing, today virtually all WLANs are based on the IEEE 802.11 standard. This technology provides users with raw transmission rates of up to 54 Mbps in IEEE 802.11a/g and 11 Mbps in IEEE 802.11b, while even "faster" standard supplements are currently under discussion. For the future, a further increase in wireless local area networks can be expected as, for example, the city of Chicago and the Bay Area currently consider the roll out of a city-wide WLAN based on IEEE 802.11 [2].

In current implementations of IEEE 802.11, a STA has to first pick (and associate with) an AP before it can access data transmission services of the WLAN cell. This process can be performed actively or passively and is referred to as scanning. In active scanning a STA sends a “Probe Request” frame and the AP replies with a “Probe Response” frame. This frame exchange allows the STA to obtain basic information about the cell like signal strength, available transmission modes, encryption etc. The frame exchange is repeated for all APs in the vicinity, such that the STA has a list of APs at the end of the scanning process. Alternatively, in passive scanning a STA listens to “Beacon” frames which are periodically transmitted by APs. After scanning (either actively or passively) multiple APs, a STA always associates to the AP from which it has received the strongest signal. Afterwards, it stays associated until the STA is powered down or the AP shuts down its service.

It is obvious that this rather “simple” selection process can lead to problems regarding the network performance of larger areas with many STAs and several APs [3, 4]. This has initiated research activity in an area commonly referred to as ”load balancing”. Load balancing basically tries to distribute stations (either centrally or decentrally) over multiple cells such that the network and station performance increases (a tutorial discussion of load balancing in packet-switched networks is provided in [5]). The major question among the work in this area has been how to measure load most realistically in a WLAN cell. For example, in [6] it has been proposed to simply characterize the load per AP by the number of stations associated to it. While this is easy to implement, the multiple rates provided by several IEEE 802.11 variants counteract this load metric. In principle, the advantages of multi-rate protocols have been shown in [7]: Usually, a mobile STA with a relatively low signal to noise (and interference) ratio chooses a low transmission rate to balance its frame error rate. However, the 802.11 medium access control (MAC) protocol provides “per-frame fairness”, meaning that in the long term STAs have the same chance to access the medium and send their frames (all STAs should transmit with an equal average frame rate over a longer time horizon). As the time duration required to transmit a frame with a low transmission rate is much longer than the duration for the same frame size with a higher transmission rate, a low transmission rate STA will occupy the channel for a longer time. This phenomena degrades the throughput of high rate STAs if they are associated to the same AP. In [8], it has been shown that if a STA with a transmission rate of 11 Mbps shares the channel with a STA at a transmission rate of 1 Mbps, the throughput of the 11 Mbps STA is about the same as that of the 1 Mbps STA (assuming equal flow characteristics of each STA as well as the saturation mode). A simple mathematical model of this effect and the consequence for centralized, optimal load balancing has been discussed by Kumar et al. in [9]. However, the authors suppose in their study that the transmission rate per stations equals the achieved goodput which is not the case in general. A further centralized solution with a simplified load characterization can be found in [10].

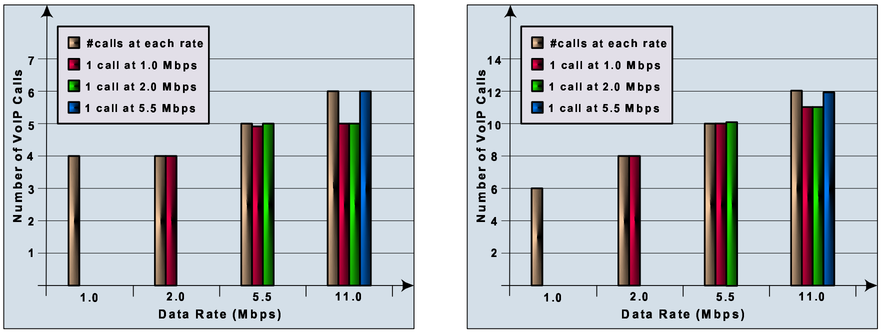
The same problem of realistic load characterization applies also to a study recently published by Ekici et al. in [11]. In contrast to Kumar, the authors propose a totally decentralized approach to load balancing where each station decides independently on its association. This decentralized approach is also referred to as AP selection. However, while the authors take the multi-rate effect into account, again the transmission rate per terminal is assumed to equal goodput.

In this paper we present a more realistic approach to AP selection. We identify the core problem of AP selection to be the choice of metric to consider and whether the choice should be (periodically) reevaluated or not. Obviously, a more effective AP selection mechanism significantly improves the overall network performance while also improving the STAs rates and delays. We firstly motivate a novel metric for AP Selection by showing that even under non-saturated conditions, low-rate stations have a significant impact on WLAN cell's performance. Secondly, a novel metric is derived that is beneficial in combination with a new static AP selection scheme for IEEE 802.11 WLANs. This new metric encapsulates several important cell and connection parameters into a single value. These parameters are easily distributed by the APs via the Beacon and Probe Response frames. Thirdly, the static scheme is enhanced towards a periodic reevaluation of the current association decision which we refer to as dynamic AP selection scheme in the following. For the static as well as the dynamic scheme, this work demonstrates the efficiency of both approaches for IEEE 802.11b WLANs.

The remainder of this paper is organized as follows: Section 2 illustrates the multi-rate effect for an example IEEE 802.11b cell. Section 3 describes the system model and the basic assumptions, derives the new metric, and presents our static as well as dynamic selection mechanism. Finally, we evaluate the performance increase of our schemes in Section 4 before we conclude our paper in Section 5.

2 Motivation for novel metric design

In IEEE 802.11 WLANs, stations choose the modulation scheme for the transmission of an MPDU dependent on channel conditions. This rate adaptation has not been standardized by IEEE, thus vendors use proprietary schemes.



(a) G.729-coded VoIP with 10 *ms* packetization

(b) G.729-coded VoIP with 20 *ms* packetization

Figure 1. Maximum number of calls at homogeneous rates and call reduction due to single low-rate transmissions - refer to Section 2 for a detailed description

Assuming even perfect adaptation schemes, i.e., neglecting effects due to false adaptation, the modulation choice of single stations has an impact on the performance of the whole WLAN cell. Heusse et al. [8] present theoretical analysis validated by simulation experiments for TCP as well as UDP traffic. Here, we illustrate this impact by an example result from VoIP simulations (see [17] for the details and more results). Consider a single WLAN cell where multiple stations are associated to the AP. Every station has one ongoing VoIP call which it tries to convey via the IEEE 802.11b system with an acceptable quality. Dependent on the average channel condition the stations set their rate accordingly. Four transmission rates with 11, 5.5, 2.0 and 1 Mbps can be used by each station.

Initially, the maximum number of calls at each transmission rate has been obtained assuming that each station transmits with the same rate. The maximum number of calls is determined such that VoIP QoS limits are not violated. Notice how the total number of VoIP calls that can be accommodated increases the higher the transmit rate is which all stations use. For example, in Figure 1(a) a total of 6 VoIP calls can be served by the cell if *all* stations transmit with 11 Mbps.

Next, assume that among all stations in the cell one of them cannot select the same rate as all other stations (due to a lower SNR, for example), but applies a lower rate to transmit its VoIP call. Clearly, the total amount of VoIP calls that can be served decreases, depending on how low the rate selection for the single station is. For example, if the specific station can only achieve a rate of 1 or 2 Mbps, the total amount of VoIP calls served by the cell drops in Figure 1(a) to the level as if *all* stations would transmit with 5.5 Mbps (despite the fact that the remaining stations even transmit with 11 Mbps).

These results show that especially stations with low transmission rates have a significant impact on the total cell performance as well as on the individual rate performance of other stations. While in this example the impact is moderate, it is much stronger for larger frame sizes.

3 Improving AP Selection

The key aspect for AP selection is the design of an appropriate decision metric. In this section, we first describe the system model and discuss afterwards a much more appropriate metric for AP selection. Then, we present two schemes and discuss how to implement the selection mechanisms based on the proposed metric.

3.1 *System model*

We consider an area where several different access points compliant to IEEE 802.11 offer service to STAs. Each AP forms a cell, whereby cells of adjacent APs overlap significantly. A STA k might decide to request service from some distinctive access point. In this case it associates to the access point and may start to send or receive data afterwards. Denote the number of STAs associated at time t to access point i by $U_i^{(t)}$. We assume that all STAs continuously have data frames queued for transmission, i.e., we consider the saturation mode.

For the medium access we consider only the distributed coordination function (DCF) (thus, the medium access is governed by the CSMA/CA protocol). Each cell is in infrastructure mode, hence, all data transmission involves the access point even if one STA might transmit data to some other STA in the same cell.

For the physical layer, we assume similar wireless channel conditions in both uplink and downlink directions. Depending on the channel gain between any transmitter and receiver pair, the transmission rate is selected from the available rates of the physical layer. This rate is denoted by R_k Mbps, where k denotes the STA.

Traditionally, if a STA is powered up, it first scans the current available access points on all channels and chooses the one with the best signal-to-noise ratio (SNR). This choice is made once and no other parameters are considered. We refer to this scheme as the legacy AP selection scheme.

3.2 *Decision Metric for AP Selection*

Given a certain traffic characterization of a STA, ultimately a STA needs to join the cell which can serve this traffic stream best. This might be the access point which has the highest SNR. However, if the cell is crowded with other STAs operating at a much lower SNR than the currently observed STA, it might not be served very well even though the SNR indicates a good service. Furthermore, a STA might join a cell (which has the best SNR regarding this STA), but the cell is already serving some other STAs with a much higher SNR than the observed STA. Then, the new STA will significantly degrade the throughput of the other STAs. That is because APs generally consume longer time serving low rate STAs. Hence, a metric is required which can capture the impact of the STA joining the cell on the performance of the other (already joined) STAs. In addition, the metric should capture the expected throughput of the STA itself if it joins a cell.

We start deriving such a metric by considering the traffic in a certain cell i. There are currently $U_i^{(t)}$ STAs associated to the access point. Some new STA k considers the association to cell i. Hence, it has to evaluate the data rate it will receive from joining cell i. However, it also has to take into consideration how much it will degrade the data rate of other STAs, which have already joined cell i. We start with deriving the data rate STA k will receive if joining cell i.

Following the analysis in [13], STA k in cell i successfully transmits a frame of length L bits after j consecutive unsuccessful transmissions within a time period of $T_{k,i}(j)$, given by:

$$T_{k,i}(j) = T_{\mathrm{P}} + T_{\mathrm{H}} + T_{\mathrm{DIFS}} + \frac{L}{R_k} + T_{\mathrm{SIFS}} + T_{\mathrm{ack}} + T_{\mathrm{backoff}}(j) \qquad (1)$$

T_{P} and T_{H} represent the time duration of the physical layer preamble and header. T_{DIFS} is the Distributed Coordination Function Inter-frame Space and T_{SIFS} is the Short Inter Frame Spacing, $L = (28 + L_{\mathrm{MSDU}}) \cdot 8$ is the length of the MAC packet in bits (where L_{MSDU} is the MAC Service Data Unit length and the 28 bytes stem from the MAC header), $T_{\mathrm{ack}} = (T_{\mathrm{P}} + T_{\mathrm{H}} + \frac{112}{R_k})$ is the duration of the ACK frame, and $T_{\mathrm{backoff}}(j)$ is the average backoff interval in μs after j consecutive unsuccessful transmission attempts given as:

$$T_{\mathrm{backoff}}(j) = \begin{cases} \frac{2^j (T_{\mathrm{CWmin}}+1)-1}{2} \cdot T_{\mathrm{Slot}} & 0 \leq j < 6 \\ \frac{T_{\mathrm{CWmax}}}{2} \cdot T_{\mathrm{Slot}} & \mathrm{j} \geq 6 \end{cases} \qquad (2)$$

where T_{Slot} is the basic slot duration, T_{CWmin} and T_{CWmax} are the minimum and maximum contention window sizes respectively.

However, $T_{k,i}(j)$ is only the *raw* average transmission time of a frame. The frame is still subject to frame errors, which requires one or several retransmissions. The average time span that STA k requires to transmit a single frame *correctly* is [13]:

$$\overline{T_{k,i}} = T_{k,i}(0) + \sum_{j=1}^{\infty}(1 - P_{k,i})P_{k,i}^{j}\left[\sum_{m=0}^{j-1} T_f(m) + T_{k,i}(j)\right] \tag{3}$$

where $P_{k,i}$ is the frame error probability[1] and $T_f(m) = T_{\mathrm{P}} + T_{\mathrm{H}} + T_{\mathrm{DIFS}} + T_{\mathrm{backoff}}(m) + \frac{L}{R_k} + T_{\mathrm{SIFS}} + T_{\mathrm{ack}} + T_{\mathrm{Slot}}$ is the time between two consecutive transmissions if the frame transmission fails. In fact the transmission can fail due to frame errors or collisions. Assuming that frame errors are independent from collisions, we could write $P_{k,i}$ as follows:

$$P_{k,i} = Pe_{k,i} + Pc_{k,i} - Pe_{k,i}Pc_{k,i} \tag{4}$$

where $Pe_{k,i}$ is the frame error rate due to the channel and $Pc_{k,i}$ is the error rate due to collisions. In [14], Bianchi et al. derived as expression for the $Pc_{k,i}$ as follows:

$$Pc_{k,i} = 1 - (1 - \tau)^{U_i - 1} \tag{5}$$

where τ represents the probability that a station transmits in a randomly chosen time slot expressed as:

$$\tau = \frac{2(1 - 2Pc_{k,i})}{2(1 - 2Pc_{k,i})(CW_{min} + 1) + Pc_{k,i}CW_{min}(1 - (2Pc_{k,i})^m)} \tag{6}$$

where $CW_{max} = 2^m CW_{min}$. (i.e m=5 if CW_{min}=32 and CW_{max}=1024).

Bianchi proposed to solve the non linear system in 5 and 6 using numerical techniques.

Assuming that STA k is the only STA in the cell i, the fraction $L/\overline{T_{k,i}}$ would yield the average throughput of STA k in the cell (assuming also that the access point does not transmit any data). However, we assume that there are in general $U_i^{(t)}$ active STAs in cell i, therefore it is quite likely that the channel is occupied by some other STA if STA k wants to start a

[1]See Section 3.3 for details on the acquisition of the frame error rate.

data transmission. In general, STA k's throughput depends on the channel occupancy time of other STAs in the cell. We capture this effect by modeling the average rate of correctly transmitted bits by:

$$G_{k,i} = \frac{\mu_{k,i} L}{\overline{T_{k,i}}} \tag{7}$$

where $\mu_{k,i}$ is the fraction of channel time consumption "left over" for STA k, given as:

$$\mu_{k,i} = \frac{\overline{T_{k,i}}}{\overline{T_{k,i}} + \sum_{j=1}^{U_i^{(t)}} \overline{T_{j,i}}} \quad . \tag{8}$$

Recall that STA k has not joined cell i yet. Finally, we obtain the average throughput of correct bits $G_{k,i}$ (combining Equations (7) and (8)) as:

$$G_{k,i} = \frac{L}{\overline{T_{k,i}} + \sum_{j=1}^{U_i^{(t)}} \overline{T_{j,i}}} \tag{9}$$

From Equation (9) it is clear that STA k's throughput depends on the channel occupancy time of frames transmitted by other STAs in cell i (apart from other issues like the SNR and chosen rate). However, this equation does not consider the effect of STA k on the average throughput of all other STAs in cell i. The current average channel occupancy time of all STAs in cell i is simply given by:

$$\frac{\sum_{j=1}^{U_i^{(t)}} \overline{T_{j,i}}}{U_i^{(t)}} \tag{10}$$

Hence, the STA can compute the new average channel occupancy time once it joined the cell. If this new value increases, the STA can deduce that it will decrease the throughput of all STAs in the cell. If the new value decreases, the STA will not harm other STAs but could achieve a higher throughput if the other STAs had a better SNR, for example. Hence, the difference between current average channel occupancy time of the STAs in cell i and the new average channel occupancy time per STA in cell i (if STA k joined this cell) is a measure for the impact of STA k on cell i. This measure is given by:

$$\frac{\sum_{j=1}^{U_i^{(t)}} \overline{T_{j,i}}}{U_i^{(t)}} - \frac{\overline{T_{k,i}} + \sum_{j=1}^{U_i^{(t)}} \overline{T_{j,i}}}{U_i^{(t)} + 1} = \frac{\sum_{j=1}^{U_i^{(t)}} \overline{T_{j,i}} - U_i^{(t)} \cdot \overline{T_{k,i}}}{U_i^{(t)} \cdot \left(U_i^{(t)} + 1\right)} \quad (11)$$

3.3 *Static AP Selection*

To select an AP, a STA will try to maximize its throughput in the cell it wishes to associate to—but should also minimize its impact on already associated STAs. Therefore, it computes the following “impact” value for all candidate APs and sends its association request frame to the AP that maximizes this function.

$$W(i) = \alpha \frac{L}{\overline{T_{k,i}} + \sum_{j=1}^{U_i} \overline{T_{j,i}}} + (1-\alpha) \frac{\sum_{j=1}^{U_i} \overline{T_{j,i}} - U_i \cdot \overline{T_{k,i}}}{U_i(U_i+1)} \quad (12)$$

where α is a weighting factor between 0 and 1, whose value depends on the variance of (9). Therefore, a STA tries to minimize its negative effect on other STAs if its theoretical throughput from the candidate APs does not differ significantly. Note that the two parts of the cost function represent different quantities. For this reason, each part has been normalized by the maximum among all values obtained from the different APs.

The first part of the cost function in (12) constitutes the theoretical throughput a STA will get in some cell i, while the second part is a measure of its impact on the other STAs in the same cell. To do so, a STA needs three pieces of information, which are the number of STAs the AP currently accommodates $U_i^{(t)}$, the summation value in equation (9) and the current SNR between AP and STA. The AP could include the first two values in a new information field in the Beacon and Probe Response frames. Obviously, the length of this field is only a few bytes, so that it does not impose a significant overhead. For computing $\overline{T_{k,i}}$ and $\overline{T_{j,i}}$, $P_{k,i}$ and $P_{j,i}$ can be evaluated as described in [15], where a STA can use the perceived SNR and the AP can use the uplink SNRs for the stations (as we assume similar wireless channel conditions in both uplink and downlink directions).

3.4 *Dynamic AP Selection*

As the wireless environment is dynamically changing, after a period of time any selected AP may no longer remain the best one. Therefore, a STA should be required to evaluate the function $W(i)$ after some time period T_c and re-associate if it found a better AP. Unlike previous solutions in which the time period T_c is constant and STAs always scan all supported channels [12] we propose an enhancement. Specifically, the value of T_c is dynamically adjusted in order to avoid unnecessary scanning and a frequent “ping-pong” effect. Moreover, a STA scans all channels only after it has been powered up. Afterwards, a mask is used to filter out all channels from which a Beacon or Probe Response frame have not been received. However, the set of all non-overlapping channels (channels 1,6, and 11 in IEEE 802.11b) are not masked since those channels are most likely to be used by APs. This considerably reduces both the number of scans and the scanning time, which is obviously very critical for delay sensitive applications. For example, considering voice over IP (VoIP traffic), which has very strict requirements, the maximum tolerable end-to-end networking delay is about 150 ms. The typical scanning and re-association time with the legacy 802.11 approach is between 1 and 2 seconds, which will hardly allow a good VoIP quality. This time is reduced by using the combination of dynamic selection and masking. However, this reduction is still not satisfactory.

The scanning time span can be mitigated further by “spreading” the scanning process over time. Specifically, once the STA has a list of neighboring APs, it can sequentially send probe request frames to the APs. The time separation between two successive probe request frames T_w could be selected such that the quality of service of ongoing calls is not degraded significantly. After receiving probe responses from all neighboring APs, the STA has the required information to decide whether to join a new AP or not.

Apart from that, the emerging 802.11k and the 802.11r standards as well as the 802.11 power-save mechanism can further contribute in resolving the scanning latency problem. 802.11k [18] enables a STA to request a list of candidate neighboring APs from the AP to which the STA is associated. 802.11r [19] defines mechanisms for fast and secure transition between APs. Dependent on the duration of the scanning latency, a station will experience packet losses since it is not able to receive frames being sent by its old AP. With the power-save mechanism of legacy 802.11, one is able to minimize packet losses while the station scans for APs on other channels. During this period, the AP buffers frames dedicated to the scanning station such that they can be conveyed after STA’s reappearance. Huang et al. [16] used this scheme to optimize selective scanning prior to fast WLAN handoffs.

Therefore, by combining the spreading scanning concept with the 802.11 power-save mechanism as well as with IEEE802.11k and IEEE802.11r technologies, we expect dynamic selection mechanism to be possible with VoIP without harming ongoing calls.

We summarize our Dynamic AP Selection algorithm as follows:

Algorithm 1 Dynamic AP Selection Algorithm

1: ChannelList $\Leftarrow$ {All Supported Channels}
2: Non-Overlapped $\Leftarrow$ {Non-overlapping Channels}
3: Send Probe Request Frames or Listen to Beacons over ChannelList
4: $AP_{new} \Leftarrow$ Select AP based on $W(i)$ (as explained in Section 3).
5: **for** $(time = 0$ to $T_c)$ **do**
6: do normal communication
7: **end for**
8: Send Probe Request Frames or Listen to Beacons over ChannelList.
Alternatively:
9: **for all** Channels $\in$ ChannelList **do**
10: {
11: Send Probe Request over a Channel;
12: do normal communication for Some Time T_w;
13: }
14: **end for**
15: ChannelList $\Leftarrow$ Non-Overlapped $\bigcup$ {Channels over which Probe Responses or Beacons were received}
16: $AP_{new} \Leftarrow$ Select AP based on $W(i)$
17: **if** $(AP_{new}$ better than current $AP)$ **then**
18: $T_c \Leftarrow T_c/2$
19: Re-associate
20: **else**
21: $T_c \Leftarrow 2 \cdot T_c$
22: **end if**
23: go to step 5

4 Performance evaluation

This section compares the performance of the proposed static and dynamic selection mechanisms with the default one implemented in IEEE 802.11b WLAN cards, in which STA-AP selection is based on the Strongest Received Signal. The performance results are obtained by means of simulation using the NCTUns [20] simulator.

4.1 *Simulation scenario and metrics*

We consider a large area like a departure hall in an airport as basic setting. Due to a potentially high number of users, four 802.11b APs that operate on different IEEE 802.11b channels are placed within this area. Users appear in this hall at different points in time and at different places. Two different user types may be present: either FTP or VoIP clients. Both have nomadic mobility degree, i.e., users start their devices and stay at a constant position during their active session.

Depending on the distance between AP and STA the wireless channel is attenuated more ore less severely. However, we assume that radio signals are not only attenuated by path loss, but are also affected by fading due to multi-path propagation. In order to accurately model these effects, a path loss component as well as a Rayleigh-distributed fading component is considered. For the path loss we have used a two-ray ground reflection model with the received power P_{rx} given as:

$$P_{\text{rx}} = \frac{P_{\text{tx}} G_{\text{tx}} G_{\text{rx}} h_{\text{tx}} h_{\text{rx}}}{d^2} \tag{13}$$

where P_{tx} is the transmit power (in mW), G_{tx},G_{rx} denote the transmitter and receiver antenna gains respectively, h_{tx} and h_{rx} are the antenna heights of transmitter and receiver, and d is the distance between them. We assume that APs and STAs use the same transmission power level. The fading impact is regenerated for every frame transmission per station. An internal fading process of the NCTUns simulation tool is used which takes as parameter a variance coefficient. Stations choose their transmission rates depending on the perceived, average SNR (i.e. without the fading impact) and try to assure a bit error rate (BER) less than 10^{-5}. This rate remains constant during the simulation, i.e, no rate adaptation mechanism has been implemented. For IEEE 802.11b the possible rates actually are 1 Mbit/s, 2 Mbit/s, 5.5 Mbit/s and 11 Mbit/s.

As the throughput of the whole system should be maximized, every AP measures the throughput every second in up- as well as downlink directions. The sum of the four AP throughputs is the first metric, which is denoted as *aggregated throughput* in the following. Despite packet loss, the end-to-end delay is the most critical component for VoIP. In simulations investigating VoIP, the *average round-trip time* is measured additionally. The value averaged over all STAs is a second metric.

4.2 *Parameterization*

STAs are uniformly distributed in a square area with an edge length of 500 meters while their arrival time is uniformly distributed over 40 seconds. APs are placed 250m apart from each other. In all simulations, the simulation time was set to 350 seconds.

We consider two different traffic mixtures: First, we investigate a setting with 60 FTP users, while in a second setting 30 VoIP users are present within the

WLAN area. FTP as well as VoIP sessions terminate at the wired part of the network at a single server. The latency for packets between APs and the server was set to 10μs. The cables connecting the APs to the server (via an 802.3 switch) have a 100 Mbit/s bandwidth. For each user traffic setting, simulations have been carried out for 5 different cases of random STA placements. Each case has been simulated 15 times where the appearance of users was randomly generated.

In the case of the FTP setting, every user downloads a file, whereby its size is indefinitely large. All TCP users utilize greedy TCP with packet length of 1000 bytes. The TCP traffic is generated with Jugi's Traffic Generator (jtg) [21]. In the case of the VoIP setting, each VoIP call is modeled by a bi-directional, isochronous audio flow. With ITU-T's G729 codec (which is widely used in 802.11 devices) and an audio frame length of 10ms, this results in an audio packet size of 10 bytes. The VoIP traffic was generated using the RTP/UDP traffic generator which comes with NCTUns simulator [20].

Table 1 lists other values of the parameters used in the simulations.

Table 1. Constant parameters

Parameter	**Value**	**Parameter**	**Value**
PLCP header T_{H}	48 μs	T_{SIFS}	10 μs
PLCP preamble T_{P}	144 μs	T_{DIFS}	50 μs
Cell overlap	20 %	T_{Slot}	20 μs
Fading Variance	10 dB	T_{CWmin}	31
APs/STAs Tx Power	100 mW	T_{CWmax}	1023
APs/STAs Tx Range	300 m	G_{tx} , G_{rx}	0 dBi
h_{tx} and h_{rx}	1 m	T_c	20 s

4.3 *Results*

In this section we present the simulation results of the proposed AP selection mechanisms. First, we discuss the results of the FTP setting and of the VoIP setting. Then, we present some deeper analysis of certain aspects of the new selection schemes. Note that parts of these results have already been presented in [22].

Let us initially consider the case of the FTP setting. We compare the aggregate throughput of the static and dynamic selection mechanisms discussed in this paper versus the legacy selection mechanism. The two graphs in Figure 2 presents the minimum and maximum aggregate throughput improvement (among the 5 different placement instances) obtained for the FTP setting for the static selection scheme. We observe that the aggregate throughput obtained from the proposed mechanism is almost always improved, in the best case this improvement is about 14.7%.

Next, consider the minimum and maximum improvements in case of the dynamic selection scheme for the FTP setting in Figure 3. Notice that for the dynamic selection scheme a significantly higher throughput gain (up to 33%)

is achieved than in the case of the static selection scheme. Hence, periodically reevaluating the AP association decisions appears to be quite beneficial for the performance of the considered network. Table 2 shows minimum, maximum and average throughput of both new selection mechanisms compared to the legacy mechanism for all five STA placements (notice that all presented results for a single placement are already the average over the 15 different instances of station appearance times).

Next, we consider the results regarding the VoIP setting. The two graphs in Figure 4 present the minimum and maximum improvements with respect to aggregate throughput for the static selection scheme. Clearly, the aggregate throughput is improved, however, the maximum improvement is only about 10%, which is in contrast to the FTP setting. Despite this small improvement in aggregate throughput, notice the significant improvement of the average round trip times. In Table 3 the average round trip times (in milliseconds) are given for all five station placements regarding the static and legacy selection scheme. We observe that if the round trip times of the legacy scheme become quite high (above 100 ms), the static selection scheme improves the round trip times substantially, since large round trip times are an indication for a highly loaded network. Thus, the lower the round trip times are for the legacy scheme, the smaller becomes the space for potential improvements with the static selection.

An important question relates to the impact of the dynamic selection scheme on the performance of a station during its scanning process. Graph (a) of Figure 5 presents the effect of periodic selective scanning (where only the three non-overlapping channels are considered) compared to default scanning (where all channels are considered) for a FTP client. The graph shows clearly that the throughput is less interrupted for the selective scanning process as only a few channels are considered. This leads to a greater minimum throughput while the recovery time is also shorter when selective scanning is used.

In the above discussion, the dynamic selection scheme has not been evaluated in combination with the VoIP setting. However, in Section 3.4 we argued that by spreading the scanning process in time (every once in a while a certain channel is scanned while the association to some AP is maintained), it might be feasible to apply dynamic selection even for VoIP scenarios. Graph (b) of Figure 5 shows the feasibility of this spreading solution. The throughput of a VoIP session is almost not effected by the modified dynamic selection scheme in comparison to the dynamic selection scheme without spreading. This result shows that the disruption time can be reduced by the spreading solution in comparison with successive scanning.

As especially the dynamic selection scheme seems to provide a significant performance improvement, an obvious issue to consider is stability. In order to capture the stability of the dynamic selection mechanism, we investigated the percentage of dynamic selection decisions in which a STA was required handoff from one AP to a new one. The results are given in Figure 6. They show that only about 15% of the re-evaluations require a STA to handoff from one AP to a new one. In order to further investigate the stability characteristics, we also considered the STAs's total distribution among the four APs during one sample simulation run for the dynamic selection scheme. These results are plotted in Figure 7, showing that with the dynamic selection scheme the total number of stations associated to each AP stays roughly constant after some initialization phase of the system.

Furthermore, we have been interested in the behavior of the static selection scheme in comparison to the legacy selection if the network density varies (either gets lighter or denser). In order to do so, we keep the number of stations in the simulation fixed and vary the distance between the AP's. Then, we collect the association decision by both approaches. Graph (a) of Figure 8 shows the percentage of equal association decisions between the two approaches for several different densities. The figure shows clearly, that the lighter the network density gets, the less often do the approaches perform equal association decisions. This is due to a much higher number of low-rate stations in the area, as the distance between APs increases. As our decision metric not only considers the throughput achievable per station but also the impact on other stations, the association decisions become less similar to the legacy decisions. Graph (b) of Figure 8 shows the comparison of the corresponding throughputs showing that the aggregate throughput was always higher with the static selection approach regardless of the distance between APs. It can be expected that the dynamic selection scheme would further improve this throughput advantage.

While a saturated network has been assumed in the presented analysis only, additional simulations have been conducted also for a non-saturated network with Poisson traffic in order to determine the gain in the throughput performance for a more realistic setting. Figure 9 plots the gain of the static selection policy over the legacy selection scheme for different packet inter-arrival times. Again, the results show that the static policy outperforms the legacy strategy even in a low loaded network while the gain increases as the load increases.

In order to capture the impact of the "weight" parameter α (which balances the throughput performance of each station to the impact to other stations in Equation 12) on the overall performance of the proposed selection policy, we evaluated the throughput performance of the static selection policy applying different values of α ranging from 0 to 1. Figure 10 provides the results of the simulation experiments carried out. The figure shows that a maximum throughput has been obtained with $\alpha = 0.4$. However, the fluctuations in the curve of figure 10 indicate that a precise estimate of an optimal value of α would require further extensive simulations which are left for future work.

Finally, we have been interested in comparing the performance of our selection policy with the ones proposed in [11] and [12] by Ekici et al. and Fukuda et al., respectively. We have included both policies in the simulator and conducted simulations with our scenario and traffic patterns as described above. In our multi rate environment, the policy of [12] does not improve the throughput overall, while the strategy of [11], however, has improved the throughput by 6.46%. The differences between our values and the improvements published in [11] are simply due to a different scenario choice. With a setting in which most of the STAs are originally positioned close to some APs and few STAs were distributed around the other APs, Ekici et al. were able to gain improvements of up to 50 percent. Moreover, our dynamic selection policy is more stable than the dynamic one of [12]. This is due to the dynamic setting of the periodicity of the re-evaluation of the best AP. Furthermore, with selective scanning our dynamic selection policy reduces the scanning time for the policy of [12] by almost 50%.

Table 2. Minimum, maximum, and average APs throughput (FTP scenario)

Min., Max., and Avg. Throughput (KB/s)				
Case	**Scheme**	**Min.**	**Max.**	**AVG.**
1	Legacy	821.79	897.82	855.81
1	Static	912.72	968.77	936.74
1	Dynamic	889.14	1075.46	992.73
2	Legacy	813.17	893.56	856.18
2	Static	952.79	1022.81	982.8
2	Dynamic	1044.45	1225.2	1141.24
3	Legacy	793.46	848.92	819.81
3	Static	839.45	908.3	871.1
3	Dynamic	839.66	1054.22	976.81
4	Legacy	858.63	955.81	910.92
4	Static	983.54	1063.06	1021.86
4	Dynamic	1076.25	1187.94	1139.11
5	Legacy	833.6	903.47	866.24
5	Static	927.13	982.33	949.52
5	Dynamic	932.45	1074.82	1011.88

Table 3. Average round trip delay of RTP packets (VoIP scenario)

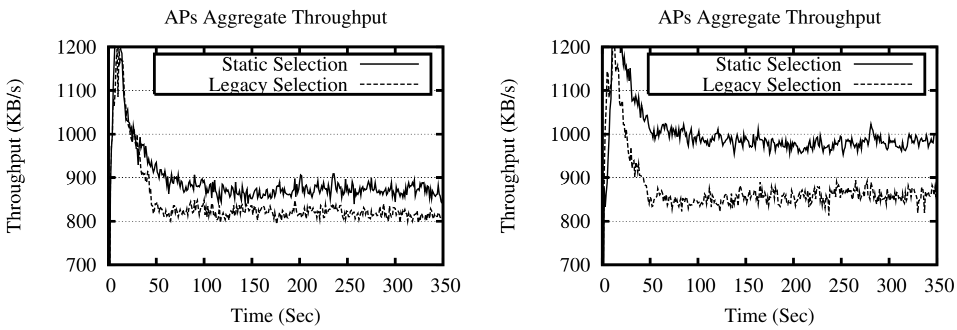
Average Round Trip Delay of RTP packets (ms)		
Case	**Legacy**	**Static**
1	68.7	62
2	101	64.3
3	112	84.3
4	151.46	81.3
5	53	49.5

5 Conclusions

The simple selection mechanism implemented currently in IEEE 802.11 WLANs does not effectively utilize WLAN resources as it bases the selection decision on RSSI values only and ignores other important parameters. In addition, as several IEEE 802.11 physical layer variants support multiple transmission rates, the currently implemented selection mechanism of IEEE 802.11 WLANs also does not consider the impact of association on the effective throughput of other (already associated) stations.

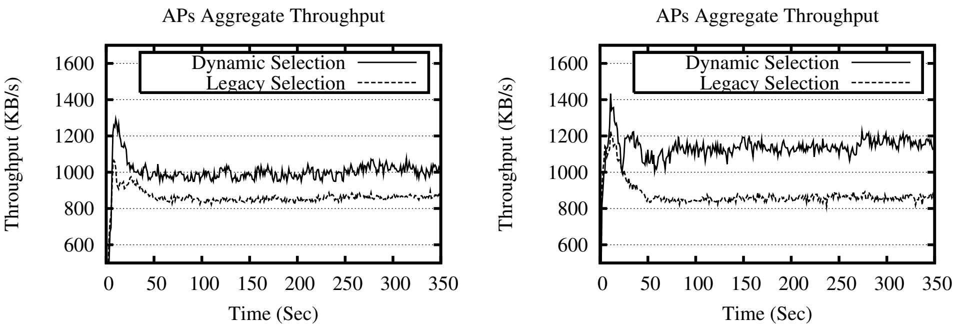
In this paper we propose to take both issues into account when selecting an AP: the "own" effective throughput as well as the impact on other (already associated) stations. We present a new AP Selection policy to mitigate this problem where the selection metric encapsulates several cell and connection parameters into a single value. Basically the mechanism tries to maximize STA's throughput as well as minimize its negative effect on high rate STAs currently accommodated by the AP to which it wishes to associate. We propose two selection schemes based on this metric: a static one where stations only consider their association once as well as a dynamic scheme where associations are reevaluated from time to time. Simulation results show that the proposed mechanisms can utilize WLAN resources much better and enhances users QoS (by improving aggregate network throughput and reducing the round trip delay). Especially the dynamic selection scheme provides a significantly performance improvement, while we have discussed several ways to mitigate problems related to channel scanning latency with the dynamic scheme.

Regarding future work, we expect further improvements when an optimal value of the weighing coefficient α in the cost function (equation 12) is used. It is our future goal to investigate this issue. Also, we are interested in the "globally" optimum performance of a WLAN network if all associations can be evaluated at a central controller in the network.



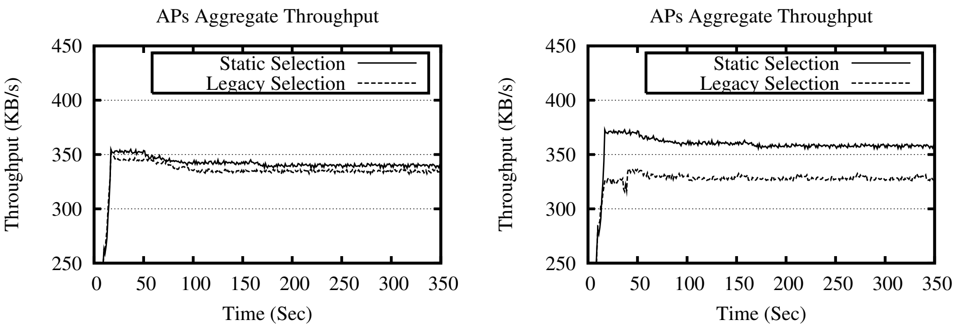
(a) Minimum improvement (b) Maximum improvement

Figure 2. Static selection- FTP scenario



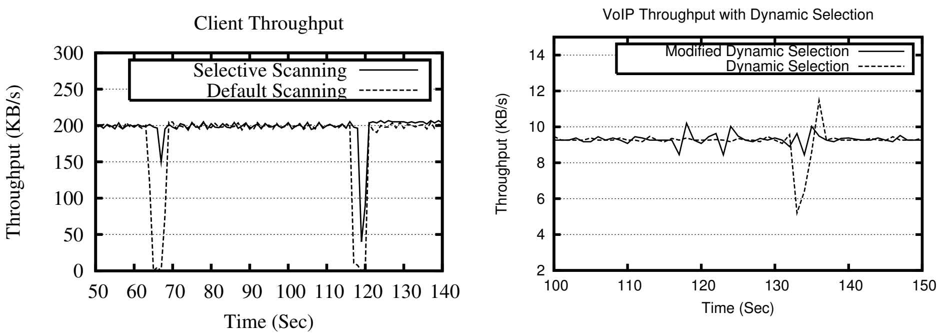
(a) Minimum improvement (b) Maximum improvement

Figure 3. Dynamic selection- FTP scenario



(a) Minimum improvement (b) Maximum improvement

Figure 4. Static selection- VoIP scenario



(a) Comparison between selective and default scanning

(b) VoIP with dynamic AP selection criterion

Figure 5. Impact of scanning on STA's throughput performance

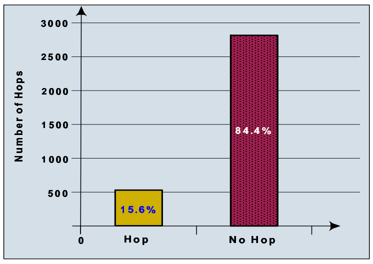


Figure 6. Hopping statistics for dynamic channel scanning

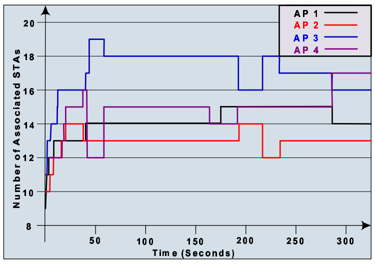
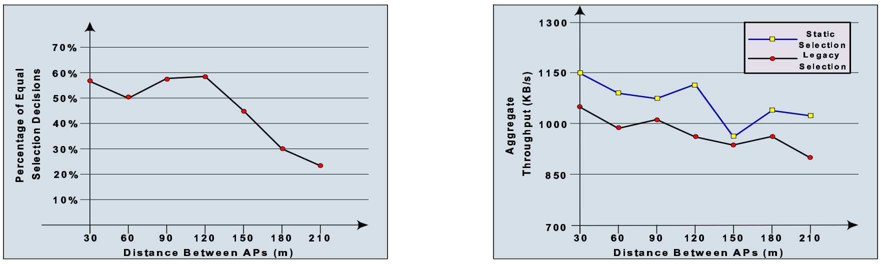


Figure 7. STAs's distribution among the APs



(a) Percentage of same selection decisions

(b) Performance of legacy and static selection

Figure 8. Legacy and static selection in dense and light WLANs

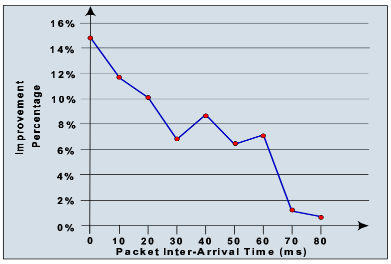


Figure 9. Static selection performance with non-saturated traffic

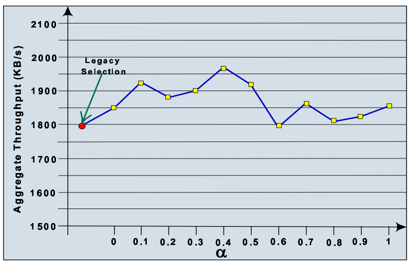


Figure 10. Impact of α on the performance of the static selection policy

References

[1] IEEE Std. 802.11-1999, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, Reference number ISO/IEC 8802-11:1999(E), IEEE Std. 802.11, 1999 edition.

[2] Wi-Pie in the sky?, The Economist, March, 2006.

[3] W. Arbaugh, A. Mishra, and M. Shin, An Empirical Analysis of the IEEE 802.11 MAC Layer Handoff Process, ACM SIGCOMM Computer Communication Review, March, 2003.

[4] Y. Bejerano and R. Bhatia, MIFI: A Framework for Fairness and QoS Assurance in Current IEEE 802.11 Networks with Multiple Access Points, in Proc. of IEEE Infocom, March 2004.

[5] G. Bianchi and I. Tinnirello, Channel-dependent Load Balancing in Wireless Packet Networks, Wireless Communications and Mobile Computing, Number 4, 43-53, 2004.

[6] A. Sang, X. Wang, M. Madihian, and R. Gitlin, Coordinated Load Balancing Handoff/Cell-Cite Selection and Scheduling in Multi-Cell Packet Data Systems, in Proc. of ACM/IEEE Mobicom, September 2004.

[7] G. Holland, N. Vaidya, and P. Bahl, A Rate-Adaptive MAC Protocol for Multi-Hop Wireless Networks, in Proc. of ACM/IEEE Mobicom, July 2001.

[8] M. Heusse, F. Rousseau, G. Berge-Dabbatel, and A. Duda, Performance Anomaly of 802.11b, in Proc. of IEEE Infocom, March 2003.

[9] A. Kumar and V. Kumar, Optimal Association of Stations and APs in an IEEE 802.11 WLAN, in Proc. of Nat. Conf. on Communications (NCC), February 2005.

[10] F. Guo and T. Chiueh, Scalable and Robust WLAN Connectivity Using Access Point Array, in Proc. of Inter. Conf. on Dependable Systems and Networks (DNS'05), June 2005.

[11] O. Ekici and A. Yongacoglu, Predictive Association Algorithm for IEEE 802.11 WLAN, in Proc. of IEEE Inter. Conference on Information and Communication Technologies (ICTTA), April 2006.

[12] Y. Fukuda and Y. Oie, Decentralized Access Point Architecture for Wireless LANs: Deployability and Robustness, in Proc. of IEEE Vehicular Technology Conference, September 2004.

[13] L. Gavrilovska and V. Atanasovski, Influence of Packet Length on IEEE 802.11b Throughput Performance in Noisy Channels, in Proc. of 1st Inter. MAGNET Workshop, November 2004.

[14] Giuseppe Bianchi, Performance Analysis of the IEEE 802.11 Distributed Coordination Function, IEEE Journal on Selected Areas in Communications, vol. 18, No. 3, March 2000.

[15] D. Qiao, S. Choi, A. Soomro, and K. Shin, Energy-Efficient PCF Operation of IEEE802.11a Wireless LAN, in Proc. of IEEE Infocom, June 2002.

[16] P. Huang, Y. Tseng, and K. Tsai, A Fast Handoff Mechanism for IEEE 802.11 and IAPP Networks, in Proc. of IEEE Vehicular Technology Conference, May 2006.

[17] S. Wiethoelter,Virtual Utilization and VoIP Capacity of WLANs Supporting a Mix of Data Rates, Technical Report TKN-05-004, Telecommunication Networks Group, Technische Universität Berlin, September 2005.

[18] IEEE 802.11 TGk, Radio Resource Measurement, March 2006, P802.11k/D4.0.

[19] IEEE 802.11 TGr, Fast BSS Transition, September 2006, P802.11r/D3.0.

[20] S. Y. Wang et al., NCTUns 3.0 Network Simulator and Emulator, http://nsl.csie.nctu.edu.tw/nctuns.html.

[21] J. Manner, Jugi's Traffic Generator (jtg), http://www.cs.helsinki.fi/u/jmanner/software/.

[22] M.Abusubaih, J. Gross, S. Wiethoelter, and A. Wolisz, On Access Point Selection in IEEE 802.11 Wireless Local Area Networks, in Proc. of the Sixth Inter. Workshop on Wireless Local Networks (WLN 2006), November 2006.