MiFi: A Framework for Fairness and QoS Assurance in Current IEEE 802.11 Networks with Multiple Access Points

Yigal Bejerano and **Randeep S. Bhatia***

Abstract—**In this paper we present a framework for providing fair service and supporting QoS requirements in IEEE 802.11 networks with multiple access-points (APs). These issues becomes critical as IEEE 802.11 wireless LAN are widely deployed in nationwide networks, linking tens of thousands of "hot-spots" for providing both real-time (voice) and non real-time (data) services to a large population of mobile users. However, both fairness and QoS guarantees cannot be supported in the current 802.11 standard.**

Our system, termed MiFi, relies on centralized coordination of the APs. During any given time of the "contention-free" period only a set of non-interfering APs is activated while the others are silenced. Moreover the amount of service granted to an AP is proportional to its load and the system's performance is optimized by employing efficient scheduling algorithms. We show that such a system can be implemented without requiring any modification of the underlying MAC protocol standard or the behavior of the mobile stations and it guarantees to overcome the hidden node and the overlapping cell problems. Our simulations establish that the system supports fairness and hence can provide QoS guarantees for real-time traffic, while maintaining a relative high throughput.

Keywords—**IEEE 802.11, QoS, fairness, Wireless LAN, Approximation Algorithms.**

I. INTRODUCTION

In the recent years we are witnessing a tremendous gain in popularity of IEEE 802.11b wireless LAN (WLAN) based networks, known by the name of WiFi. WLANs are getting rapidly deployed all over the world. In the U.S. alone several companies, including Boingo, Cometa and T-mobile have declared their intentions to build and support nationwide WLAN networks, with tens of thousands of hot spots for providing broadband IP connectivity to the traveling professionals. Recently, Cisco introduced its WiFi phone aimed at providing wireless phone service at nominal charge in these hot spots. However, to make these efforts a success, many serious shortcomings of WiFi must be overcome. Some of these include the lack of fairness of the WiFi technology and its inability to provide any kind of quality of service (QoS) guarantees making it incapable of supporting real-time (RT) services such as voice and video conferencing.

Generally speaking, *fairness* is the ability of a network to provide the same level of service to all its users, while *Quality of Service* (QoS) is the ability of providing a service with some level of assurance for data delivery. This assurance is usually given in terms of guaranteed bandwidth, delay bounds and jitter, which are essential for RT applications like voice. Providing QoS assurance and fairness are two related problems, and a system that cannot provide a certain degree of fair service to its users, *i.e.,* minimal allocated bandwidth, cannot provide QoS guarantees. However, both fairness and QoS guarantees cannot be supported in the current 802.11 standard.

The IEEE 802.11 MAC standard defines two operations modes. The first is a "best effort" Distributed Coordination Function (DCF) that employs a Carrier Sense Multiple Access/ Collision Avoidance (CSMA/CD) scheme for channel access. In this mode the users compete for their opportunity to transmit data in a contention period. The DCF mode is known to exhibit both short and long term unfairness [1],[2]. That is the MAC layer in DCF mode may fail to provide equitable allocation of channel resources to competing stations, which has detrimental effect on RT applications [1]. Numerous schemes have been proposed in the literature to overcome these shortcomings of the DCF mode (see survey papers [3], [4]). Most of these schemes propose one or a combination of the following techniques; modifying the DCF backoff mechanism [5], [2], employing adaptive size for contention window [6], using different interframe spacing (IFS) [7] or various frame sizes [8]. Other papers [9],[10] propose distributed schemes for obtaining "soft" QoS guarantees. Generally, these papers have shown that the distributed schemes are capable of providing fair service and supporting real-time traffic when the network load is low and when all the stations are in the transmission range of each other. However, the distributed schemes can provide only "soft" QoS assurance without strict bounds on the message delay and jitter [9]. In addition, their efficiency is reduced at heavy load and when high data rate is used [11], which motivates the usage of centralized access methods.

The second operation mode is the Point Coordination Function (PCF) that was designed to support real-time (RT) traffic. Here, a network access-point (AP) periodically initiates collision free periods (CFPs) in which it polls its associated stations. Unlike the DCF mode, the PCF is a centralized access mode and several recent papers [11], [12], have shown that it is suitable for supporting real-time traffic. The "connection-oriented" behavior of the PCF mode allows the network to provide bandwidth and delay guarantees that are necessary to support real-

---

* The authors are with Bell Laboratories, Lucent Technologies.

time applications. These papers consider only networks with a single access point and assume that all the user are in the transmission range of the access point. However, this property is not sustained in networks with multiple access-points that have overlapping transmission ranges or *cells*. In such networks transmission collisions may occur during a CFP as a result of either "hidden nodes" that have not received the beacon messages announcing the beginning of a CFP, the so-called *hidden node* problem, or when two adjacent access points schedule their CFPs simultaneously, known as the *overlapping cell* problem [13]. Thus, a station may fail to send or receive data when polled during the CFP as a result of interferences from other cells. Recently, some work has addressed the hidden node and the overlapping cell problems However these schemes cannot ensure either fair service or QoS guarantees. These problems are still unsolved also in this new QoS extension of the standard, the so called IEEE 802.11-e proposal [3].

In this paper we present the Managed WiFi system, called *MiFi*, for supporting fairness and QoS in the existing IEEE 802.11 MAC layer. To the best of our knowledge this is the first comprehensive system that overcomes both the hidden node and the overlapping cell problems in multiple-AP WLAN networks. Our scheme ensures a fair service for the mobile users and provides the required QoS guarantees for RT applications. It is inspired by the recent results of [11], [12] that have shown that the PCF mode can efficiently by used for supporting RT sessions and for providing fairness in WLAN when all the users are in the transmission range of a *single* AP. In this work, we extend their result also to networks with *multiple* APs. Our system uses a combination of MAC layer techniques and centralized management and coordination of the Access Points (APs). All the access points simultaneously toggle between a *Contention Free Period* (CFP) using PCF mode and a *Contention Period* (CP) using DCF mode. In the CFP, time is divided into equal sized slots such that within each slot only a subset of Access Points are activated, while the rest are silenced. By ensuring that the APs activated in any slot of the CFP are "non-interfering", our scheme guarantees that the system is free from the overlapping cell and the hidden node problems. Thus, in every slot of the CFP the system imitates the behavior of a single AP([11], [12]), and therefore guarantees fairness and QoS. For determining the slot assignment, we designed an efficient scheduling algorithm with proven guarantees on its performance. The algorithm ensures that the number of slots allocated to each access point is proportional to its load and moreover it maximizes each AP's share of time when it is activated. In the CFP the system ensures fairness and QoS to both RT and non RT sessions by providing service to backlogged sessions. The CP is also used for data transmission and as a signaling channel for management messages and initiating new sessions. By regulating the proportion of time spent in the CFP versus the CP the system is able to tradeoff the achieved fairness with the overall network throughput. Our scheme requires modifications only at the access-points and it can be implemented without requiring any changes to 802.11 MAC layer standard. We do not require any additional software or any changes in the mobile stations. Moreover, the scheme can be viewed as a complementary mechanism for enhancing the 802.11-E proposal. By extensive simulations we show that the system indeed provides fair service to its users and can support RT-sessions even in large networks, while the network throughput is at least half of the throughput of the regular WiFi networks.

II. THE NETWORK AND THE CHANNEL MODELS

A. The Network Model

We consider a standard 802.11 network operating in the infrastructure mode and supporting a large number of access-points (APs). All the APs are attached to a single distribution system (DS) that connects them via gateways to wired data and voice networks such as the Internet and the PSTN. The network provides connectivity services to the mobile users that reside in its coverage area, comprised of the APs transmission ranges. At every given time, each mobile user is associated with a single AP, and we denote by m_v the number of mobile user that are associated with AP v. We consider two types of mobile users, *real-time* (RT) and *non-real-time* (NRT) users. In the following we use the term *stations* for denoting both APs and mobile users.

For the sake of clarity of the presentation, we make several simplifying assumptions. These assumption mainly simplify the fairness definition and are not essential for the proper operation of the system. We assume that all the stations transmit with the same bit rate and all the packets has the same length. Thus, the transmission time of each message is one *time unit* (TU). We assume that NRT-users always have pending message for either sending or receiving while RT-users are occasionally engaged in a RT-session, *i.e.,* voice calls. Each RT-session requires a constant bit rate (CBR) and an RT-user needs to either receive or send a RT-message every D time units. For initiating a RT-session, a user needs to send a request to its associated AP, and the latter determines whether to accept or reject the request.

B. The Wireless Channel Model

We now turn to describe the main characteristics of the wireless channel. The WiFi standard allows us to use three non-interfering channels. Each AP and its associated mobile users utilize only a single channel and use it as a shared medium. As a result, simultaneous transmissions may interfere with each other. The ability of a station to decode a message depends only on the Signal-to-Interference Ratio (SIR) defined by the strength of the received signal over the accumulated strength of the other interfering transmissions. The received signal at a desti-

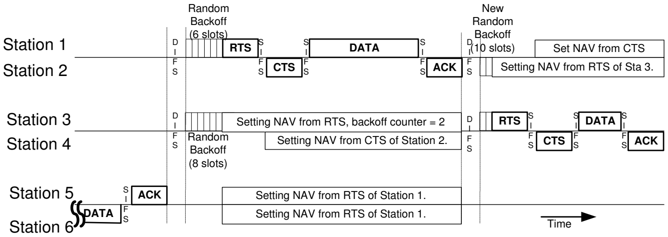

Fig. 1. An example of station transmissions in DCF mode.

nation station depends on the transmission power at the source station and the *gain* between them. The latter depends on the attenuation of the signal power with distance. We assume that the main factor that determines the gain is the path loss and we ignore secondary effects like multipath and shadowing. We model the path loss of two stations s and t with distance $d = d(s,t)$ from each other as $g_{s,t} = (d_0/d)^{\beta}$, for a given $\beta \in [2..4]$ and d_0, as described in [14]. For simplicity, we assume that the distance units are normalized such that $d_0 = 1$. Furthermore, we assume that all stations transmit with the same power, denoted by P_{max}. Therefore, the strength of a received signal at distance d from a sender is $P_r(d) = \frac{P_{max}}{d^{\beta}}$.

We denote by h_S the minimal signal strength needed for sensing a signal. Thus, anything below h_S is background noise. Let SIR_{min} denote the minimal SIR for decoding a message. Then, we must have $SIR_{min} = \frac{h_T}{h_S}$ where h_T is the minimum signal strength needed for decoding a received signal. Note that $h_T > h_S$. Thus, each station v can be associated with two circular regions around it. The first is its *transmission range* that defines the zone in which any message sent by station v can be correctly decoded and its radius is $R_T = \sqrt[\beta]{P_{max}/h_T}$. The second is its *sensing range* and any station included in this range can sense every transmission of v. The radius of the sensing range is $R_T = \sqrt[\beta]{P_{max}/h_S}$ and in practice $R_S > 2 \cdot R_T$.

III. THE IEEE 802.11 STANDARD

A. *The MAC layer of the IEEE 802.11 standard*

In this section we provide a brief description of the IEEE 802.11 MAC protocol [15] and its limitations regarding QoS support and fairness. For our need, we present only the *infrastructure operation mode* of the standard. In this mode, the network contains *access points* (APs), that relay messages between mobile stations and connect them to a wired backbone. The standard supports two coordination functions for accessing the media. The *Distributed Coordination Function* (DCF) provides distributed asynchronous access to the wireless channel based on a collision avoidance scheme. The *Point Coordination Function* (PCF) offers synchronous, collision free access to the media by using a centralized polling approach.

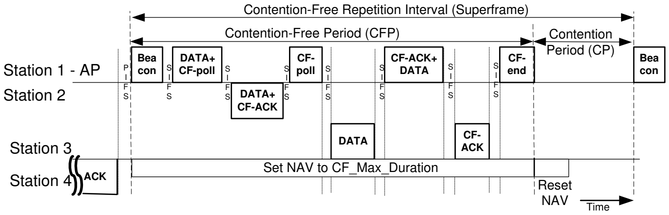

Fig. 2. An example of station transmissions in PCF mode.

A.1 The Distributed Coordination Function

The DCF is a distributed medium access method that employs the Carrier Sense Multiple Access/ Collision avoidance (CSMA/CA) mechanism. Each station senses the wireless channel before transmission by using both physical and virtual carrier sensing mechanisms, and it sends its message only if the channel is idle at least for a period of DCF Interframe Space (DIFS). At the end of each successful transmission, the receiving station acknowledges the reception by sending an ACK message after a Short Interframe Space (SIFS). The SIFS is much shorter then the DIFS for preventing other stations from initiating a new transmission between the data and the ACK messages. In this approach, collisions occurs as a result of several stations detecting idle channel at the same time and initiating simultaneous transmissions.

For reducing the interferences that result from the hidden node problem (see Section III-B for details), the standard defines an optional Request-to-Send/Clear-to-Send (RTS/CTS) handshake[1]. A station initiates its transmission by sending a short RTS message that contains the message destination and a *duration field*. The latter defines the required period for data transmission and the CTS and ACK replies. Upon reception of a RTS message, the destination replies after a SIFS with a CTS message that also carries a duration field. Each station that decodes the RTS or CTS message, sets an internal timer, termed the *network allocation vector* (NAV), for this period and defers its transmission appropriately. An example of a DCF message transmission is illustrated in Figure 1. Recall that a station may sense a busy channel but may not be able to decode the message as a result of a collision or low SIR. In such a case, the station defers its transmission for a period of at least Extended Interframe Space (EIFS) which is much longer than the DIFS period.

A.2 The Point Coordination Function

The Point Coordination Function (PCF) is used for supporting delay sensitive applications. In this mode, time is divided into repeated periods that are called *superframes*. Each superframe contains both a *Contention Free Period* (CFP) managed by the PCF and a *Contention Period* (CP) in which the DCF is used for accessing the channel. For starting a CFP, the AP transmits a beacon frame for all the

---

[1] In practice, most WiFi devices utilize the RTS/CTS-handshake.

stations in its transmission range to enter the PCF mode. During the CFP, the AP polls its associated stations according to a pre-determined order, called *polling-list*, usually based on round-robin scheduling. No station is allowed to transmit unless it is polled and it enables the AP to receive and deliver messages without interference. The CFP ends when the AP sends a CF-end message. An example of a superframe is depicted in Figure 2. Currently, the PCF mode is optional part of the standard, however, it is a mandatory mode in the IEEE 802.11-E QoS Extension [3]. In this work we assume that the PCF mode is supported by all stations.

B. The Fairness and QoS Limitations of the Standard

Generally speaking, fairness is the network ability to provide the same service level to all its users during a given time internal. This property is essential for providing Quality of Service (QoS) assurance for real-time applications, usually given in terms of guaranteed bandwidth, delay bound and jitter. However, these guarantees cannot be supported in the current 802.11 standard. The DCF is designed to support only best-effort services and the distributed MAC protocol cannot provide any guarantee on message delays. The PCF mode provides certain degree of fairness in the case of a single AP, however it cannot support the required guarantees in networks with several APs. In such networks, transmission collisions may occur during a CFP as a result of the hidden node and overlapping cell problems [13]. The *overlapping cell problem* refers to the issue of interference during transmissions in a CFP due to the transmissions in adjacent cells. and it is a special case of a more subtle *hidden node* problem. A station is called *hidden* when it is in the sensing range of the intended receiver but out of the sensing range of the transmitter. Thus, a transmission of the hidden station may prevent the receiver from decoding the intended message. The 802.11 standard addresses this problem by using both physical and virtual sensing mechanisms (the latter is supported by the RTS/CTS-handshake). However, these mechanisms provide only partial solution to these problems.

The hidden node problem causes a significant reduction in the system throughput and prevent the APs from providing fair service to their users even during their CFPs. In this case, appropriate synchronization of the CFPs is not sufficient, since the hidden node may be a mobile user operating in DCF mode that has not received the beacon message at the beginning of a CFP. Our simulations[2] indicate that the affect of hidden nodes is more severe on users near the boundaries of the AP transmission range, as illustrated in Figure 3. For such users the signal from their associated AP is very weak, while the aggregated noise from adjacent cells is elevated. Thus, making them highly prone to interference. Moreover, the backoff mechanism itself may have a detrimental effect on the service level obtained by these users, as the contention window is doubled after each failed transmission. Consequently, the service level that a mobile user experience depends on its distance from its associated AP. While closer users benefit from high throughput and low delays, remote users are actually starved.

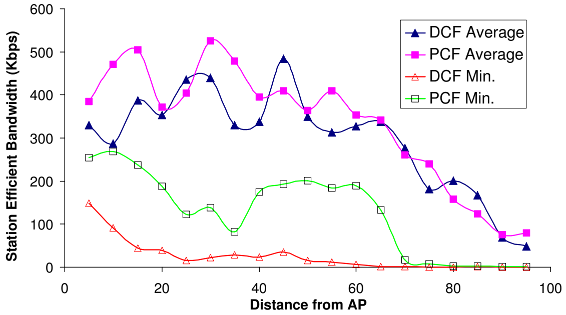

Fig. 3. The provided service level of the DCF and PCF modes with frequency planning, where the transmission range of an AP is 100.

IV. THE SYSTEM GOALS

In the following we use the following definition. We say the *efficient bandwidth* B_v of a station v, also termed the station *flow*, is the amount of data that station v transmits or receives successfully in a given time period. Thus, the efficient bandwidth of an AP is the aggregated flow that the AP successfully exchanges with its associated users and the networks throughput is the total flow of all its APs.

The main goal of our system is to provide fair service to all its mobile users and to ensure QoS guarantees to real-time sessions, while maximizing the achievable overall network throughput. Intuitively, a system provides a fair service if every user experiences the same network usage and has the same flow as any other user. This notion of fairness assumes that all users are identical and have the same requirements. However, in our system, RT users with very strict QoS requirements for latency and bandwidth co-exist with NRT users that would only like to maximize their average throughput. Thus, *our notion of fairness is to ensure that users of a single type experience the same network usage and the network resources are proportionally allocated among the users of the two types.* Moreover, the experienced service level should be independent of the distance between the users and their associated APs. In other words, we require *spatial fairness*. To achieve our fairness goals we need to consider both *inter-AP* and *intra-AP* fairness. An AP provides *intra-AP fairness* by balancing between resource allocation to its NRT and RT-users according to a given *fairness criteria*. For instance, a plausible fairness criteria may require that the success probability of a RT-session request be proportional to the efficient bandwidth given to each NRT-user. In addition, an *inter-AP* fairness is obtained when the efficient band-

---

[2] The details of our simulations can be found in Section VII.

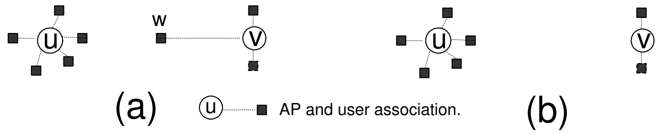

Fig. 4. The relationship between fairness and throughput.

width B_v of an AP v is directly proportional to the total number of users m_v associated with it. Thus, we say the network is *inter-AP fair* if the ratio B_v/m_v is the same for all APs. In wireless networks achieving intra-AP or inter-AP fairness is contradictory to achieving high throughput as we illustrate with the following examples.

Example 1: Consider two APs u, v, as depicted in Figure 4-(a). Here user w is associated with AP v. Due to the large distance between the two APs, both of them can simultaneously exchange messages with their adjacent users (except user w). However, when v is exchanging messages with user w then u and all its associated users must be silent. Consequently, the network throughput is maximized by starving user w. In other words, in order to maximize the network throughput the intra-AP fairness (spatial fairness) requirement must be violated. ■

Example 1 illustrates that inter-AP fairness requires coordinating the operations of adjacent APs, resulting in a reduction of the overall network throughput.

Example 2: In Figure 4-(b) the two APs do not interfere with each others communication and both have the same efficient bandwidth, (assuming $B_v = B_u = $ 1Mbps). In this scenario, inter-AP fairness requires that all the users in the system experience the same flow allocation, $1/5$, since AP u has 5 associated users. However, the network utilization can be increased by allowing increased flow allocation of $1/2$, to the users associated with AP v without affecting the flow allocations of the users attached to AP u. ■

We resolve such issues as illustrated in Example 2, by considering a *maxmin fairness* requirement that seeks to find a flow allocation that maximizes the ratio $\min B_v/m_v$ for all the APs v. After maximizing this ratio, our scheme uses a heuristic to maximize the system overall throughput by allocating the residual efficient bandwidth of the network to the different APs. Thus, the two examples illustrate the need for a system that would help the network administrator to strike a balance between fairness and throughput. We show later how MiFi achieves this goal.

V. AN OVERVIEW OF THE MIFI SYSTEM

In this section we present an overview of the MiFi system. The system contains a *network operation center* (NOC) that coordinates the APs. For management purpose, we require the internal clocks of the NOC and the APs to be synchronized for a certain degree of accuracy[3], and let assume that the clock gap between any pair of APs is negligible. On-board each AP, a special software is used to control its behavior for providing QoS and fairness to the attached users and for communicating with the NOC. However, the MiFi system does not require any modification of the IEEE 802.11 standard or the software of the mobile users. We only assume that mobile users are able to convey, via request messages, the type of session they wish to initiate, which can be either a RT or a NRT session.

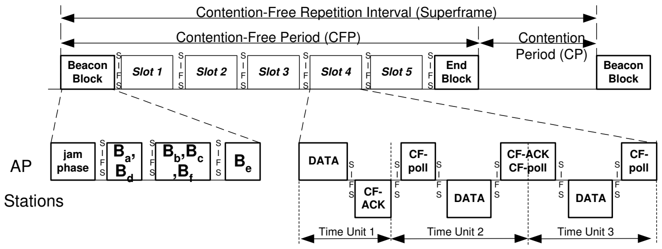

Fig. 5. A MiFi superframe with 5 slots, each 3 time units long.

A. General Description

Our scheme is inspired by the recent results of [11], and [12]. These papers have shown that the PCF mode can efficiently support real-time sessions and provide fairness in WLAN when all the users are in the transmission range of a *single* AP. In this work, we extend this result also to networks with *multiple* APs. This challenging goal cannot be obtained easily, mainly due to the overlapping cell and the hidden node problems. For achieving this goal, the MiFi system imitates the behavior of a single AP. We partition time into repeated periods or *superframes*. Each superframe has a fix length D and it contains a *Contention Free Period* (CFP) followed by a *Contention Period* (CP). The CFP is used for data transmission of both RT and NRT sessions, for obtaining inter-AP fairness and QoS support, while the CP is used for serving the AP nearby users and as a signaling channel for initiating new sessions and sending management messages. The proportion of time allocated to each period is determined by the system needs to balance between fairness and network throughput. The CFP starts with a *beacon block* (BB) in which all the APs transmit 'almost' at the same time beacon messages for initiating a CFP in their vicinity. It ends with an *end block* (EB) in which all the APs send CF-end messages approximately at the same time to end their CFPs. This is illustrated in Figure 5. During the CFP, the APs poll their associated users according to their *polling-lists*, and only stations that are polled are allowed to transmit. *By ensuring that the CFPs of all the APs start and end at the same time, our scheme solves the hidden node problem that results from spontaneous transmissions of mobile users operating in the DCF mode.*

Our scheme also overcomes the overlapping cell problem which is achieved as follows. We divide the CFP into slots, each of size at least Δ time units, for a specified

---

[3] Such synchronization can be obtain since all the APs are connected to the same local network.

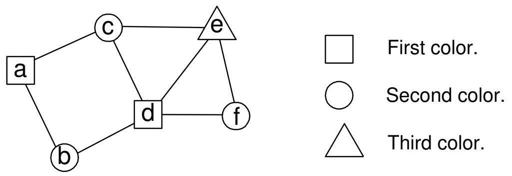

Fig. 6. The interference graph of a WLAN with 6 APs.

parameter Δ. The slots are efficiently allocated by the NOC to the APs such that no two APs whose transmissions may interfere get the same slot. Every AP is allowed to poll its users only during its allocated slots. In addition, it is ensured that transmissions in a slot terminate by the end of the slot, thus avoiding collisions with transmissions in the following slots. In our system the NOC only determines the slot assignments and synchronizes the APs. Each AP manages its own admission control mechanism for accepting new RT-sessions and determines its own order for polling its users. We now provide more details of our scheme.

B. The Interference Graph

We say that two AP are *interfering* when a message exchanged by one AP may prevent a proper message decoding in the vicinity of the other. The MiFi system overcomes the overlapping cell problem, by allocating disjoint sets of slots to interfering APs. From our channel model it follows that,

Property 1: The distance between any pair of interfering APs is at most $2 \cdot R_T + Rs$.

Proof: We make the simplifying assumption that interference results mainly from simultaneous transmissions from a single interfering station. Consider a station u with distance R_T from an AP i. Node u cannot decode a message from AP i when its SIR is less than SIR_{min}. This results from simultaneous transmission of another node w whose distance from u is at most R_S. Since w may be a user associated with another AP, it follows that two APs are at most $2 \cdot R_T + Rs$ apart. ■

Note that even if two APs are $2 \cdot R_T + Rs$ apart they many not be interfering. However we attempt to provide "hard" QoS guarantees by assuming that two APs are interfering if their distance is at most $2 \cdot R_T + Rs$. These interfering relationships are represented by an *interference graph* defined as follows,

Definition 1 (interference graph) : The *interference graph*, $G(V, E)$, is defined by the set V of APs and a set of edges E between every pair of APs $u, v \in V$ that are at most $2 \cdot R_T + Rs$ apart, *i.e.,* $d(u, v) \leq 2 \cdot R_T + Rs$.

Definition 1 implies that the interference graph is actually a unit disk graph [16]. An example of an interference graph $G(V, E)$ is depicted in Figure 6.

The achievable throughput of the MiFi system depends on the density of the interfering graph G. General speaking, the number of slots allocated to each AP are inverse proportional to its degree in G. In practice, we may eliminate from G the edges whose endpoints have low probability of interfering with each other, and still get adequate QoS guarantees. We leave it for a future work the problem of constructing an interfering graph that allows the MiFi system to provide sufficient QoS assurance while exploiting spatial reuse more efficiently.

C. The Beacon and the End Blocks

In the 802.11 standard each CFP starts with a beacon message and ends with a CF-end message. In our scheme we ensure that every mobile user successfully receives these messages from its associated AP by the usage of beacon-blocks and end-blocks. For simplicity, we describe only the beacon block (BB). The end block (EB) has similar description. Recall that a station in DCF mode initiates a transmission by sending a RTS packet with a duration field. A station, that senses a signal, which it is unable to decode, postpone its transmission for a period of EIFS. The BB contains two phases, a *jamming* phase followed by a *beacon transmission* phase. The jamming phase silences the network for a period of EIFS, so that no users are allowed to transmit. In the beacon transmission phase, APs send their beacon messages. Since the users are silent in the beacon transmission phase the beacon messages do not suffer from collisions with messages transmitted by mobile users operating in DCF mode.

Let us assume, for illustration, that the beacon block starts at time t_0. Any RTS messages that originate before time t_0 and whose data and ACK transmission are supposed to end after time t_0 are ignored by the APs. Thus at time t_0 there are no transmissions of data or ACK messages in the network and the only possible transmissions are for RTS messages. At time t_0, all the APs start to jam the channel, which is done for a period of $RTS\_TIME + DIFS$, where RTS_TIME is the time required for sending RTS message at the lowest bit rate. As a results, all the mobile users, including those that have sent RTS messages at time t_0, sense the jammed signal and set their NAV to EIFS. At the end of the jamming phase the APs send their beacon messages in the beacon transmission phase. Since beacon messages from two interfering APs may collide, the beacon transmissions of APs are synchronized such that two adjacent APs in the interference graph do not send their beacon messages simultaneously. For reducing the overhead of the beacon block, we would like to send the beacon messages as quickly as possible. Thus, we map the beacon synchronization problem into a graph coloring problem that seeks to find the minimal number of colors that are needed to color the interference graph, such that all the nodes with the same color send their beacon messages simultaneously, as described in Example 3. The details of the coloring algorithm are given in Section VI. Note that the beacon transmission phase may be required to be longer than an EIFS period. In this case, after several beacon transmissions, the APs simultaneously transmit a short jamming signal, for resetting the NAV of the mobile users to an additional EIFS period.

Example 3: An example of a beacon block is shown in Figure 5 for the interference graph $G(V, E)$ presented in Figure 6. Since, $G(V, E)$ is 3-colorable, the beacon block contains 3 *beacon slots*. In the first slot nodes a and d transmit their beacon messages, in the second slots a beacon is sent by nodes b, c and f, and finally in the third slot node e sends a beacon message. ■

D. The Slot Assignment Mechanism

As described above, all the APs share a common CFP that is partitioned into slots, each of size at least Δ time units. Slots are assigned to APs such that no two interfering APs get scheduled in the same slot. Thus, the goal of the *slot assignment mechanism* is to maximize the network throughput while ensuring inter-AP fairness.

Let the CFP be divided into R slots enumerated from 1 to R. We denote by S_v the set of slots that are allocated to AP v and by r_v the number of slots in S_v. A *slots assignment* is a vector $\mathcal{S} = \{S_{v_1}, S_{v_2}, \cdots, S_{v_{|V|}}\}$, of the sets S_{v_i} for every AP $v_i \in V$. A slots assignment is termed *feasible* if for every AP v, $S_v \subseteq [1..R]$ and any pair of adjacent nodes in the interference graph $G(V, E)$ do not have any common slot, *i.e.,* for every $(u, v) \in E$ it follows that $S_u \bigcap S_v = \emptyset$. For obtaining both inter-AP fairness and high throughput, we say that a feasible slot assignment $\mathcal{S}$ is *optimal* if it maximizes the *min-slot-to-user ratio* defined by $\rho = \min_{v \in V} \frac{r_v}{m_v}$. We show, in Section VI-C, that the problem of finding the optimal slot assignment is NP-hard and we present efficient approximation algorithms for the problem.

Our slot assignment scheme has two components. The first is a *coloring algorithm* that given a graph $G(V, E)$ and the number of colors, r_v, required by every node $v \in V$, finds a feasible color assignment with minimal (within a factor 3) number of colors. The second component uses the coloring scheme as a building block for finding an efficient slot assignment. It performs a binary search for finding the maximal min-slot-to-user ratio ρ that requires no more than R slots. At each iteration, it selects a ratio ρ and sets the requirement of every node $v \in V$ to $r_v = \lceil \rho \cdot m_v \rceil$ colors. The algorithm then uses the coloring algorithm to check whether or not there is a feasible slot assignment with R slots (colors). Based on the result, the algorithm picks lower or higher value for the ratio ρ until it quickly converges to the optimal ratio ρ. The details of slot assignment algorithm and its approximation analysis are given in Section VI-C.

Example 4: In Figure 7-(a) we present a slot assignment that maximizes the min-slot-to-user ratio of the given interfering graph $G(V, E)$. Each superframe contains 5 slots and the number of users m_v attached to each AP $v \in V$ is depicted near each node v. The figure shows the allocated slots S_v as well as the slot-to-user ratio $\frac{|S_v|}{m_v}$ of each node v. In this case the maximal min-slots-to-users ratio is $1/5$ due to nodes c and e. ■

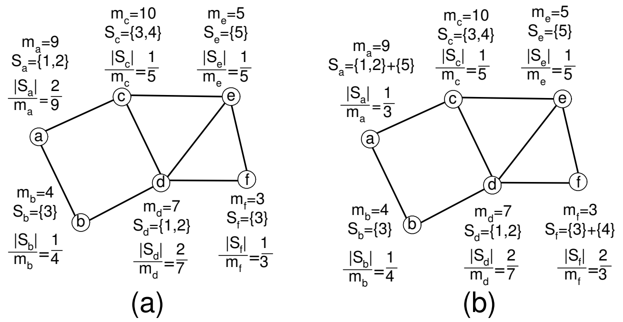

Fig. 7. An example slot assignment for maximizing the min-slots-to-users ratio.

Here, we consider further enhancements to the slot assignment mechanism for improving the system throughput. Consider a slot j and let V_j be the set of APs that are allowed to use slot j by the slot assignment algorithm. The network utilization is maximized during the period of slot j only if the set V_j is a *dominating set* of the interfering graph $G(V, E)$, *i.e.,* every node $v \in V$ is either included in V_j or is adjacent to a node in V_j. Note that if V_j is not a dominating set then there is an AP $u \notin V_j$ that is not adjacent to any other AP in V_j. Thus, node u may use slot j without interfering with the APs in V_j, resulting in an increase in the network utilization. Consequently, for maximizing the network throughput at each slot j without causing interferences, we seek a dominating independent set V_j' that includes all the nodes in the set V_j. We calculate the set V_j' by constructing a unit disk subgraph $\tilde{G}(\tilde{V}, \tilde{E})$ of graph G. The nodes $\tilde{V} \subset V$ are the ones that are not included in V_j and are not adjacent to any node in V_j in G. We use the algorithm in [17] for finding a maximal independent set of $\tilde{G}$, denoted by $\tilde{V}'$. The nodes that then acquire slot j are $V_j' = V_j \bigcup \tilde{V}'$. Recall that the set V_j' is a dominating independent set of G and therefore the throughput of the network is increased without violating the inter-AP fairness, as illustrated in Example 5

Example 5: In the slot assignment of Example 4, the sets V_4 and V_5 defined by the colors 4 and 5, respectively, are not dominating sets. These sets can be augmented with additional nodes while keeping the independent set property. After this augmentation, the new set V_4' also includes node f and node a is included in the set V_5'. Consequently, the network flow is increased by 22%, from 9 to 11 allocated slots per superframe, as depicted in Figure 7-(b). ■

The 802.11 standard allows us to use three non-overlapping frequencies F for reducing interference and thus increasing the network throughput. Every AP is associated with a frequency $f \in F$ and a slot is identified by a pair (f, c) of a frequency $f \in [1..|F|]$ and a color $c \in [1..R]$. Here, a slot assignment is feasible only if all

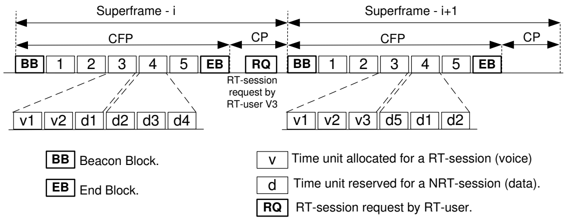

Fig. 8. An example of a polling list of node c.

the slots allocated to an AP belong to the same frequency but have different colors and any two adjacent APs do not share the same (f, c) pair. In Section VI-B we present an efficient slot assignment algorithm for the case when there are multiple non-overlapping frequencies.

E. The Admission Control and the Polling-List

In a MiFi system, each AP needs to maximize its efficient bandwidth while providing fair service to its RT and NRT users. For ensuring intra-AP fairness, each AP employs an *admission control* mechanism that enforces a given *fairness criteria*. For instance, consider an AP v that has r_v slots and is associated with m_v users, where m_v^{RT} of them are RT-users. Let Δ be the number of time units in every slots. Our admission control approves new RT-session request only while the aggregated flow allocated for RT-users does not exceed a threshold of $H_v = c \cdot \frac{m_v^{RT}}{m_v} \cdot r_v \cdot \Delta$, for a given configuration parameter $c \geq 1$ and a requirement that $H_v < r_v \cdot \Delta$. Such admission control balances between the success probability of RT-session requests versus the average flow given to each NRT-users. Other mechanisms can be found in [12], [11].

An RT-user initiates a new RT-session by sending a request to its AP during the CP. If the AP approves the request then it allocates a time unit to this user and adds the users address to its polling list. In the CFP the AP first polls all the RT-users with active RT-sessions and in the remaining time of its slots it polls its NRT-users. Since, H_v is smaller than the time units available to AP v, *i.e.,* $r_v \cdot \Delta$, every RT-user engaged in an active RT-session is either polled or receives data at each superframe. For ensuring intra-AP fairness, the AP employs a sliding window method for determining the next NRT-user to poll at time t. The AP keeps records of the number of successfully served messages (either sent or received) by each NRT-user during a time period of $[t - T, t]$, for a sliding window of size T. The polled NRT-user is the one which has the minimal number of served messages during that time period. Note that the intra-AP fairness is improved by increasing the size of the sliding window, T. An illustration of the polling mechanism is given in Figure 8.

The main concern about the efficiency of the polling mechanism is the number of unsuccessful polling attempts. In practice, NRT-stations do not always have packets to send and polling these stations decreases the system utilization. Our method for reducing this number is based on the following observations. Most of the traffic is from the APs to the mobile-users, *i.e.,* web browsing and email. Moreover, the majority of the packets are originated as a response to a message reception, like TCP acknowledgment. In our approach, the AP polls a user after sending it a packet and the users use the CP for sending or initiating sessions or for resuming their operations. The latter may be sent with high priority, as proposed by the IEEE 802.11-E proposal [3], for expediting the session initiations.

F. Mobility Management

Generally speaking user mobility does not raise any major concern for the system operation. Upon movement from one cell to another, each user scans all the 802.11 wireless channels for detecting beacon messages and it uses these messages to identify the APs in its vicinity. Then, the user sends a re-association request to one of the detected APs during its CP. Consequently, user mobility does not interfere to the CFPs of the MiFi system. However, mobility raises the following two challenges. First, preventing the disconnection of ongoing RT-sessions. This may happened when an RT-user moves to a cell of an AP v that is already supporting the maximal number of permitted RT-sessions, H_v. This problem can be solved by allowing a violation of H_v threshold for a short period of time, until one of the RT-session ends. The second challenge is maintaining the inter-AP fairness as the users change their attachment points. For that purpose, the NOC periodically re-calculates the current slots-to-users ratio and compares it with best possible ratio. If the gap between the two ratios is significant, then the slot assignment is modified.

VI. THE SLOT ASSIGNMENT AND THE FREQUENCY PLANNING ALGORITHMS

We now turn to describe our algorithms for the slot assignment problem for the two cases, when we have a single frequency and when we have multiple frequencies. First we present two relevant graph coloring problems both of which are NP-hard, and we present efficient approximation algorithms for solving them. These algorithms are then used as a building block for constructing solutions for the single and multiple frequency slot assignment problems.

A. The Graph Coloring and Frequency Assignment Problem

Let $G = (V, E)$ be the interference graph defined in Section V-B. Recall that G is a unit disk graph. We first define a coloring problem for G. We assume that each node $v \in V$ is associated with an integer requirement $r_v \geq 1$, which stands for the number of distinct colors required by node v. We define the coloring problem for G as an assignment of r_v distinct colors S_v to every node $v \in V$, such that no common color is assigned to the two end nodes of any edge $(u, v) \in E$, i.e., $S_u \cap S_v = \emptyset$ and the total number of colors used $|\bigcup_{v \in V} S_v|$ is minimized. Next, we define a joint coloring and frequency assignment problem for G. Here, in addition to finding a coloring for nodes of G we need to assign frequencies from a given set F to the nodes of G. Each node v is assigned a single frequency $f_v \in F$ and $S_u \cap S_v = \emptyset$ for only those edges $(u, v) \in E$, such that $f_u = f_v$. The objective is still to minimize the total number of colors used: $|\bigcup_{v \in V} S_v|$.

It was shown in [18] that the problem of deciding whether a unit disk graph with unit requirements can be colored with 3 colors is NP-complete. This implies that the joint coloring and frequency assignment problem as defined above cannot be approximated to a ratio $4/3$ or better unless $P = NP$. It was also shown in [19] that deciding whether a unit disk graph with unit requirements can be colored with k colors is also NP-complete for all $k \geq 3$. Note that this implies that the more general coloring and frequency assignment problem is also NP-complete for $k \geq 3$. On a positive side, it was shown in [16] that there is a 3-approximation algorithm for the unit disk graph coloring problem with unit requirements and there results also imply a 3-approximation algorithm for the unit disk graph coloring problem with general requirements.

B. The Frequencies and Slot Assignment Algorithm

We now present a 4-approximation algorithm A for the joint coloring and frequency assignment problem. Note that this problem reduces to just the coloring problem, when there is a single frequency. As mentioned in Section VI-A the following is known for the coloring problem for unit disk graphs and therefore directly applicable to the case of a single frequency.

Claim 1: ([16]) There exists a 3-approximation algorithm for the unit disk coloring problem with general requirements.

We denote by $n(v)$ the set of neighbors of node v in the graph G. We assume that the colors are numbered $1, 2, 3 \ldots$ etc. We make the assumption that the location of the APs (*i.e.,* the center of the unit disks) are known. This assumption is used to simplify the presentation and for lack of space. Our results also extend to the case where the AP locations may not be known. Let $v_1, v_2 \ldots v_n$ be the nodes of G ordered in non-decreasing X-coordinate of their locations. Algorithm A processes the vertices in the reverse order $v_n, v_{n-1} \ldots v_1$ and uses a generalized First-Fit for its frequency and color assignment. Let $N(v_i) \subseteq n(v_i)$ be the set of neighbors of node v_i among the nodes $v_{i+1}, v_{i+2} \ldots v_n$. Thus $N(v_n) = \emptyset$. Note that when v_i is considered by A, all the nodes in $N(v_i)$ have already been assigned colors and frequencies by the algorithm. Let $N_f(v_i) \subseteq N(v_i)$ be the set of neighbors of node v_i in $N(v_i)$ that have been assigned frequency $f \in F$ by A. Note that the sets $N_f(v)$, $f \in F$ form a partition of $N(v)$ for all v in G.

To determine the frequency assignment for node v_i, algorithm A applies First-Fit, for each possible frequency $f \in F$. In other words, assuming frequency f, the algorithm computes the least r_{v_i} colors that can be assigned to v_i, while considering only the nodes in $N_f(v_i)$. Then, the algorithm A selects that frequency f for which the maximum color assigned to v_i by First-Fit is minimized (ties broken arbitrarily) and assigns, accordingly, the set of colors to node v_i.

```
Alg. $A$
For $i = n$ down to 1
    $minMaxColor = \infty$
    $f = 0$
    For $j = 1$ up to $|F|$ /* Compute best freq. $f$ for $v_i$ */
        $C = r_{v_i}$-th largest avail. color for $v_i$ for freq. $f_j$
        If $minMaxColor > C$ then
            $minMaxColor = C$
            $f = f_j$
        End If
    End For
    Assign $r_{v_i}$ minimum avail. colors for freq. $f$ to $v_i$
    Assign freq. $f$ to $v_i$
End For
```

Example 6: We consider again the graph G shown in Figure 6. The nodes are ordered in non-decreasing X-coordinate of their locations, which yield the sequence a, b, c, d, e, f. Let F consists of two frequencies f_1 and f_2. Algorithm A, first, assigns color 1 and frequency f_1 to node f. For the next node (node e) the smallest available color for frequency f_1 is 2 and for frequency f_2 is 1. Hence, algorithm A assigns color 1 and frequency f_2 to node e. Note that at this point the set of available colors for node d are all colors except 1 for both frequencies f_1 and f_2. Thus, A assigns colors 2 and 3 and frequency f_1 to node d. Next, algorithm A assigns colors to node c and the minimum available colors are 2 and 3 for frequency f_2 and 1 and 4 for frequency f_2. Hence algorithm A assigns colors 2 and 3 and frequency f_2 to node c. Node b is next and for it the best frequency is f_2 and for this frequency algorithm A assigns it color 1. Finally, for node a the minimum available colors are 1 and 2 for frequency f_1 and 1 and 4 for frequency f_2. Hence, algorithm A assigns colors 1 and 2 and frequency f_1 to node a. Thus, the total number of colors used by algorithm A for G with two frequencies is 3, which is optimal. ■

We now show that A is a 4-approximation algorithm for the unit disk coloring and frequency assignment problem.

Proposition 1: The maximum requirement $\max_v r_v$ is a lower bound on the number of colors used by any optimal solution.

Proof: This trivially follows from the fact that the colors assigned to any node must all be distinct ■

Claim 2: The number of colors used by A is at most

$$\max_i \left\lceil \frac{\sum_{u \in N(v_i)} r_u + r_{v_i}}{|F|} \right\rceil + \max_v r_v.$$

Proof: For contradiction, let this not be the case and let the violation occur first time for node v_j. Then for every frequency f the maximum color among the set of colors obtained by applying First-Fit to the set of colors assigned by A to nodes in $N_f(v_j)$ is strictly greater than

$$\lceil \frac{\sum_{u \in N(v_j)} r_u + r_{v_j}}{|F|} \rceil + r_{v_j}.$$

Thus for every frequency f the number of distinct colors assigned by A to nodes in $N_f(v_j)$ is strictly greater than $\lceil \frac{\sum_{u \in N(v_j)} r_u}{|F|} \rceil$, implying that

$$\sum_{u \in N_f(v_j)} r_u > \lceil \frac{\sum_{u \in N(v_j)} r_u}{|F|} \rceil$$

for all f. But this cannot happen since the sets $N_f(v_j)$, $f \in F$ form a partition of the set $N(v_j)$ and would thus imply that

$$\sum_{u \in N(v_j)} r_u = \sum_{f} \sum_{u \in N_f(v_j)} r_u > |F| \lceil \frac{\sum_{u \in N(v_j)} r_u}{|F|} \rceil \geq \sum_{u \in N(v_j)} r_u.$$

■

Claim 3: Any coloring and frequency assignment algorithm must use $\max_i \lceil (\sum_{u \in N(v_i)} r_u + r_{v_i})/(3|F|) \rceil$ colors.

Proof: Let $v_1, v_2 \ldots v_n$ be the nodes of G ordered in non-decreasing X-coordinate of their locations. Thus node v_i has the minimum X-coordinate among all nodes in $\{v_i\} \cup N(v_i)$. Thus as shown in [16] the set of nodes $\{v_i\} \cup N(v_i)$ do not contain an independent set of size more than 3 in G. Thus the nodes $\{v_i\} \cup N(v_i)$ require at least $\lceil (\sum_{u \in N(v_i)} r_u + r_{v_i})/(3|F|) \rceil$ distinct colors in any feasible coloring of G, since the nodes assigned a particular color and a particular frequency must form an independent set of G. Thus, any coloring and frequency assignment algorithm must use at least $\lceil (\sum_{u \in N(v_i)} r_u + r_{v_i})/(3|F|) \rceil$ colors. Taking the maximum for all i we get the desired result. ■

Claim 4: A is a 4-approximation algorithm for the coloring and frequency assignment for unit disk graphs.

Proof: It can be shown that $\lceil 3x \rceil \leq 3\lceil x \rceil$ for all non-negative real numbers x. Thus we have

$$\max_i \lceil \frac{\sum_{u \in N(v_i)} r_u + r_{v_i}}{|F|} \rceil \leq 3 \max_i \lceil \frac{\sum_{u \in N(v_i)} r_u + r_{v_i}}{3|F|} \rceil.$$

Recall that from Claim 2 the number of colors used by algorithm A is at most $\max_i \lceil (\sum_{u \in N(v_i)} r_u + r_{v_i})/|F| \rceil + \max_v r_v$, which is at most $3 \max_i \lceil (\sum_{u \in N(v_i)} r_u + r_{v_i})/(3|F|) \rceil + \max_v r_v$. By Claim 3 and Proposition 1, it is at most $3 + 1 = 4$ times the number of colors used by an optimal coloring and frequency allocation. ■

C. The Fair Slot Assignment Scheme

As described earlier, the problem of providing fairness in our model is equivalent to assigning slots (colors) to access points in a superframe such that the total number of slots (colors) assigned to an access point v in a superframe is proportional to the number of stations m_v associated with the access point. In addition in our model fairness is provided with the additional goal of maximizing the throughput of the system. Recall that in our model the slot sizes are assumed to be at least Δ time units. Let $\tilde{D}$ denote the configured CFP in time units. Thus the maximum number of slots in a CFP cannot exceed $R = \lfloor \tilde{D}/\Delta \rfloor$. Thus the problem of providing fairness in our model can be formulated as the problem of finding the largest *min-slots-to-users ratio* ρ such that there is a feasible superframe slot assignment in which access point $v \in V$ is assigned $r_v = \lceil \rho \cdot m_v \rceil$ slots. Note that a superframe slot assignment is feasible if there is a coloring and frequency assignment of the underlying graph $G = (V, E)$ in which node $v \in V$ has requirement r_v and the total number of colors used is at most $R = \lfloor \tilde{D}/\Delta \rfloor$. We now show that the slot-assignment problem as stated above is very hard not just to solve optimally but even to solve approximately. We then turn to a relaxation of the problem for which we develop efficient algorithms.

Claim 5: No polynomial time constant approximation algorithm is possible for the slot-assignment problem unless $P = NP$.

Proof: For contradiction let us assume a c-approximation algorithm X for some constant c for the slot-assignment problem. This implies that algorithm X outputs a $\rho > c\rho^*$ where ρ^* is the optimal *min-slots-to-users ratio*. We show that X can be used to decide whether a unit disk graph with all unit requirements and a single frequency can be colored with 3 colors. As mentioned in Section VI-A this is a NP-hard problem thus implying that X does not exist unless $P = NP$. Given a unit disk graph $H = (V, E)$ an instance of the slot-assignment problem is created with $G = H$, $m_v = 1$ for all $v \in V$, total number of slots (colors) $R = 3$ and a single frequency. Note that if H is 3-colorable then for G the value of $\rho = k \geq 1$ for some k and if H is not 3-colorable then $\rho = 0$. In the former case X outputs $\rho \geq kc$ while in the latter case X must output $\rho = 0$. Since $0 < kc$ the return value of algorithm X can be used to determine if H is 3-colorable or not. Thus establishing that the slot-assignment problem is hard to approximate. ■

In light of Claim 5 we turn to a different bi-criteria approximation for the slot-assignment problem. We define an (α, β) approximation algorithm to be one that computes the value of ρ to within a factor $\alpha \leq 1$ of the optimal *min-slots-to-users ratio* where the optimal is only allowed at most $\lfloor R/\beta \rfloor$ slots. Note that Claim 5 implies that a $(c, 1)$-approximation is not possible for any constant c. Note that allowing optimal to use less total number of slots is equivalent to restraining the optimal to use a bigger slot size since the CFP size is fixed at $\tilde{D}$ time units.

We first present a $(1, 3)$-approximation algorithm X for the slot-assignment problem when there is a single fre-

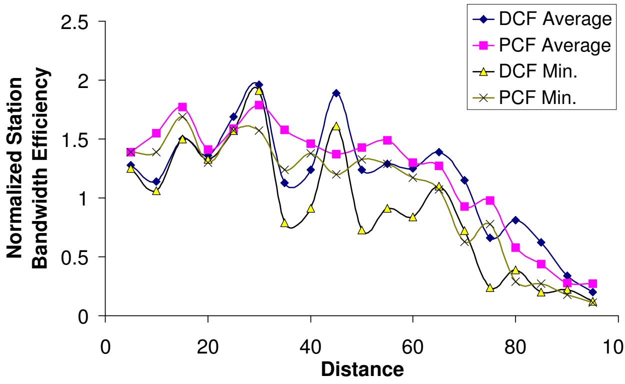

Fig. 9. The average and minimum normalized network efficiency in DCF and PCF modes.

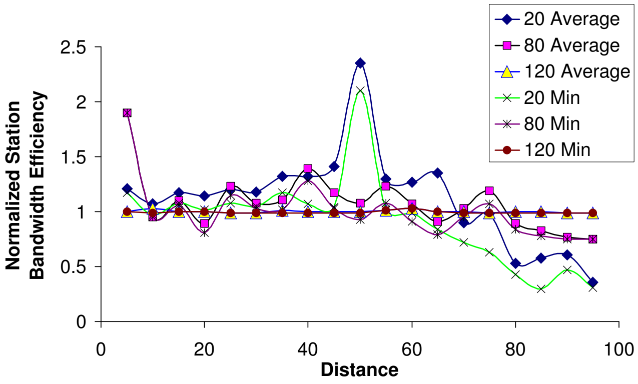

Fig. 10. The average and minimum normalized network efficiency of the MiFi system.

quency. The algorithm X works by guessing a value for the *min-slots-to-users ratio* ρ. For a particular guess for ρ it sets $r_v = \lceil \rho \cdot m_v \rceil$ for all $v \in V$. It then uses the unit-disk graph coloring algorithm referred to in the Claim 1 to color the underlying graph. Let $f(\rho)$ denote the number of colors used by this algorithm for coloring this graph for the choice of ρ. Note that $f(\rho)$ is a monotonically non-decreasing function of ρ. Algorithm X uses a binary search over ρ to compute the largest value ρ^* for which $f(\rho^*) \leq R$. Algorithm X then outputs $\rho = \rho^*$. Note that algorithm X need only consider those values of ρ for which $\max_v \lceil \rho \cdot m_v \rceil \leq R$.

Claim 6: Algorithm X is a $(1, 3)$-approximation for the slot-assignment problem when only one frequency is used.

Proof: Let us denote by ρ_{OPT} the optimal value of ρ for the slot assignment problem with $\lfloor R/3 \rfloor$ total slots. Thus, when the node requirements are set to $r_v = \lceil \rho_{OPT} \cdot m_v \rceil$ for all $v \in V$ then the underlying graph G can be colored with at most $\lfloor R/3 \rfloor$ colors. The algorithm referred to in the Claim 1 is then able to color G for these node requirements with at most $3\lfloor R/3 \rfloor \leq R$ colors implying that $\rho = \rho_{OPT}$ is feasible for algorithm X, and $f(\rho_{OPT}) \leq R$. Thus algorithm X outputs $\rho \geq \rho_{OPT}$, implying that algorithm X is a $(1, 3)$-approximation for the slot-assignment problem. ■

For the case when there can be more than one frequency the slot assignment algorithm Y works exactly like algorithm X described earlier, except that it uses algorithm A defined in Section VI-B, instead of the the algorithm referred to in the Claim 1. The proof that Y is a $(1, 4)$-approximation algorithm, follows along the same line as the proof that X is a $(1, 3)$-approximation algorithm and is omitted for lack of space.

VII. SIMULATION RESULTS

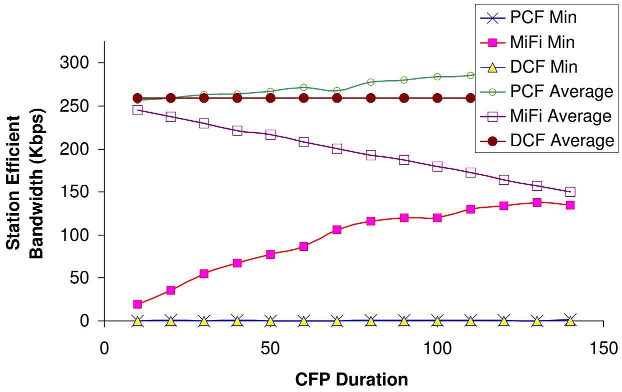

Fig. 11. The minimum and average efficient bandwidth of *all* the stations in date rate of 10Mpbs.

We did extensive simulation to compare the performance of the MiFi system with that of a WiFi. Our simulation compares the performance of both the PCF and the DCF modes of 802.11 with that of the MiFi system, and we used three different performance metrics, described below. We quantify these performance metrics in terms of the *efficient bandwidth* of the stations, which reflects the amount of data that the station transmits or receives successfully in a given time period. This parameter is a good measure of both the fairness and the throughput of the system.

Simulation Setup: For the simulation we used a 802.11 network with 50 APs, uniformly distributed over a grid of size 1000×1000. Each AP has a transmission range of 100 units. The AP distribution is picked to ensure complete coverage of the grid and we follow the layout approach of [20]. We assumed 1000 mobile stations in the system that always have pending message to send and considered a message length of 1500 bytes (12000bits). For both the PCF mode and the MiFi system we used a superframe size of 150ms. Each simulation involved running the system for 1 minute and observing its behavior.

We tried our simulation both with a single frequency and with optimal frequency planning when there are 3 non-interfering frequencies, at different bit rates ranging from 1 to 10 Mbps per AP and for different CFP window sizes ranging from 10ms to 140ms. For lack of space the results presented, unless mentioned otherwise, are for a system with optimal frequency planning of 3 frequencies at 10 Mbps bit-rate, and CFP window set to 80ms. We note that the results we observed are typical for all our simulations. We used an ideal network planning to illustrate the significant unfair behavior of the DCF and PCF mode even in optimally planned networks.

The first metric that we use to measure the relative performance of the MiFi system with respect to the 802.11 based WLAN is *intra-AP* fairness. Here we measure the *normalized* efficient bandwidth of the stations as a function of their distance from their associated APs. The normalized efficient bandwidth for a user is computed by dividing the efficient bandwidth of the user by the average efficient bandwidth of all the users that are associated with its AP. We measured both the average and the minimum values for the normalized efficient bandwidth for both the PCF and DCF modes of 802.11, and we compare it to that for the MiFi system. Note that an ideal (fair) system should have both a average and minimum normalized station efficient bandwidth of 1 at all distances from the AP. Figure 9 shows our results for the 802.11 PCF and DCF mode for both the average and minimum normalized network efficiency of the stations within the transmission range of the AP. While, Figure 10 shows the MiFi system results for the same parameters for three different choices of the CFP window size: 20ms, 80ms, and 120ms. Our results show that in the 802.11 case the normalized efficient bandwidth of the stations far away from the AP is close to 0, both for the average and the minimum metrics. On the other hand for the MiFi system both the average and the minimum normalized station efficient bandwidths are close to 1 at all distances even for the CFP window of 80ms and the behavior of the MiFi system is very close to that of an ideal system even when the CFP window is of size 120ms.

Our second metric is *inter-AP* fairness. Here we measure the minimum and average efficient bandwidth of *all* the stations in the system as a function of the size of the CFP window size (for PCF and MiFi) for both the MiFi system and the 802.11 DCF and PCF modes. The results are presented for the rate of 10 Mbps in Figure 11. Our results indicate that the minimum value is almost 0, indicating starvation, for both PCF and DCF mode for all CFP window sizes. For the MiFi system however as the CFP window size is increased the minimum value rapidly approaches the average efficient bandwidth of *all* stations. Even for small CFP window sizes the gap between the minimum and the average is not significant for the MiFi system. These results show that the MiFi system is starvation free and provides excellent fairness.

Our third metric is the overall system throughput. Note that the system throughput is the average efficient bandwidth of *all* stations times the number of stations. Thus the results for the average efficient bandwidth in Figure 11, when multiplied by 1000 gives the system throughput, as a function of the CFP window size, for both the MiFi system and the 802.11 DCF and PCF modes for bit rates of 10 Mbps. These results show that even for large CFP window size the overall system throughput of the MiFi system is comparable to that of the 802.11 networks. Moreover, we noticed that there is an optimal MiFi CFP size of 130ms, at which the minimal efficient bandwidth of any station is maximized.

REFERENCES

[1] C. E. Koksal, H. Kassab, and H. Balakrishnan. An analysis of short-term fairness in wireless media access protocols (poster). In *Measurement and Modeling of Computer Systems*, 2000.

[2] T. Nandagopal, T-E. Kim, X. Gao, and V. Bharghavan. Achieving mac layer fairness in wireless packet networks. In *Proceedings of MobiCom '2000*, 2000.

[3] Q. Ni, L. Romdhani, T. Turletti, and I. Aad. Qos issues and enhancements for ieee 802.11 wireless lan. Technical report, INRIA, 2002.

[4] D. Chalmers and M. Sloman. A survey of quality of service in mobile computing environments. *IEEE Communications Surveys*, 2(2), 1999.

[5] N. H. Vaidya, P. Bahl, and S. Gupta. Distributed fair scheduling in a wireless lan. In *Proceedings of MobiCom'2000*, 2000.

[6] T. Ozugur, M. Naghshineh, P. Kermani, C. M. Olsen, B. Rezvani, and J. A. Copeland. Balanced media access methods for wireless networks. In *Proceedings of MobiCom '98*, 1998.

[7] I. Aad and C. Castelluccia. Differentiation mechanisms for ieee 802.11. In *proceeding of INFOCOM'01*, 2001.

[8] D. Qiao and K. G. Shin. Achieving efficient channel utilization and weighted fairness for data communications in ieee 802.11 wlan under the dcf. In *Proceedings of IWQoS'2002*, 2002.

[9] A. Veres, A. T. Campbell, M. Barry, and L-H. Sun. Supporting service differentiation in wireless packet networks using distributed control. *IEEE JSAC*, 19(10):2081–2093, 2001.

[10] S-T. Sheu and T-F Sheu. Dbase : A distributed bandwidth allocation/sharing/extension protocol for multimedia over ieee 802.11 ad hoc wireless lan. In *proceeding of INFOCOM'01*, 2001.

[11] A. Kopsel and A. Wolisz. Voice transmission in an ieee 802.11 wlan based access network. In *Proceedings of WoWMoM*, 2001.

[12] C. Coutras, S. Gupta, and N. B. Shroff. Scheduling of real-time traffic in ieee 802.11 wireless lans. *Wireless Networks*, 6(6):457–466, 2000.

[13] S. Mangold. Coexistence of overlapping basic service sets. In *Proceedings of the Mobile Venue'02*, 2002.

[14] T. S. Rappaport. *Wireless Communication Principle and Practice*. Prentice Hall, 1996.

[15] B. O'Hara and A. Petrick. *The IEEE 802.11 Handbook: A Designer's Companion*. Standards Information Network, IEEE press, 1999.

[16] M. V. Marathe, H. Breu, H. B. Hunt III, S. S. Ravi, and D. J. Rosenkrantz. Simple heuristics for unit disk graphs. *Networks*, 25:59–68, 1995.

[17] D. S. Hochbaum and W. Maass. Approximation schemes for covering and packing problems in image processing and vlsi. *Journal of the ACM*, 31:130–136, 1985.

[18] B. N. Clark, C.J. Colburn, and D.S. Johnson. Unit disk graphs. *Discrete Mathematics*, 86(1-3):165–177, 1990.

[19] M. Stumpf A. Graf and G. Weibenfels. On coloring unit disk graphs. *Algorithmica*, 20(3):277–293, 1998.

[20] A. Hills. Large-scale wireless lan design. *IEEE Communications Magazine*, 39(11):98–107, 2001.