

A Dynamic Radio Resource Management Technique for Multiple APs in WLANs

Ming Yu, *Senior Member, IEEE*, Hui Luo, *Member, IEEE* and Kin K. Leung, *Fellow, IEEE*

Abstract—As more and more wireless LANs are deployed in many popular locations, there is a need for dynamic radio resource management (RRM) schemes. This paper proposes a new dynamic RRM technique for multiple APs network. First, we develop a real-time algorithm to estimate the number of active stations from the standpoint of an AP, instead of a station. Second, we propose a dynamic RRM algorithm that not only significantly increases network performance but also reduces the co-channel interference. Third, we find the optimal switching probability that minimizes the transient time for an AP to reach its equilibrium state. For uniform traffic conditions, our scheme produces classical optimal frequency assignments. For non-uniform traffic conditions, it produces sub-optimal frequency assignments. Our simulations have shown that the proposed RRM technique adapts to dynamic network conditions and increases the channel throughput up to 65% in less than 20 iterations. The results can be applied to practical WLANs to significantly improve network performance.

Index Terms—Wireless LAN, resource management, radio spectrum management, cochannel interference.

I. INTRODUCTION

RECENTLY, there have been many research efforts in the IEEE 802.11 WLANs, which have been widely deployed in many locations such as homes, buildings, and campuses. The existing research is mainly focused on the protocol modeling and performance tuning to improve the capacity of the protocols [1]–[3]. Because the IEEE 802.11 standard only provides a limited number of usable channels, the signal coverage among many neighboring APs has to be overlapped in these popular locations. It has been observed that the channel contentions among neighboring APs result in worse performance than expected, such as very low throughput. Therefore, how to improve the network performance and also reduce the co-channel interference is still a problem.

There are two practical reasons why WLANs have been densely deployed in many popular locations. The first is to eliminate coverage holes for a large-scale WLAN that covers an entire building or campus. The WLAN has to maintain a high SNR to assure high data rates everywhere. Therefore, the coverage areas of APs have to be overlapped. The second is that many APs in these locations are owned by different

service providers and there is no coordination during the deployment.

The current solutions could only partially solve the problem, by static Radio Resource Management (RRM) techniques [4], [5], such as, a careful frequency planning and AP placement, a scan for radio energy by an AP to choose a channel with the least level, thus to separate the neighboring APs that share the same channel. This is usually done in the initial installation or when the AP is initialized. It cannot adapt to local environment changes and traffic conditions afterwards. In [6], the authors proposed a channel allocation technique that maximizes the per-user throughput in dense WLANs, which is based on an assumption that each AP knows and periodically broadcasts the number of active stations that are associated with it, which is too strong for practical networks, especially when the APs are owned by different companies in a densely deployed area.

In this paper, we present a new dynamic RRM technique based on the dynamic IEEE 802.11 model [1]. We develop a real-time algorithm to estimate the number of active stations from the standpoint of an AP [7], which include those that are formally associated with the AP and also those that use the same channel in the neighboring area of the AP. In the dynamic RRM technique, an AP finds an optimal radio channel that is less overloaded by actively monitoring its channel utilization. By switching to this channel, the overall network performance will be increased. At the same time, the co-channel interference among different APs will be reduced. We develop a heuristic algorithm to optimize the frequency assignment locally for each AP in the network. This way the AP determines the best channel it should use for the next transmission. Then, it switches to that channel with an optimal switching probability. We also derive an optimal switching probability that can minimize the transient time for an AP to reach the equilibrium state. Our simulations have shown that the proposed scheme quickly adapts to dynamic network environment without relying on message exchanges among neighboring APs and outperforms the existing frequency assignment schemes.

The remainder of this paper is organized as follows. In Section II, the dynamic IEEE 802.11 protocol will be briefly discussed. In Section III, a real-time estimation algorithm will be presented to estimate the number of active stations from the standpoint of an AP, instead of a station. In section IV, the dynamic frequency assignment scheme is introduced. The optimal switching probability is investigated in Section V. The simulation results are presented in Section VI. Section VII discusses the advantages of this proposed technique and concludes this paper.

Part of this paper has been presented in IEEE International Conference On Networks, Singapore, 2004.

M. Yu is with the Department of Electrical & Computer Engineering, State University of New York, Binghamton, NY 13902. He can be reached via email at: mingyu@binghamton.edu.

H. Luo is with Broadcom Inc., Matawan, NJ. He can be reached via email at: hluo@broadcom.com.

K. Leung is with the Department of Electrical & Electronic Engineering, Imperial College, London, UK. He can be reached via email at: kin.leung@imperial.ac.uk.

II. THE DYNAMIC IEEE 802.11 PROTOCOL MODEL

It is well known that the standard IEEE 802.11 protocol capacity can be significantly increased by either appropriately tuning the backoff window size according to the number of active stations, or modifying the backoff distribution to uniformly spread the channel access in a contention window. For this reason, the authors in [1] proposed a dynamic IEEE 802.11 protocol, which is based on the p -persistent CSMA protocol for radio wireless. In the dynamic protocol, a station transmits at the beginning of an empty slot with a probability p , and defers the transmission with probability $1 - p$. The p value is dynamically adjusted to determine the average backoff time, based on the network configuration and load conditions. In the standard protocol, the backoff algorithm uses a set of predefined slotted windows and doubles the window size after each collision. The length of the backoff interval decreases only when the channel is sensed idle.

It is worth pointing out that the only difference between the dynamic and standard protocol is in the selection of the backoff interval. Identifying an optimal p value in the dynamic protocol is equivalent to identifying the average backoff window size in the standard protocol. Therefore, the methods used to adjust the average backoff time in the dynamic protocol can be also used in the standard protocol to select the appropriate size of the contention window. The dynamic protocol provides a mathematical model for the standard protocol to tune its protocol performance. Therefore, we base our RRM technique on the dynamic protocol.

In this paper, we assume that a station can dynamically adjust the p value within the contention domain of a single AP, in order to operate near or at the protocol capacity, which has been extensively investigated by existing methods [1], [2]. Our focus is to reassign the radio channels for those APs that are overloaded with stations in order to improve the overall network performance. The APs will be responsible for collecting network status, estimating the number of active stations, and computing the channel utilization. Each AP and its stations still use the dynamic protocol.

The time interval between two successful transmissions is referred to as virtual-transmission time (VTT). A VTT includes one successful transmission and maybe several collisions and idle slots, as shown in Fig 1. The channel utilization is defined as the fraction of channel bandwidth used by successfully transmitted messages. The maximum value of the channel utilization is known as the protocol capacity. The utilization has been defined in [1] that

$$\rho = \bar{t}_s / \bar{t}_v, \quad (1)$$

where \bar{t}_s is the average time interval that the channel spent on a successful transmission; \bar{t}_v is the average VTT of the channel. The probability that a station is in idle, or successful transmission, or collision is the probability of no station, or exactly one station, or more than one station transmitting at the beginning of an empty slot. For a network with M stations, each transmits with a p probability, the three probabilities can

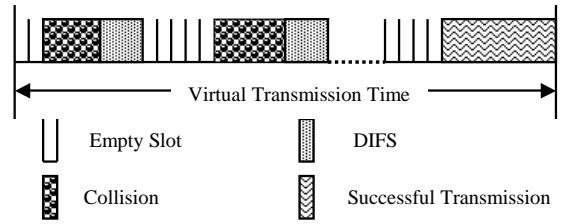


Fig. 1. IEEE 802.11 MAC Model

be easily found as p_0 , p_1 , and $1 - p_0 - p_1$, respectively, where

$$p_0 = (1 - p)^M, \quad (2)$$

$$p_1 = Mp(1 - p)^{M-1}. \quad (3)$$

Thus, the two values \bar{t}_s and \bar{t}_v can be calculated as

$$\bar{t}_s = \bar{L}_m t_B p_1, \quad (4)$$

$$\bar{t}_v = t_{slot} p_0 + E_{suc} p_1 + (E_{col|col} + O_H)(1 - p_0 - p_1), \quad (5)$$

where \bar{L}_m is the average message length in bytes; E_{suc} is the average duration of a successful transmission, given a transmission attempt; $E_{col|col}$ is the conditional average duration of a collision, given that a collision occurs; t_B and t_{slot} are the time needed to transmit a byte and a slot, respectively; O_H is the collision overhead that includes the propagation delay and interframe spacing time. For the DCF access method, $O_H = \tau + DIFS$, where τ is the propagation delay over the wireless channel; $DIFS$ is the DCF interframe spacing time that a station has to wait before attempting to acquire the channel to send a new frame. For the RTS/CTS access method, $O_H = \tau + EIFS$, where $EIFS$ is the extended interframe spacing time, which is used by a station that has just received a bad or unknown frame to report the bad frame.

It is worth mentioning that E_{suc} is only a function of traffic load, i. e., the message length distribution. But $E_{col|col}$ is a function of network configuration, i. e., the number of active stations M , and the p value, in addition to the message length distribution. The detail expressions for E_{suc} and $E_{col|col}$ can be found in [3].

By substitute Eqns. (4) and (5) into Eqn. (1), we can find

$$\rho = \bar{L}_m t_B / \left[\frac{p_0}{p_1} t_{slot} + E_{suc} + \frac{1 - p_0 - p_1}{p_1} (E_{col|col} + O_H) \right], \quad (6)$$

Let's denote N_I and N_C as the average number of idle slots, and the average number of collisions in a VTT, respectively. We can find that

$$N_I = \frac{p_0}{p_1} = \frac{1 - p}{Mp} \quad (7)$$

$$N_C = \frac{1 - p_0 - p_1}{p_1} = \frac{1 - (1 - p)^M}{Mp(1 - p)^{M-1}} - 1 \quad (8)$$

As we know that the average number of idles in a VTT is greater than the average number of collisions by one. If we denote average duration of an idle period as E_{idl} , then $N_I t_{slot} = E_{idl}(N_C + 1)$, i. e.,

$$E_{idl} = \frac{N_I t_{slot}}{N_C + 1} = \frac{(1-p)^M}{1-(1-p)^M} t_{slot}. \quad (9)$$

By substitute Eqns. (9) and (8) into Eqn. (6), we can find

$$\rho = \frac{\bar{L}_m t_B}{(N_C + 1)E_{idl} + E_{suc} + N_C(E_{col|col} + O_H)}, \quad (10)$$

We know $N_C/(N_C + 1)$ is the probability that a collision occurs given a transmission attempt. If we denote average duration of a collision as E_{col} , we have

$$E_{col} = \frac{N_C}{N_C + 1} E_{col|col}.$$

We can find

$$E_{col|col} = \frac{N_C + 1}{N_C} E_{col}. \quad (11)$$

By substitute Eqn. (11) into Eqn. (10), we have

$$\rho = \frac{\bar{L}_m t_B}{(N_C + 1)E_{idl} + E_{suc} + (N_C + 1)E_{col} + N_C O_H}, \quad (12)$$

It has been observed that the same expected overall time is wasted on idles and collisions when the CSMA Aloha reaches its maximum departure rate [8]. In [2], the authors proposed to approximate the condition when the IEEE 802.11 MAC protocol reaches its capacity by

$$E_{col|col} N_C = (N_C + 1)E_{idl},$$

which has been verified by extensive simulations. Compared to Eqn. (11), it is clear that the utilization reaches maximum value if

$$E_{idl} = E_{col}, \quad (13)$$

holds. In the dynamic IEEE 802.11 protocol, by choosing a p value such that $E_{idl} = E_{col}$ to force the network operate at maximum throughput.

III. REAL-TIME ESTIMATION OF AP STATUS

In a CSMA/CA protocol, every time an AP successfully receives a data frame from a station, the AP will send a positive acknowledgement frame to the station, after a short interframe spacing (SIFS) time, which is less than the DIFS. If the station has not receive the ACK, it will consider the previously observed busy period as a collision. Similarly for the downlink traffic the AP observes all the busy and idle periods of the channel. If the AP has not received an acknowledgment after sending out a data frame, it will schedule a retransmission of the data frame in the next available slot. In this way, the AP can explicitly detect a collision period, in addition to the successful transmission and idle periods. Therefore, we can assume that an AP can obtain a channel status by measuring the collision, successful transmission, and idle periods.

In order to develop a dynamic RRM technique that can optimally assign radio channels among different APs, we need to calculate the utilization of a channel from the standpoint of an AP, instead of a station. Although the AP can also get the estimated values of E_{idl} and E_{col} by using measurements, as a station in [2], it is not responsible for choosing the optimal p value. Therefore, Eqn. (13) cannot be used to estimate the M

value. Initially, each AP does have the count of its associated stations. But in a dynamic network environment, or with co-channel neighboring APs, the active number of stations that have messages to send will still contend for a same radio channel even if these stations may be associated with different APs. Thus, we still need to estimate the M value. To estimate the number of active stations associated with an AP, we decide to use the measurement of E_{idl} and N_I , instead of E_{col} , which involves more computations.

Let's denote the time interval between two consecutive transmission attempts as transmission interval. The transmission attempt can be either a successful or a colliding attempt. We assume that at the end of the n th transmission interval, the AP has the values of $\hat{E}_{idl}(n)$ and $\hat{N}_I(n)$, which are denoted as the measurements of E_{idl} and N_I , respectively. During the $(n + 1)$ th transmission interval, these measured values will be used to obtain the estimations of E_{idl} and N_I , which are denoted as $\hat{E}_{idl}(n + 1)$ and $\hat{N}_I(n + 1)$, respectively.

To obtain smooth estimations, we adopt the moving averaging model:

$$\hat{N}_I(n + 1) = \alpha_1 \hat{N}_I(n) + (1 - \alpha_1) \hat{N}_I(n + 1), \quad (14)$$

$$\hat{E}_{idl}(n + 1) = \alpha_2 \hat{E}_{idl}(n) + (1 - \alpha_2) \hat{E}_{idl}(n + 1), \quad (15)$$

where $0 < \alpha_1 < 1$ and $0 < \alpha_2 < 1$ are smoothing factors used to adjust the estimation speed and accuracy. The values of α_1 and α_2 can be appropriately chosen in terms of specific application and requirement. In practice, after a few times of trying, or if we have observed the changing pattern of the variables, we can find a good compromise between accuracy and promptness. In our simulation, we find that $\alpha_1 = 0.9$ and $\alpha_2 = 0.9$ is a good value for the smoothing factors.

From Eqns. (7) and (9), we can see that N_I and E_{idl} are functions of p_0 and p_1 . After we obtain the estimations of N_I and E_{idl} , we can also find the estimations of p_0 and p_1 . Specifically, at the end of the $(n + 1)$ th transmission interval, each AP computes $\hat{p}_0(n + 1)$ and $\hat{p}_1(n + 1)$, which are denoted as the estimations of p_0 and p_1 , respectively. By some rearrangements on Eqns. (7) and (9), we have

$$\hat{p}_0(n + 1) = \frac{\hat{E}_{idl}(n + 1)}{\hat{E}_{idl}(n + 1) + t_{slot}} \quad (16)$$

$$\hat{p}_1(n + 1) = \frac{\hat{E}_{idl}(n + 1)}{\hat{N}_I(n + 1)[\hat{E}_{idl}(n + 1) + t_{slot}]} \quad (17)$$

In order to find the M value, by some manipulations on Eqns. (2) and (3), we have

$$\left(\frac{M p_0}{M p_0 + p_1}\right)^M = p_0. \quad (18)$$

We can see that only numeric solution exists for M if we know p_0 and p_1 . Therefore, we propose an iterative method by replacing the M 's in Eqn. (18) with estimations obtained at different transmission intervals. By combining Eqns. (16) and (17), we can develop an algorithm as follows:

$$\tilde{M}(n + 1) = \log \hat{p}_0(n + 1) / [\log(\tilde{M}(n) \hat{p}_0(n + 1)) - \log(\tilde{M}(n) \hat{p}_0(n + 1) + \hat{p}_1(n + 1))]. \quad (19)$$

In this way, we can compute a measurement value \tilde{M} for M in terms of Eqns. (16) and (17). Similar to Eqn. (14), we also use a moving average model to smooth the fluctuations of the estimations:

$$\hat{M}(n+1) = \gamma_0 \hat{M}(n) + \gamma_1 \tilde{M}(n+1) + (1 - \gamma_0 - \gamma_1) \tilde{M}(n), \quad (20)$$

where $\hat{M}(n+1)$ is the estimated M value at the $(n+1)$ th transmission interval; $0 < \gamma_0 < 1$ and $0 < \gamma_1 < 1$ are smoothing factors, which can be chosen in terms of operation requirement. To further smooth the fluctuations of estimations, we use the second-order moving average model in Eqn. (20), where $0 < \gamma_0 + \gamma_1 < 1$.

We summarize that each AP estimates its number of active stations in real-time by using the measurement about the average number of idle slots in a VTT and the average duration of an idle period. These measurements are obtained by using the AP's carrier sensing mechanism.

When an AP switches to a different radio channel, it first estimates the number of active stations associated with the channel. After a short transition period, the estimation processes quickly converge to a value around the real value, which will be demonstrated by simulation results. Therefore, the AP does not need a prior knowledge about the M value.

IV. THE DYNAMIC FREQUENCY ASSIGNMENT SCHEME

A. The RRM Problem Formulation

The goal of RRM in WLANs is to effectively allocate the available radio channels to the APs such that the overall network performance can be maximized and the co-channel interference can be minimized. Here, the channel utilization of an AP represents the network performance in the WLAN service area. This is different from the traditional coverage-oriented WLAN capacity design and planning, which mainly addresses how to deliver the maximum signal to every WLAN service area. The AP placement problem [5], is more like a static RRM technique, which addresses how to assign APs to the best location to meet traffic demands and also assign channels to the APs. The channel assignment is usually formulated as an integer linear programming (ILP) problem and found to be NP-complete or NP-hard. In a dynamic network environment, the RRM needs to consider the changes in network and traffic conditions over time, including co-channel interference.

In this paper, the interference is considered in two aspects. First, the estimated M value includes both the formally associated stations and the co-channel stations. Second, the impact of the co-channel APs on the channel utilization of an AP is taken into account by defining an effective channel utilization. The transmissions of one or more co-channel APs can cause enough interference to be sensed as busy by the AP. For simplicity, we consider only the strongest interfering set of APs, S , in which a single such AP can cause the channel to be sensed as busy.

Assume that there are J non-overlapping channels, indexed from 1 to J , $I_J = \{1, 2, \dots, J\}$, available for allocation to I APs, indexed from 1 to I , $I_I = \{1, 2, \dots, I\}$. Since the CSMA protocol used in IEEE 802.11 prohibits APs from transmitting

when a given channel is sensed busy, we define the effective channel utilization U_i^j as the fraction of time at which channel j can be sensed busy or is used for transmission by AP_i , that is,

$$U_i^j(n) = \rho_i^j(n) + \sum_{k=1}^J C_i^k(n) \sum_{l \in S_i} C_l^k(n) \rho_l^k(n), \quad (21)$$

where $\rho_i^j(n)$ is the channel utilization of AP_i when it operates on channel j at the n th channel updating cycle; C_i^j is the interference coefficient that $C_i^j = 1$ if AP_i uses channel j and 0 otherwise. The number of APs co-channelled with AP_i on channel j , is $C_i^j(n)$. The interfering set for AP_i is denoted as S_i . An AP can identify its interfering set by observing the received signal power level from its neighboring APs. For simplicity, we assume that, in Eqn. (21), each co-channelled APs transmits at the same probability as AP_i . Therefore, a co-channelled AP is assumed to have the same channel utilization as AP_i .

Mathematically, the problem can be stated similarly to the static RRM [4]. The objective function for optimal channel assignment can be defined as follows:

$$\min_{i \in I_J} \max_{i \in I_I} \{U_i^j(n)\}, \quad \text{for } i \in I_I, j \in I_J; \quad (22)$$

Under the constraint that

$$U_i^j(n) < 1, \quad \text{for } i \in I_I, j \in I_J; \quad (23)$$

at the n th channel adjustment interval during channel allocation and reallocation period. The objective function in Eqn. (22) is to assign channels such that the effective utilization of the most heavily loaded AP is minimized. This results in more resource available for the most heavily loaded AP for a given traffic load, thus improves the overall network throughput. The stability condition in Eqn. (23) is to make sure that all traffic can be sent eventually.

B. The Dynamic Frequency Assignment Algorithm

Generally, the solution to the optimization problem in Eqns. (22) and (23) may not exist by using ILP methods. But for our goal in this paper, the problem can be solved by using the explicitly computed utilization obtained in Eqn. (12), for a small scale, i. e., $J = 3$.

The new dynamic frequency assignment algorithm can be outlined as follows:

1. Initially, each AP is randomly assigned a radio channel from the J available channels. The current assignment is treated as the best assignment so far.

2. Each AP estimates M value, the number of active stations in real-time, including both its associated stations and the co-channelled stations that use the same channel but with other APs in the vicinity area.

Note that the co-channel interference is taken into consideration implicitly by the fact that the N_J and E_{idl} values, which are sensed by the AP and used to estimate the M value, are directly impacted by the co-channel stations. The estimated M value reflects the number of active stations that are associated with a virtually formed AP.

3. Each AP computes its effective channel utilization by using Eqn. (21), under the current assigned channel. Denote the utilization of AP_i is $U_{i,j}(n)$ when using channel j , $i \in \{1, 2, \dots, I\}$, and $j \in \{1, 2, \dots, J\}$, at the n th channel adjustment period.

4. Based on all the channels an AP has used so far, AP_i finds its the maximum channel utilization, which corresponds to minimum number of active stations. For AP_i , let's denote $V_{i,k}(n) = \min \{U_{i,j}(n)\}$, where $j = 1, 2, \dots, J_n$ are the channels AP_i that has used so far. By using channel k , AP_i reaches its maximum utilization.

5. For AP_i , compare the current maximum value, $V_{i,k_1}(n)$, to its previous maximum value $V_{i,k_2}(n-1)$, where $k_2 = k_1$ and $k_1 = k$, k_1 and k_2 are the channels that maximize the utilization of AP_i at the n th and $(n-1)$ th channel adjustment periods, respectively. Let $W_{i,k_{min}}(n) = \min \{V_{i,k_1}(n), V_{i,k_2}(n-1)\}$, where $k_{min} \in \{k_1, k_2\}$ is the channel that minimizes the maximum utilization of AP_i .

6. Check the inequality:

$$|W_{i,k_1}(n) - W_{i,k_2}(n-1)| \leq \beta, \quad (24)$$

where β is a predefined constant that determines the degradation of quality of service (QoS) of the network. If Eqn. (24) holds, stop the current channel adjustment phase. Go to step 2, to monitor the network performance.

7. The AP_i switches to the channel k_{min} , with a switching probability p_s , where p_s is an optimal probability that can minimize the overall adjustment periods of AP_i . The optimal p_s will be derived in the next section.

8. Let $n = n + 1$, go to step 2.

It is worth pointing out that the algorithm is not trying to search a global optimization solution to Eqns. (22) and (23) because of the complexity of the problem. In the above algorithm, each AP attempts to find and stay at its optimal channel locally. The algorithm is carried out by all APs simultaneously and independently. If the changes in network performance are within the predefined level, there will be no channel adjustment. If the changes result in network performance degradation worse than the predefined level, the AP will automatically adjust its channel to adapt to the changes in real-time.

We note that choosing the β value in Eqn. (24) gives the algorithm a flexibility to accommodate the need of specific network environment. To be more sensitive to the changes in utilization, the algorithm needs to use a relatively small β value, which could result in frequent updating of channel assignment. Therefore, to avoid such frequent updating, a relatively large β value is preferred, which could make the algorithm insensitive to changes in utilization. In our simulations, we find that $\beta = 25\% \sim 50\%$ of the utilization value at normal operation results in a satisfactory balance for slow to moderate changing network environment. However, for general network environment, how to choose an appropriate β value is an operational issue that can be addressed in terms of specific network environment.

The convergence of the algorithm can be proved by using the Theorem 2 in [4], which says that the algorithm does not have infinite looping if $0 < p_s < 1$. In the next section, we

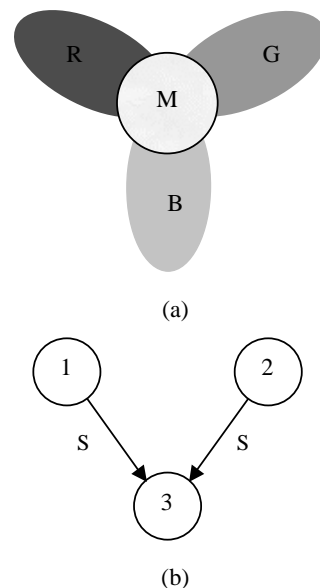


Fig. 2. Channel Allocations in IEEE 802.11b

find that $p_s = 1/2$ if the initial layout of APs is randomly and uniformly distributed. Therefore, the proposed algorithm can reach a local optimal solution with a finite amount of steps for the uniformly distributed APs. See the Appendix for detailed proof of the theorem.

V. OPTIMAL SWITCHING PROBABILITY

A. Solution to Chapman-Kolmogorov Backward Equations

The proposed RRM algorithm can be modeled by Fig. 2 (a), in which an AP_i that has M_i stations will be assigned one of the three channels. Each channel is represented by a color. The state transition process is shown in Fig. 2 (b), in which 1 and 2 are the transient states and 3 is the optimal state that corresponds to the optimal channel assignment. We assume that the transition rate for the transition from each of transient states to the optimal state is the same and denoted by $s \geq 0$. The Q -matrix that describes the transitions has the form

$$Q = \begin{bmatrix} -s & 0 & s \\ 0 & -s & s \\ 0 & 0 & 0 \end{bmatrix}; \quad s \geq 0.$$

In this finite state spaces, the transient probability functions $P_{i,j}(t)$, of the three-state Markov process are determined by solving the Chapman-Kolmogorov backward equations [9]. The result is a system of differential equations that may be written in matrix form as

$$P'(t) = QP(t) = P(t)Q, \quad (25)$$

where $P(t) = \{P_{i,j}(t)\}$, $i, j \in \{1, 2, 3\}$ and Q is the transition rate matrix of the process. Then the solution to the Chapman-Kolmogorov differential equation can be written

$$P(t) = e^{Qt}. \quad (26)$$

The power series can be expanded into matrices operation. In particular

$$e^{Qt} = \sum_{k=0}^{\infty} \frac{(Qt)^k}{k!} = I + Qt + \frac{Q^2 t^2}{2!} + \frac{Q^3 t^3}{3!} + \dots, \quad (27)$$

converges for any square matrix Q , where I is the identity matrix. By substituting the Q -matrix into the above equation, the matrix exponential is found by

$$e^{Qt} = \begin{bmatrix} e^{-st} & 0 & 1 - e^{-st} \\ 0 & e^{-st} & 1 - e^{-st} \\ 0 & 0 & 1 \end{bmatrix}. \quad (28)$$

B. Optimal Switching Probability

As shown in Figure 2 (a), in a channel updating cycle, the AP_i has three choices, to which it can select one as its optimal channel and switch to the channel. If its current state is 1 or 2, the switching probability is $1 - e^{-st}$. The probability to stay in its current state is e^{-st} . If its current state is 3, then it won't switch but stay in state 3. For the two transient states, the switching and staying probability will reach their steady state values after an infinite long time. For the optimal state, the staying action will happen instantly, i. e., without costing any time.

Assume the cycle time to update the channel allocation is T . Let's denote the staying and switching probabilities

$$p_{st}(t) = e^{-st}. \quad (29)$$

$$p_{sw}(t) = 1 - e^{-st}. \quad (30)$$

where the subscripts i for the AP_i is omitted for simplicity.

Initially, at time $t = 0$, $p_{st}(0) = 0$, $p_{sw}(0) = 1$. At time $t = T$, $p_{st}(T) = e^{-sT}$, $p_{sw}(T) = 1 - e^{-sT}$. After that, if a staying event happens for the AP , the transient probability will decay by the factor of e^{-sT} . If a switching event happens for the AP , the transient probability will decay by the factor of $1 - e^{-sT}$. For the state 3, there is no decaying. During these transitions, there may be also some staying actions happened in state 3. Although each time a staying in state 3 will also cost time T , the time is not related the total convergence time that an AP spends until it stabilizes to its optimal channel. The time staying in state 3 is what we hope and does not impact the total convergence time of the channel allocation algorithm. Therefore, we consider the staying in state 3 does not cost any time.

To simplify the notation, let's also define the probability

$$p_s = 1 - e^{-sT}. \quad (31)$$

Note that, p_s is time-invariant and its value is determined by the transition rate and the updating cycle time.

We assume that up to now, there have been n total times of channel transitions, including staying and switching actions happened for the two states. Among them, n_1 times are staying actions, n_2 times are switching actions. Each of the action costs a time of T .

The decaying function for the transient probabilities of the state 1 and 2 can be written as

$$D(s) = (e^{-sT})^{n_1} (1 - e^{-sT})^{n_2}. \quad (32)$$

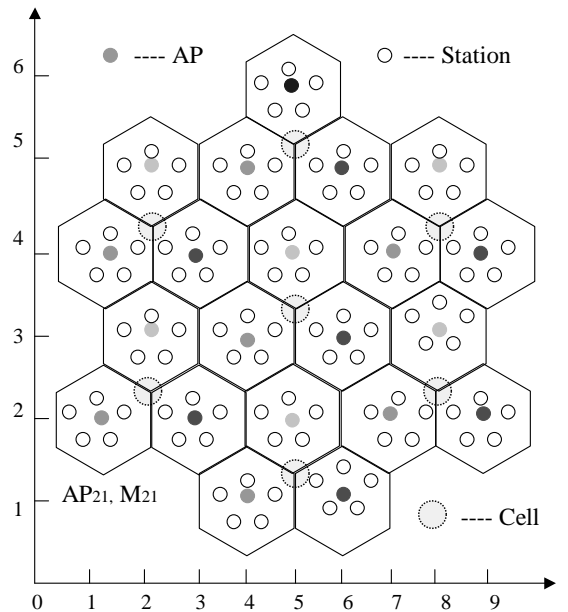


Fig. 3. Network Model

By substituting Eqn. (31), into the above equation, we can find

$$D(s) = (1 - p_s)^{n_1} p_s^{n_2}. \quad (33)$$

By taking $dD(s)/dp_s = 0$, we have

$$p_s = \frac{n_2}{n_1 + n_2}. \quad (34)$$

Let's define a r.v. X that has two possible values, $X = 1$ or $X = 0$, with probabilities q_0 and $1 - q_0$, respectively. Here, $X = 1$ means a switching action and $X = 0$ means a staying action for the AP . Among the n total times of transitions happened, the numbers of the staying and switching actions are $n_1 = nq_0$, and $n_2 = n(1 - q_0)$, respectively. If X has a uniform distribution over $[0, 1]$, the p.d.f. for X is $q_0 = 1/2$. Therefore,

$$p_s = \frac{n(1 - q_0)}{nq_0 + n(1 - q_0)} = 1 - q_0 = \frac{1}{2}. \quad (35)$$

It can be easily verified that $d^2D(s)/d^2p_s < 0$, which means that $D(s)$ is indeed maximized at $p_s = 1/2$, the optimal switching probability. Similarly, for non-uniform distributed APs layout, we can also find an optimal switching probability by using Eqn. (34) for a given p.d.f. of X .

VI. VALIDATION AND SIMULATION RESULTS

A. Network Environment

In this section, we simulate an example network, as shown in Fig. 3. The network has 7 cells, each cell is divided into 3 hexagon coverage areas. Each hexagon area represents the coverage area of an AP, which is located in the center of the hexagon area. The AP is responsible for allocating a channel for the WLAN stations associated with it. In the figure, AP_{ij} represents an AP that is located at (i, j) in the network. The number of stations associated with AP_{ij} is denoted as M_{ij} . We consider two types of initial load for the network. The

first type is $M_{ij} = 25$ for all AP_{ij} , $i = 1, 2, \dots, 6$; $j = 1, 2, \dots, 9$, which is called uniform load. The second type is $M_{ij} = 5 \sim 50$, among which AP_{ij} will be randomly allocated a load, which is called non-uniform load. Totally there are 21 APs in the network. We also simulate a larger size network with the same architecture as the one shown in Fig. 3, which has 37 cells and 115 APs.

The simulation works as follows. First, all the APs and stations run the dynamic IEEE 802.11 protocol [1] in the MAC sublayer. Second, in order to estimate the M value by using our algorithm proposed in Section III, an AP has to collect its operation statistics, such as N_I and E_{idl} , as defined in Eqns. (7) and (9), respectively. Third, the estimated M value will be used by the application layer of the AP to compute its effective channel utilization, as defined in Eqn. (21), and select a frequency channel for the AP. Finally, the AP switches to the selected channel with probability of p_s . Therefore, for the MAC sublayer of each AP and station in the network, we simulate the dynamic IEEE 802.11 protocol without modification, as we analyzed in Section II. The channel allocation algorithm runs in the application layer.

The MAC parameters use the same values as reported in [3]. The related parameters are listed as follows: PHY_{hdr} is the physical header that precedes the transmission of a MAC frame, MAC_{hdr} is the MAC header added to the data payload, $H = PHY_{hdr} + MAC_{hdr}$ is the total overhead added to the data payload. Also, t_B , t_{slot} , t_H , t_{RTS} , t_{CTS} , and t_{ACK} are the time needed to transmit a byte, a slot, the overhead H , the RTS , the CTS , and the ACK message. Also, τ is the maximum propagation delay over the wireless channel.

The values of these parameters are: $\tau = 1 \mu s$, $t_{slot} = 20 \mu s$, $SIFS = 10 \mu s$, $DIFS = 50 \mu s$, $EIFS = 364 \mu s$, $MAC_{hdr} = 14.55 \mu s$, $PHY_{hdr} = 192 \mu s$, $t_{ACK} = 192 \mu s$, $t_{RTS} = 214 \mu s$, $t_{CTS} = 202 \mu s$, $CW_{MIN} = 31 t_{slot}$, and $CW_{MAX} = 1023 t_{slot}$. The traffic associated with each AP is assumed to have a bimodal message-length distribution, where the data payload is 40 bytes long with probability q , and 1500 bytes long with probability $1 - q$, and $q = 0.5$.

B. Simulation Results

In our first example, each AP is allocated the uniform load, to verify the correctness of the proposed RRM algorithm. Starting from a random initial assignment, the proposed algorithm generates an optimal channel assignment map, as shown in Fig. 4, in less than 10 iterations. On an Intel P4 2.8 GHz PC, one iteration costs about 0.1302 seconds of the CPU time. It can be seen that the channel assignment is the classical frequency reuse factor of 3.

As a second example, the APs in the same network are allocated the non-uniform load. To improve the network throughput, some neighboring APs may form a virtual AP that uses the same channel. This has been verified by the second allocation results, as shown in Fig. 5. It can be seen that a heavily loaded AP can use a single channel that is different from those in adjacent APs, while many lightly loaded APs have to share a channel.

As a third example, the same algorithm is applied to the large network with 115 APs. For the uniform load, the

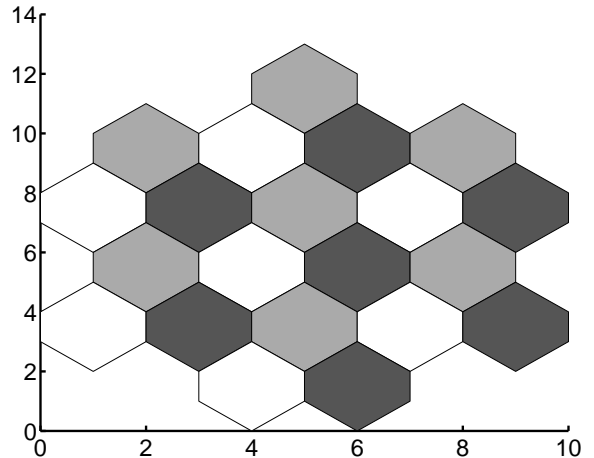


Fig. 4. Channel Allocations For Uniform APs Layout

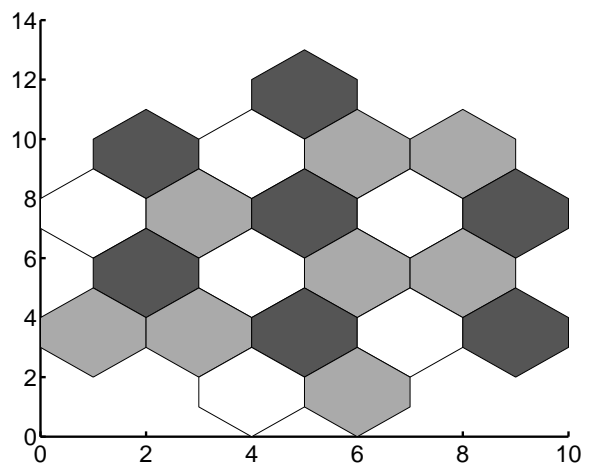


Fig. 5. Channel Allocations For Non-Uniform APs Layout

allocation results are shown in Fig. 6. This time the algorithm is not able to generate the optimal channel assignment of reuse factor of 3. The reason is that the algorithm only searches the optimal solution locally and thus results in a suboptimal solution. It can be seen that most APs can use a channel that is different from those in adjacent APs, while less than 10% of APs have to share a channel with their neighboring APs. In the worst case, at most two neighboring APs share the same channel. For the non-uniform load, a similar result is shown in Fig. 7, in which more lightly loaded APs have to share a channel with their neighboring APs. As a dynamic channel assignment scheme, this suboptimal solution is quite acceptable in practices.

Our fourth example is to demonstrate the effectiveness of real-time estimation algorithm for M values, three different load are used in the simulation: $M = 25$, 50, and 75, respectively. As shown in Fig. 8, the estimated M value approximately converges to the real value within about 200 virtual-transmission times (VTTs). The estimation error falls below 10% of the real value. For the bimodal message-length distribution simulated in this study, the VTT can be computed in terms of Eqn. (5). For $M = 25$, 50, and 75, the VTT equals

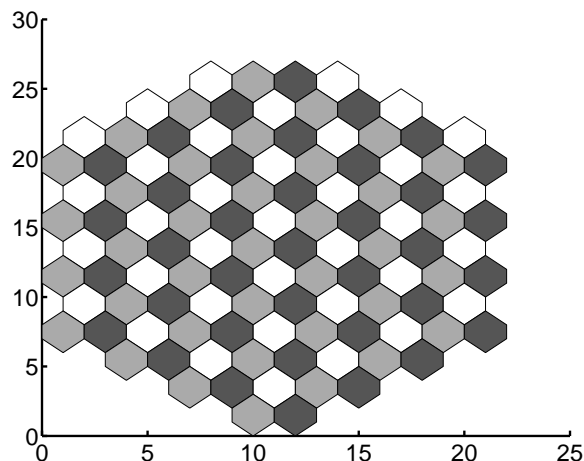


Fig. 6. Channel Allocations For Uniform Large APs Layout

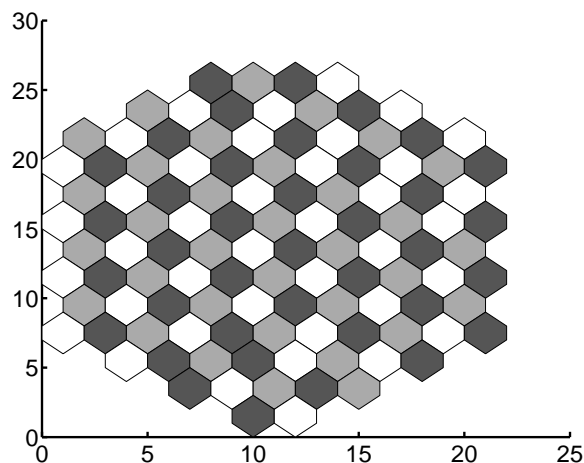


Fig. 7. Channel Allocations For Non-Uniform Large APs Layout

453.3876 μs , 534.8905 μs , and 573.8196 μs , respectively. The estimation algorithm that was used in the adaptive backoff mechanism [1] just could not converge to the real value within an error of 50%, even after 2000 VTTs. Therefore, our algorithm is more efficient.

To investigate the adaptiveness of the M estimation algorithm to a dynamic network environment, we assume that many stations continually join and leave the network. Thus, the M value keeps changing. For the 21-AP-Layout network, we assume that M value changes during a period of 3000 VTTs, from 40 to 65, then to 40 again, after that it goes to 20, 60, and 75. This scenario represents a rapid change in network environment because the interval between changes is only 500 VTTs, which equals about 250 milliseconds. When applying the M estimation algorithm, we choose an initial M value of 100, which is far from the real value. The estimated M values are shown in Fig. 9. Clearly, the proposed algorithm adapts to the changes in M values very well.

It is worth pointing out that the channel assignment algorithm takes a much longer time to reach an optimal solution than the time needed for M estimation algorithm. For example, for the 21-AP-Layout network with uniform load, it takes

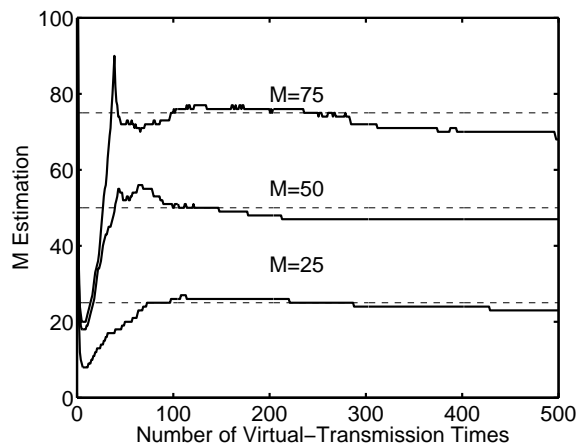


Fig. 8. Estimations of M Values

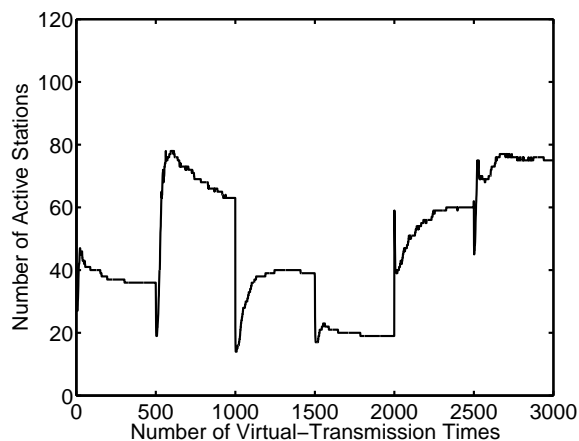


Fig. 9. Estimations of Variable M Values

about 10 iterations or 1.3020 seconds to reach the optimal assignment. If the M value changes again before the channel assignment reaches its optimal solution, we will not see significant throughput improvement, as shown in Fig. 10. In practice, it takes tens of seconds to activate or deactivate the WLAN applications, such as laptop computers. Therefore, it is acceptable for a channel assignment to take a few seconds to reach an optimal solution.

Our fifth example is to investigate the overall network performance improvement after using the real-time M estimation and dynamic RRM algorithms. For the network shown in Fig. 3, each AP associates a fixed number of stations. Initially, we randomly assign each AP a channel. After that, each AP switches to its optimal channel that min-max its throughput with the optimal switching probability of $p_s = 0.5$. Although the throughput of some APs may be reduced, the overall throughput should have been significantly improved. This is confirmed by the curve of normalized throughput, as shown in Fig. 10, which shows nearly a 65% improvement achieved by the proposed algorithms in less than 20 iterations.

Our last example is to verify the optimal switching probability. For the same 21-AP-Layout network, we simulate the channel algorithm with different switching probabilities, i. e., $p_s = 0.1, 0.2, 0.3, \dots$, and 1.0. For each p_s value, the

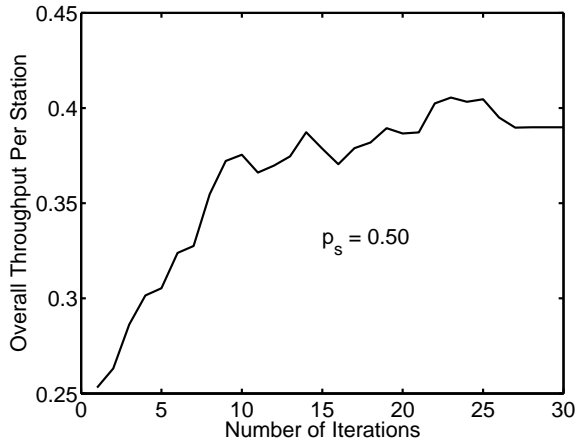


Fig. 10. Throughput Improvement

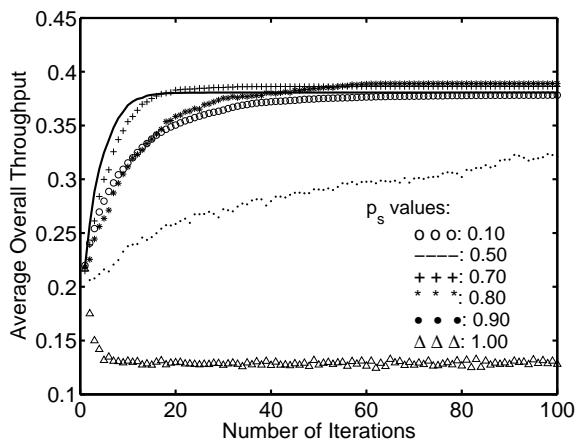


Fig. 11. Optimal Switching Probability

simulation has been conducted 1000 times. Each time, we randomly assign each AP an initial channel. Each AP is randomly allocated a load by a uniformly distributed random number generator. The average results from the 1000 simulation runs are shown in Fig. 11. It can be seen that the solid line, which corresponds to $p_s = 0.5$, has the shortest convergence time, as compared to other p_s values. The simulation results have demonstrated the effectiveness of our finding of the optimal switching probability in Eqn. (35).

VII. DISCUSSION AND CONCLUSIONS

In this paper, we develop a new dynamic frequency assignment technique for the multiple APs in WLANs. Without the need of prior knowledge about network status, each AP can be dynamically assigned to an optimal channel. For the hexagon coverage with uniform AP layout, the optimal assignment matches the classical result of a reuse factor 3. The proposed algorithm not only adapts to the changes in network status but also significantly increases the overall network performance.

In a dynamic environment, a typical issue is how to trade-off between the accuracy and promptness of the estimation algorithm. Another issue is how to tradeoff between the sensitiveness and frequency of the channel assignment algorithm.

As we mentioned before, the proposed algorithms provide the flexibilities to tune the values of the related parameters in operation, in terms of specific application requirements. But it would be better if we can develop a self-learning algorithms.

A third issue is related to the implementation of the dynamic RRM technique. After an AP has chosen a best channel for itself, how to inform the stations that are currently associated with it? One solution is like the hard handoff procedure that has been implemented in the existing mobile telephone systems. A disconnected station has to re-associate its AP again after receives the beacon messages from the AP. Another solution is like the soft handoff procedure. A station will be allocated to another AP before the previous one signs it off. This may need to develop a signaling protocol that could be implemented in the future and more power WLANs.

Our future work would be investigating the above issues and applying the technique to a larger number of channel allocation problems.

ACKNOWLEDGMENT

The authors would like to thank Aniket Malvankar and Nachiket Karmar for their help in doing some of the simulation studies. Also thanks to the editor and reviewers for their constructional comments and suggestions.

APPENDIX

PROOF OF THE DYNAMIC FREQUENCY ASSIGNMENT ALGORITHM

In this appendix, we present the proof for the Theorem 2 in [4], with minor modifications when it is applied to prove the convergence of our frequency assignment algorithm.

Theorem 1: The dynamic frequency assignment algorithm presented in Section IV-B does not have infinite looping if $0 < p_s < 1$.

Proof: Given that the number of APs I and available channels J in the system are finite, Steps 2 to 5 can be completed in a finite amount of time. The only possibility that the algorithm has an infinite loop is that Steps 2 to 6 are executed repeatedly without stop. Assume that such looping can happen and the W value after the n th iteration of Step 6 be denoted by $W(n)$. To proceed, let $p_s = 0$ in Step 7 for a moment. In order to form the infinite looping, we must have $W(1) > W(2) > \dots > W(n)$ with $n \rightarrow \infty$. With both I and J being finite, there are only a finite number of all possible channel assignments. Since each new assignment finalized by Step 6 has a unique maximum effective channel utilization, it is thus impossible that n goes to infinity. That is, Step 7 must be reached after a finite amount of processing.

Let's assume that infinite looping is possible with $0 < p_s < 1$. Based on the above argument, we now must have $W(1) > \dots > W(s) = W(s+1) > \dots > W(t) = W(t+1) > \dots > W(n)$ with $n \rightarrow \infty$ for some s and t . Since the argument above has already ruled out the possibility of having subsequences of W 's of infinite length between two = signs on this list, it must contain an infinite number of = signs. Since each = sign corresponds to an execution of Step 7 with probability p_s , the probability of executing this step for an infinite number of time is thus zero. Hence, the infinite looping cannot exist. ■

REFERENCES

- [1] F. Cali, M. Conti, and E. Gregori, "Dynamic Tuning of the IEEE 802.11 Protocol to Achieve a Theoretical Throughput Limit," *IEEE/ACM Trans. on Networking*, vol. 8, no. 6, pp. 785-799, December, 2000.
- [2] F. Cali, M. Conti, and E. Gregori, "IEEE 802.11 Protocol: Design and Performance Evaluation of an Adaptive Backoff Mechanism," *IEEE Journal of Selected Areas in Communications*, vol. 18, no. 9, pp. 1774-1786, September, 2000.
- [3] R. Bruno, M. Conti, and E. Gregori, "IEEE 802.11 Optimal Performances: RTS/CTS Mechanism vs. Basic Access," *PIMRC*, 2002.
- [4] K. K. Leung and B. Kim, "Frequency Assignment for Multi-Cell IEEE 802.11 Wireless Networks," *VTC'03*, 2003.
- [5] Y. Lee, K. Kim, and Y. Choi, "Optimization of AP Placement and Channel Assignment in Wireless LANs," *Proc. 27th Annual IEEE Conference on Local Computer Networks (LCN'02)*, 2002.
- [6] H. Luo and N. K. Shankaranarayanan, "A Distributed Dynamic Channel Allocation Technique for Throughput Improvement in A Dense WLAN Environment," *Int. Conf. on Acoustics, Speech, and Signal Processing*, May 17-21, 2004, Montreal, Canada.
- [7] M. Yu and H. Luo, "An Adaptive Radio Resource Management Technique for APs in WLANs," *IEEE Int. Conf. on Networks (ICON'04)*, Nov. 2004, Singapore.
- [8] D. Bertsekas and R. Gallager, *Data Networks*, Section 4.4.1, 2nd Edition, Prentice-Hall, Englewood Cliffs, NJ, 1992.
- [9] W. J. Anderson, *Continuous-Time Markov Chains, An Application-Oriented Approach*. Springer-Verlag, New York, 1991.



Ming Yu received his Doctor of Engineering from Tsinghua University, Beijing, in 1994, and Ph.D. from Rutgers University, New Brunswick, NJ, in 2002, all in Electrical and Computer Engineering.

He joined the Operation Technology Center, AT&T, Middletown, NJ, in July 1997 as a Senior Technical Staff Member. Since 1999, he was with the Dept. of ATM Network Service, AT&T Labs, Middletown, NJ. From December 2002, he worked for the Dept. of IP/Data Network Management System Engineering, AT&T Labs. As of August 2003,

he joined the Dept. of Electrical and Computer Engineering, State University of New York (SUNY), Binghamton, NY, as an assistant professor.

His research interests are in the areas of routing protocols, MAC, QoS, security, energy-efficiency, clustering, radio resource management, traffic engineering, and performance analysis, for both wired and wireless networks.

Dr. Yu is a senior member of the IEEE. He was awarded the IEEE Millennium Medal on May 2000.



Kin K. Leung received his B.S. degree from the Chinese University of Hong Kong in 1980, and his M.S. and Ph.D. degrees in computer science from University of California, Los Angeles, in 1982 and 1985, respectively.

He started his career at AT&T Bell Labs in 1986. Following Lucent Technologies spun off from AT&T in 1996, he was with AT&T Labs from 1996 to 2002. In 2002, he re-joined Bell Labs of Lucent Technologies. Since 2004, he has been the Tanaka Chair Professor in Internet Technology at Imperial

College (University of London). His research interests include radio resource allocation, MAC protocol, TCP/IP protocol, mobility management, network architecture, real-time applications and teletraffic issues for broadband wireless networks. He is also interested in a wide variety of wireless technologies, including 802.11, 802.16, and 3G and future generation wireless networks.

He received the Distinguished Member of Technical Staff Award from AT&T Bell Labs in 1994, and was a co-recipient of the 1997 Lanchester Prize Honorable Mention Award. He is an IEEE fellow. He has published widely and acquired patents in many areas of communication networks. He has actively served on conference committees, including as the committee co-chair for the Multiaccess, Mobility and Teletraffic for Wireless Communications (MMT'98) and the committee Vice-Chair for the IEEE ICC 2002. He was a guest editor for the IEEE Journal on Selected Areas in Communications (JSAC), the MONET journal and the Wireless Communications and Mobile Computing journal, and as an editor for the JSAC: Wireless Series. Currently, he is an editor for the IEEE Transactions on Communications and the Transactions on Wireless Communications.

Hui Luo received his Doctor of Engineering from Tsinghua University, Beijing, China, in Electrical Engineering, in 1994.

From 1995 to 1997 he was a postdoctoral research associate with the Department of Electrical Engineering, University of Notre Dame, Indiana, where his research area was blind signal processing. From 1999 to 2003, he was with AT&T Labs, Middletown, New Jersey. Since 2004, he joined Broadcom, Matawan, NJ, as a Senior Staff Scientist.

His research areas include wireless/mobile networking, network security, and signal processing for wireless communications.