Joint Optimal Access Point Selection and Channel Assignment in Wireless Networks

Iordanis Koutsopoulos, *Member, IEEE*, and Leandros Tassiulas, *Senior Member, IEEE*

***Abstract*—In wireless cellular networks or in other networks with single-hop communication, the fundamental access control problem pertains to access point (AP) selection and channel allocation for each user. For users in the coverage area of one AP, this involves only channel allocation. However, users that belong in the intersection of coverage areas of more than one AP can select the appropriate AP to establish connection and implicitly affect the channel assignment procedure. We address the joint problem of AP selection and channel assignment with the objective to satisfy a given user load vector with the minimum number of channels. Our major finding is that the joint problem reduces to plain channel allocation in a cellular network that emerges from the original one after executing an iterative and provably convergent clique load balancing algorithm. For linear cellular networks, our approach leads to minimum number of required channels to serve a given load vector. For 2-D cellular networks, the same approach leads to a heuristic algorithm with a suboptimal solution due to the fact that clique loads cannot be balanced. Numerical results demonstrate the performance benefits of our approach in terms of blocking probability in a dynamic scenario with time-varying number of connection requests. The presented approach constitutes the basis for addressing more composite resource allocation problems in different context.**

***Index Terms*—Access point (AP) assignment, channel allocation, load balancing, wireless access.**

I. Introduction

THE pervasiveness of wireless devices and the increasing dependability of our lifestyle on wireless networks have brought into picture the issue of efficient wireless access. Wireless devices obtain network access through single-hop or multi-hop connections. Examples of single-hop access networks are typical wide-area cellular networks and local-area networks in point coordination function (PCF) operation. In these systems, the interface between the user and the network is a base station (BS) or an access point (AP) that is connected to the backbone network. Instances of access through multi-hop connection are encountered in personal area networks, sensor networks or wireless local-area networks in distributed coordination function (DCF) mode. Furthermore, the emerging wireless mesh networks employ both single- and multi-hop connection access.

Manuscript received March 18, 2003; revised October 7, 2005; approved by IEEE/ACM TRANSACTIONS ON NETWORKING Editor M. Grossglauser. This work was supported in part by the Greek Secretariat of Research and Technology PENED Grant and in part by the European Commission NoE NEWCOM IST-507325. The work of I. Koutsopoulos was supported in part by a Marie Curie Grant SAINT-W IRG-017267. Part of the material in this paper was presented at the International Communications Conference 2000, New Orleans, LA.

The authors are with the Department of Computer and Communications Engineering, University of Thessaly, Volos GR 38221, Greece (e-mail: jordan@uth.gr; leandros@uth.gr).

Digital Object Identifier 10.1109/TNET.2007.893237

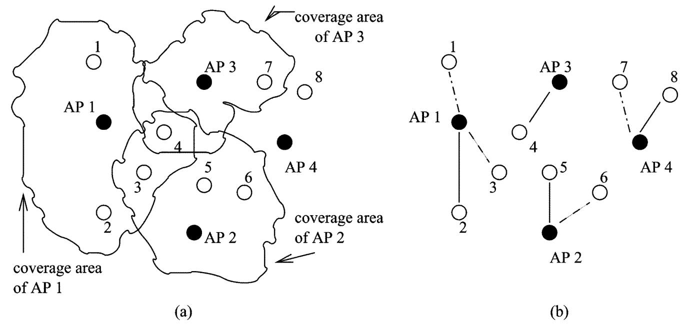

Fig. 1. Stages of single-hop access. (a) Selection of a serving AP. Users can select any AP among those in whose coverage area they belong. (b) Channel assignment for each AP-user link. APs and users are shown with black and white circles, respectively. Co-channel links are shown with the same type of line. For example, the same channel is used for links AP1—User2, AP2—User5, AP3—User4, and AP4—User8. (a) Access point selection problem. (b) Channel allocation.

Whenever single-hop connection of users to an AP or BS is involved, the fundamental problem at the access layer pertains to AP selection for users and channel allocation. Fig. 1 illustrates these two stages. First, users need to select an AP with which they establish connection. Users in the coverage area of more than one AP can choose one among these APs based on different criteria. Next, a channel needs to be allocated to the AP-user link. Depending on the employed multiple access scheme (time, frequency or code division multiple access), the channels can be time slots, frequency bands or codes. The objective of the access methods is to ensure quality of service (QoS) provisioning to users across layers. At the physical layer, QoS is synonymous to an acceptable signal-to-interference-and-noise ratio (SINR) or bit-error rate (BER) at the receiver. At the access layer, QoS can be expressed in terms of a blocking probability or packet error rate bound or as transmission rate guarantees.

Depending on the setup, channel allocation per se can be viewed from different perspectives. For given number of available channels, the objective is to maximize user capacity, namely the number of users that can be accommodated with acceptable link quality. On the other hand, for given number of users, the goal is to minimize the number of required channels so as to accommodate them with acceptable link quality. Then, the system can respond better to potential sudden load increase or link quality deterioration and the probability of blocking is minimized. If users need different number of channels due to different rate requirements, the goals are to maximize total user rate or minimize the required number of channels to satisfy rate requirements. Channel allocation algorithms can be broadly classified as fixed or dynamic (FCA, DCA). In the former, disjoint sets of channels are permanently allocated to cells and a user can only use channels that are assigned to its cell. A new user is admitted if the cell where it resides has a free channel, otherwise it is blocked. In the latter, all channels are kept in a pool and a channel can be used by any user in any cell. Due to scarcity of available spectrum, channels can be reused for AP-user links in different cells if the separation distance exceeds a minimum reuse distance that guarantees acceptable link quality. A new user is admitted if a channel can be assigned to it and the minimum reuse distance constraints are satisfied. A comprehensive survey of channel allocation algorithms can be found in [2].

With APs selected *a priori*, the channel allocation problem reduces to graph multi-coloring. Each vertex of the graph represents a cell. An edge connects two vertices if the corresponding cells must not use common channels. The problem is to assign colors (channels) to each vertex, so that no common channels are assigned to adjacent vertices and the minimum total number of colors is used to satisfy all users. This problem is known to be NP-hard [3]. Since exhaustive search over possible channel allocations to find the optimal solution is impractical for large-scale systems, most of the studies focus on designing efficient heuristics that may provide optimal solutions for simple networks or special cases but are suboptimal in general [4]–[6]. Performance bounds for different classes of dynamic channel assignment algorithms for linear cellular systems are presented in [7].

Packing algorithms represent a class of DCA methods which are based on channel reassignments for existing users so as to accommodate a new one. Depending on allowed reassignments, packing algorithms provide performance bounds for channel allocation methods in terms of number of accommodated users. The ultimate upper bound is given by the maximum packing (MP) policy [8]. MP accepts a new user whenever there exists a channel reassignment of any size such that channel reuse distance constraints are satisfied. Then the load vector with components equal to the number of users at corresponding cells is feasible. MP has the largest space of feasible load vectors over all channel allocation policies and in that sense it implies minimum blocking probability for given number of available channels or equivalently minimum number of required channels to serve a given user load vector. From an implementation point of view, knowledge of the number of occupied channels in each cell suffices for system analysis in MP and no knowledge about individual occupied channels is necessary. Raymond showed that MP can be implemented in polynomial time in linear cellular networks with nonoverlapping cells by performing at most two user rearrangements upon arrival of a new user [9]. However, this method cannot be extended to 2-D cellular networks.

Overlapping coverage areas of cells often arise in practical cellular systems when some users can listen to more than one AP. Overlapping coverage areas allow mobile users to handoff between cells and maintain connectivity. Users in the common coverage area of cells can select the AP with which they establish connection. However, AP selection affects cell populations and channel allocation, since for some AP assignments to users there may exist a channel allocation that satisfies reuse distance constraints, while for others not. Then, the problems of AP selection and channel assignment arise jointly and the objective is the same as in channel allocation. For fixed AP assignment, the joint problem reduces to a channel allocation one, while for fixed channel assignment the problem involves only AP selection. The work in [10] considers the problem of assigning consumers to resource locations with the objective to balance loads among resource locations as much as possible. It shows that there exist assignments that achieve the objective above and they also minimize certain convex functions of location loads. It also relates finding such assignments to solving network flow problems. In [11], the same location assignment problem is first formulated as a static load balancing problem. The solution minimizes a convex cost function of loads and it also minimizes the maximum load over all locations. The dynamic load balancing problem is solved by the least loaded routing policy, in which a new user is assigned to the location with minimum instantaneous load. For locations with finite capacities, the least relatively loaded routing (LRLR) policy minimizes blocking probability in the limit of large arrival rates. In other words, this policy solves the joint AP selection and channel assignment problem for fixed channel assignment to APs.

The BS selection problem was addressed together with up-link power control for a single channel in [12]. The proposed iterative algorithm finds a feasible BS and power assignment if there exists one and minimizes the total transmitted power. In [13], the joint channel, BS and power assignment problem is solved optimally in terms of number of required channels for the case of two BSs, while in [14], a greedy heuristic is provided for the same problem in a network with several BSs and users.

The contribution of this paper is the treatment of the joint AP selection and channel assignment problem in a cellular network with overlapping cells and dynamic channel assignment. This problem is fundamentally different from the one with fixed channel assignment since any channel can be assigned to any user and cell and load balancing between APs is not sufficient. Compared with the classical dynamic channel assignment problem in cellular networks with nonoverlapping cells, the challenge in our problem is that the satisfiability of reuse distance constraints depends on both channel assignment and AP selection. We identify the relation between the two separate problems and we show that the joint problem reduces to plain channel allocation in an equivalent network that arises from the original one after executing an iterative and provably convergent load balancing algorithm. We first apply our method to linear cellular networks with overlapping cells which constitute natural coverage models for certain environments. Our method finds the optimal solution in terms of minimizing the number of channels needed to serve a certain load vector. For 2-D cellular networks the same approach leads to a suboptimal, yet meaningful algorithm.

The rest of the paper is organized as follows. In Section II, we consider linear cellular networks. We propose an iterative load balancing algorithm, prove its convergence and show correspondence of the joint problem to the plain channel allocation one. In Section III, we consider 2-D networks and in Section IV, we describe the algorithm in a dynamic scenario. Section V contains numerical results and Section VI concludes our study.

Throughout the paper, we use the term "AP" to denote the interface of a user with the network. This term encompasses the BS of wide-area cellular networks. We also use interchangeably the terms user, call and connection with the same meaning.

II. Joint AP Selection and Channel Allocation in Linear Cellular Networks

We consider a cellular network with N cells, M mobile users and L orthogonal channels with down-link transmission from APs to users. The same treatment can be applied for the up-link as well. The propagation environment between each AP i and user j is characterized by a time-varying path gain G_{ij} that captures path loss, shadowing and multi-path fading. All channels are available to each AP for dynamic assignment to users. Co-channel interference is assumed to be the prevailing cause of interference, so that the SINR can be replaced by the signal-to-interference ratio (SIR). We first consider linear cellular networks that model coverage in outdoor environments such as city streets or highways and in indoor environments such as local-area networks of special structure. Unless otherwise stated, a snapshot model is used so that locations of users are frozen. In order to illustrate our approach, we assume that one channel is needed for transmission to each user as in a circuit-switched system. However, our approach is also applicable to packet-based systems with single-rate transmission and different channel requirements for each user.

A. Linear Cellular Network With Nonoverlapping Cells

A linear cellular network with N APs is depicted in Fig. 2. Each AP provides coverage to a cell of radius d and coverage areas of adjacent APs do not overlap. The AP coverage area is the set of location points at which a minimum acceptable SIR is ensured. Two or more APs can simultaneously use the same channel to transmit to users in their cells if acceptable SIRs can be ensured for co-channel users. In general, APs can reuse the same channel if they are separated by at least r cells or equivalently they have reuse distance $R = r + 1$ with $R \geq 1$. For $R = 1$, the resulting sum interference model captures the situation where a channel can be used at each AP. We consider a simple propagation model with path loss equal to $1/x^{\alpha}$ for a user at distance x from the AP, where α is a propagation constant. A user has minimum SIR if it is at the boundary of its cell and experiences co-channel interference from all APs that use the same channel subject to reuse distance constraints. Thus, the minimum SIR for a user at cell i is

$$\mathrm{SIR}_0 = \frac{d^{-\alpha} P_t}{\sum_{k=1}^{k_1^*} [(2kR+1)d]^{-\alpha} P_t + \sum_{k=1}^{k_2^*} [(2kR-1)d]^{-\alpha} P_t} \quad (1)$$

where P_t is the AP common transmit power. The sums in the denominator denote interference from cells on the left and right of cell i that use the same channel and the summation limits are $k_1^* = \max\{k : i - kR \geq 0\}$ and $k_2^* = \max\{k : i + kR \leq N\}$. The reuse distance is selected at the design phase so that $\mathrm{SIR}_0 \geq \gamma_0$, where γ_0 is a threshold that depends on the maximum tolerable BER at the receiver and on the employed modulation and coding rate. In reality, SIR fluctuates due to time-varying fading. Thus, a more meaningful requirement for acceptable link quality is that the instantaneous SIR should exceed a threshold γ for more than a fraction τ of time. Link quality requirements $\mathrm{SIR} \geq \gamma$ and $\mathrm{SIR}_0 \geq \gamma_0$ are equivalent if γ_0, γ and τ are appropriately selected.

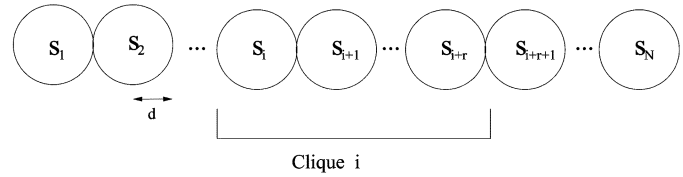

Fig. 2. Linear cellular network with nonoverlapping cells. The number of users in cell i is denoted by s_i, for $i = 1, \ldots, N$.

The occupancy of the system is described by the AP load vector $\mathbf{s} = (s_1, s_2, \ldots, s_N)$, where s_i is the number of active users in cell i. Two cells interfere with each other if they cannot use the same channel simultaneously. A *clique* is a set of cells that all interfere with each other. Clearly, different channels must be assigned to users in cells that belong to the same clique. There exist $N - r$ cliques in a linear network of N cells and reuse distance $r+1$ and clique i consists of cells $i, \ldots, i+r$. The set of cells of clique i is denoted as $\mathcal{C}_i$ and the set of all cliques is Ω. The clique load vector is $\mathbf{Q} = (Q_1, \ldots, Q_{N-r})$, where Q_i is the load of clique i.

In MP, a new user is admitted whenever there exists an assignment of channels to existing users and the new one such that channel reuse constraints are satisfied regardless of the number of required channel reassignments for existing users. An AP load vector $\mathbf{s}$ is *feasible* under MP if and only if there exists an assignment of s_i channels to cell i, $i = 1, \ldots, N$, which respects the channel reuse constraints. Define as $\mathcal{S}_{\mathrm{MP}}$ the set of feasible load vectors under MP. Consider the clique packing (CP) allocation that is defined by the set of feasible load vectors

$$\mathcal{S}_{\mathrm{CP}} = \left\{ \mathbf{s} : \sum_{j \in \mathcal{C}_i} s_j \leq L \text{ for all cliques } i \in \Omega \right\} \quad (2)$$

so that $\mathcal{S}_{\mathrm{MP}} \subseteq \mathcal{S}_{\mathrm{CP}}$. In CP, a new connection request at cell j is admitted if $\mathbf{s} + \mathbf{e}_j \in \mathcal{S}_{\mathrm{CP}}$, where $\mathbf{s}$ is the instantaneous load vector just before the new request and $\mathbf{e}_j$ is a vector of length N whose jth entry is 1 and all other entries are 0. For load vector $\mathbf{s}$ and the MP policy, let $L_{\mathrm{MP}}(\mathbf{s})$ be the minimum number of channels needed for all users so that reuse constraints are satisfied. Since a different channel must be used for each user in cells of the same clique, $L_{\mathrm{MP}}(\mathbf{s})$ must satisfy

$$L_{\mathrm{MP}}(\mathbf{s}) = \max_{i=1,\ldots,N-r} \left\{ \sum_{m=0}^{r} s_{i+m} \right\}. \quad (3)$$

Thus, the minimum number of required channels equals the maximum clique load. This lower bound on required channels over all DCA algorithms is achieved by MP in polynomial time in linear cellular networks. The minimum number of required channels under CP also equals the right-hand side of (3). In linear networks, it is $\mathcal{S}_{\mathrm{CP}} = \mathcal{S}_{\mathrm{MP}}$, but in general networks CP violates the reuse constraints and therefore has no practical merit. However, if the probability that an AP load vector belongs in $(\mathcal{S}_{\mathrm{CP}} - \mathcal{S}_{\mathrm{MP}})$ is sufficiently small, CP can serve as a good approximation to MP.

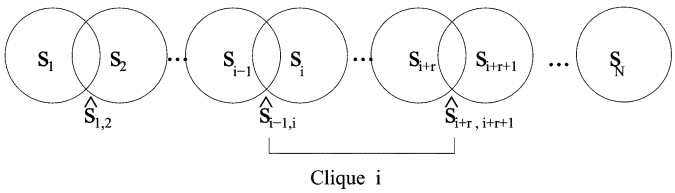

Fig. 3. Linear cellular network with overlapping cells.

B. Linear Cellular Network With Overlapping Cells

A linear cellular network of N cells with overlapping coverage areas is shown in Fig. 3. The occupancy of the system is described by the *nonoverlap area load vector* $\mathbf{s} = (s_1, \ldots, s_N)$, where s_i is the number of active users in the nonoverlap region of cell i, and the *overlap area load vector* $\hat{\mathbf{s}} = (\hat{s}_{1,2}, \ldots, \hat{s}_{N-1,N})$, where $\hat{s}_{i,i+1}$ is the number of active users in the overlap area of cells i and $i+1$. Users in nonoverlap and overlap areas are referred to as type-1 and type-2 users, respectively. Each AP broadcasts a beacon signal in a different control channel and each user measures the signal strength of these signals. Type-1 users establish connection with the corresponding AP and AP selection is not an issue. However, type-2 users can select one among the APs they can listen to in order to establish connection.

A type-2 user does not necessarily "lock" to the AP with the strongest received signal. In fact, existing AP selection policies, also termed as association policies rely solely on the received signal strength. Instead, in our model a type-2 user selects an AP such that some load is transferred from overloaded APs to adjacent less loaded ones and the resulting AP load vector can be satisfied with fewer channels. With the option of AP selection by type-2 users, some channels in APs become free and can be used by type-1 users that would be blocked otherwise. Clearly, clique loads depend on AP selection for type-2 users. For each overlap area i between cells i and $i+1$ with $\hat{s}_{i,i+1}$ users, let $k_{i,i}$ and $k_{i,i+1}$ be the number of users (out of the $\hat{s}_{i,i+1}$ ones) that are assigned to the left and right AP of that overlap area, namely to AP i and $i+1$, respectively.

The class of *admissible* integer-valued AP allocations $\mathcal{F}$ includes the set of assignments α for which $k_{i,i}(\alpha)$, $k_{i,i+1}(\alpha)$ are integer numbers that satisfy $k_{i,i}(\alpha) + k_{i,i+1}(\alpha) = \hat{s}_{i,i+1}$, for $i = 1, 2, \ldots, N-r$. The AP load vector for allocation α is $\boldsymbol{\ell}(\alpha) = [\ell_1(\alpha), \ell_2(\alpha), \ldots, \ell_N(\alpha)]$, where $\ell_i(\alpha) = s_i + k_{i-1,i}(\alpha) + k_{i,i}(\alpha)$ for $i = 2, \ldots, N-1$. Note that the loads of the first and last AP are $\ell_1(\alpha) = s_1 + k_{1,1}(\alpha)$ and $\ell_N(\alpha) = s_N + k_{N-1,N}(\alpha)$. Given an AP allocation, the AP load vector is known and a channel allocation algorithm such as MP can be applied as in the case of nonoverlapping cells. We now give the following definitions.

- An AP load vector $\boldsymbol{\ell}$ is *feasible* if there exists an allocation of ℓ_i channels to each cell i, for $i = 1, 2, \ldots, N$ which satisfies the channel reuse distance constraints.
- A system load vector $(\mathbf{s}, \hat{\mathbf{s}})$ is *feasible* if there exists an admissible AP allocation $\alpha \in \mathcal{F}$ such that the AP load vector $\boldsymbol{\ell}(\alpha)$ is feasible.

Clearly, AP selection and channel assignment are interdependent problems since for certain AP allocations there may exist channel allocations that satisfy channel reuse constraints, while for others not. The joint AP selection and channel assignment problem is stated as:

Problem (P): Given a system load vector $(\mathbf{s}, \hat{\mathbf{s}})$, allocate APs and channels to users so that the total number of required channels is minimized.

In a linear cellular network of overlapping cells, the minimum number of required channels is determined by the maximum clique load. The load of each clique i depends on AP selection and consists of two components.

- A fixed load Q_i^0 that includes type-1 users in cells $i, \ldots, i+r$ of the clique and type-2 users in overlap areas between adjacent cells of the clique, namely

$$Q_i^0 = \sum_{m=0}^{r} s_{i+m} + \sum_{m=0}^{r-1} \hat{s}_{i+m,i+m+1}. \tag{4}$$

- A load $Q_i(\alpha)$ that depends on the AP allocation α. This load consists of: (i) $k_{i-1,i}(\alpha)$ users from the overlap area between cells $i-1$ and i that are assigned to AP i and (ii) $k_{i+r,i+r}(\alpha)$ users from the overlap area between cell $i+r$ and $i+r+1$ that are assigned to AP $i+r$. Thus $Q_i(\alpha) = k_{i-1,i}(\alpha) + k_{i+r,i+r}(\alpha)$, for $i = 2, \ldots, N-r-1$.

Problem (P) is equivalent to that of identifying an admissible AP allocation that minimizes the maximum clique load. This can be formulated as

$$\min_{\alpha \in \mathcal{F}} \max_{i=1,\ldots,N-r} Q_i^0 + Q_i(\alpha) \tag{5}$$

subject to

$$k_{i,i}(\alpha) + k_{i,i+1}(\alpha) = \hat{s}_{i,i+1}, \text{ for } i = 1, \ldots, N-r. \tag{6}$$

In the special case of a linear network with reuse distance $R = 2$, the problem above becomes

$$\min_{\alpha \in \mathcal{F}} \max_{i=1,\ldots,N-1} \left(s_i + s_{i+1} + \hat{s}_{i,i+1} + k_{i-1,i}(\alpha) + k_{i+1,i+1}(\alpha) \right) \tag{7}$$

subject to $k_{i,i}(\alpha) + k_{i,i+1}(\alpha) = \hat{s}_{i,i+1}$ for $i = 1, \ldots, N-1$.

C. Joint AP Selection and Channel Assignment Algorithm

We now introduce an iterative algorithm for solving the joint AP selection and channel assignment problem. The algorithm is referred to as sequential clique load balancing (SCLB). The key idea is that AP allocation at each step should attempt to balance clique loads as much as possible. The algorithm is described here for a linear network with reuse distance $R = 2$ but it can be extended to other reuse distances as well. We consider a snapshot model of the system where the instantaneous load vector $(\mathbf{s}, \hat{\mathbf{s}})$ is assumed fixed. For notational simplicity, we define $f_{i-1,i}(\alpha) = k_{i-1,i}(\alpha)$ and $f_{i+r,i}(\alpha) = k_{i+r,i+r}(\alpha)$ as the number of users in overlap areas $i-1$ and $i+r$ that are assigned to clique i with assignment α. Fig. 4 shows the resulting clique loading from AP assignments for $R = 2$.

We start with a random initial AP selection for each type-2 user. This does not affect the outcome of the algorithm as will be shown in the sequel. The algorithm consists of a finite number

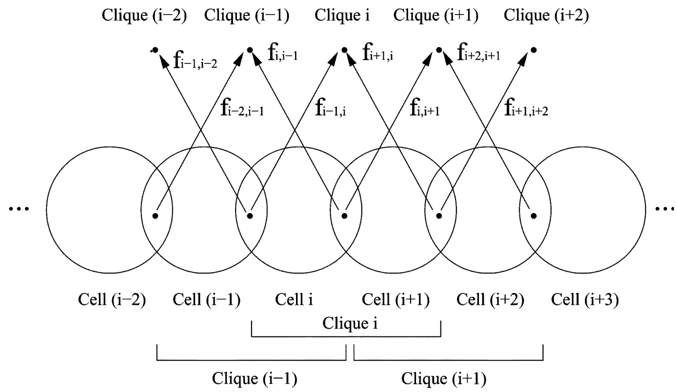

Fig. 4. Clique loading from AP selection for type-2 users in a linear cellular network with $R = 2$.

of iterations, each of which has two phases: the previous assignment cancelation (PAC) and new assignment (NA) one. Overlap areas are considered one at a time. In the PAC phase of the nth iteration for an overlap area, the AP assignments from the $(n-1)$th iteration are canceled. Next, in the NA phase, type-2 users of the overlap area are allocated to one of the two adjacent APs so that the loads of the two affected cliques are as balanced as possible after the assignment. The nth iteration finishes when PAC and NA take place for all overlap areas. The algorithm terminates when clique loads remain unchanged after two subsequent iterations.

Let $f_{i+1,i}^{(n)}$ and $f_{i-1,i}^{(n)}$ be the number of type-2 users that are assigned to clique i from overlap areas $i+1$ and $i-1$ at the nth iteration. Let $Q_i^{(n)}$ be the load of clique i at the end of the nth iteration. After the NA phase of the nth iteration, it is $Q_i^{(n)} = Q_i^0 + f_{i+1,i}^{(n)} + f_{i-1,i}^{(n)}$ with $f_{i,i-1}^{(n)} + f_{i,i+1}^{(n)} = \hat{s}_{i,i+1}$. The SCLB algorithm can be summarized as follows.

Step 0) Set $n = 0$. For each overlap region i, assign each of the $\hat{s}_{i,i+1}$ users randomly to the left or right AP to load cliques $i-1$ and $i+1$. The $\hat{s}_{1,2}$ and $\hat{s}_{N-1,N}$ users in the first and last overlap areas load only clique 2 and clique $N-2$, respectively. For each overlap area i, perform steps 1–3.

Step 1) (PAC) Set $n \leftarrow n+1$. For overlap area i, cancel the assignment of the $(n-1)$th iteration by modifying clique loads as follows:

$$Q_{i-1} \leftarrow Q_{i-1}^{(n-1)} - f_{i,i-1}^{(n-1)} \text{ and } Q_{i+1} \leftarrow Q_{i+1}^{(n-1)} - f_{i,i+1}^{(n-1)}.$$

Step 2) (NA) Consider cliques $i-1$ and $i+1$ and assume that $Q_{i-1} < Q_{i+1}$. Assign users of overlap area i to clique $i-1$ (AP i) until the two loads are equal or until all $\hat{s}_{i,i+1}$ users are assigned. That is, assign $\min\{\hat{s}_{i,i+1}, Q_{i+1} - Q_{i-1}\}$ users to clique $i-1$.

Step 3) If $\hat{s}_{i,i+1} > Q_{i+1} - Q_{i-1}$, there exist $\hat{s}_{i,i+1} - (Q_{i+1} - Q_{i-1})$ unassigned type-2 users. Assign them interchangeably to cliques $i-1$ and $i+1$ one by one so as to maintain load balance.

Step 4) Compare clique load vector $\mathbf{Q}^{(n)}$ to $\mathbf{Q}^{(n-1)}$. If $\mathbf{Q}^{(n)} = \mathbf{Q}^{(n-1)}$, terminate the algorithm, else go to step 1 and repeat the algorithm.

Note that steps 2 and 3 need to be appropriately modified in the case that $Q_{i-1} > Q_{i+1}$.

D. Main Result

Consider the resulting clique vector after the SCLB algorithm. A clique load vector is *feasible* if its maximum entry does not exceed the number of available channels L. After SCLB, an equivalent linear cellular network with nonoverlapping cells emerges, and the resulting AP load vector depends on AP selections of type-2 users. The next task is to allocate channels to users in this equivalent linear network subject to channel reuse distance constraints. The relation between the initial joint AP selection and channel assignment problem and the plain channel allocation one is illustrated in the following theorem.

Theorem 1: A system load vector $(\mathbf{s}, \hat{\mathbf{s}})$ is feasible if the maximum entry of the clique load vector in the equivalent linear network after the SCLB algorithm does not exceed the available number of channels.

Proof: The proof is given here for $R = 2$. Let i_0 be the clique with maximum load after the SCLB algorithm and let $Q_{i_0} = s_{i_0} + s_{i_0+1} \le L$. Without loss of generality assume that i_0 and N are even numbers. Let A and B be the set of even and odd indexed cells and define conditions $C1$ and $C2$ as follows:

$$C1: s_{i_0} = \max_{i \in A} s_i, \quad C2: s_{i_0+1} = \max_{i \in B} s_i.$$

Note that at least one of C1, C2 is true, otherwise the clique with the maximum load is other than i_0. Thus, either both C1 and C2 hold or one of them holds. If both C1 and C2 hold, some (or all) of the s_{i_0} channels of cell i_0 can be reused in each of the cells in A and some (or all) of the s_{i_0+1} channels of cell i_0+1 can be reused in each of the cells in B.

If one of C1 or C2 is true but not both, channel allocation is more composite. Assume for instance that C2 is true. Then, some (or all) of the s_{i_0+1} channels of cell i_0+1 can be reused in each of the cells in B. Since C1 is not true, there exists an even-indexed cell m such that $s_m > s_{i_0}$. Since i_0 is the clique with the maximum load, the loads of neighboring cells of cell m should satisfy $s_{m-1} < s_{i_0+1}$ and $s_{m+1} < s_{i_0+1}$, otherwise the maximum clique would be $m-1$ or m, respectively. Since the load of clique i_0 exceeds that of cliques $m-1$ and m, we have $s_{i_0+1} - s_{m-1} > s_m - s_{i_0}$ and $s_{i_0+1} - s_{m+1} > s_m - s_{i_0}$. Hence, $\min\{s_{i_0+1} - s_{m+1}, s_{i_0+1} - s_{m-1}\}$ channels are unused and they can be used for $s_m - s_{i_0}$ calls of cell m. The rest s_{i_0} users in cell m are served by reusing channels that serve the s_{i_0} users of cell i_0.

Since the clique load vector after the SCLB algorithm is feasible, the initial system load vector is feasible as well. Thus, there exists a solution for the joint AP selection and channel assignment problem that satisfies channel reuse distance constraints. ■

Therefore, solving the joint problem in a linear network with overlapping cells is the same as solving a plain channel allocation problem in an equivalent linear network with nonoverlapping cells, where the AP loads are determined by the SCLB algorithm. For the network of overlapping cells, the minimum number of required channels is given by the right-hand side of (3) with the AP loads s_i replaced by the AP loads after the SCLB algorithm. The minimum number of channels is achieved by MP in the equivalent linear network of nonoverlapping cells.

Examples: As an example when both C1 and C2 are true, consider a linear network of four cells with load vector $\mathbf{s} = (15, 16, 17, 18)$ after SCLB. The number of required channels equals the maximum clique load $Q_3 = 35$. Users in cell 3 are served by 17 channels and users in cell 4 are served by 18 channels. For users in cell 1, 15 out of 17 channels of cell 3 can be reused, while for users in cell 2, 16 out of 18 channels of cell 4 are reused.

As an example of channel allocation when one of C1, C2 is true, consider a system with load vector $\mathbf{s} = (17, 18, 16, 17, 18)$ after SCLB. The maximum clique load is $Q_4 = 35$ and thus 35 channels are needed. Users in cell 4 are served with 17 channels and users in cell 5 are served with 18 channels. For users in cell 3, we reuse 16 out of 18 channels that are used in cell 5 and for users in cell 1 we reuse 17 out of these 18 channels. For the 18 users in cell 2, we first reuse the 17 channels that are used in cell 4. For the remaining one user, we use that channel among the 18 channels of cell 5 that has not been used in cells 3 and 1.

E. Convergence of SCLB Algorithm

We prove the convergence of the iterative SCLB algorithm for the class of continuous (real-valued) AP assignments $\mathcal{F}_R$ which is a superset of integer-valued assignment class $\mathcal{F}$. The system load $(\mathbf{s}, \hat{\mathbf{s}})$ is treated as a deterministic divisible fluid. With real-valued AP assignments, perfect load balancing is achieved for two cliques whenever this is possible in a step of the SCLB algorithm. The fluid model captures admissibility of AP assignments but does not take into consideration channel reuse constraints. We will show that the resulting AP assignment at the end of SCLB algorithm minimizes the maximum clique load over all admissible AP assignments.

A formal proof of convergence for integral AP allocations is cumbersome since perfect load balancing may not be feasible. For instance, this occurs if the number of users in an overlap area is odd and the load difference between the two affected cliques is even at some iteration. Even if perfect balancing is achieved in one step, the balance may be canceled in next iterations. If the SCLB algorithm with continuous assignments converges to the minmax clique load, the algorithm with integral assignments is expected to have similar properties. The problem with continuous assignments is more amenable to analysis and this is the reason it is considered here.

Let $\hat{\mathcal{S}}$ denote the set of all overlap areas and let Ω be the set of cliques. Let Φ be a strictly convex, real-valued function. Define a function $G(\cdot)$ of AP assignments α, that is, $G : \mathcal{F}_R \rightarrow \mathbb{R}_+$ with

$$G(\alpha) = \sum_{i \in \Omega} \Phi\left(Q_i^0 + Q_i(\alpha)\right) \tag{8}$$

and consider the following problem:

Problem (P1): Minimize $G(\alpha)$ over all admissible AP assignments $\alpha \in \mathcal{F}_R$.

For each admissible assignment $\alpha \in \mathcal{F}_R$ and each overlap area $i \in \hat{\mathcal{S}}$ with $\hat{s}_{i,i+1}$ users, we define a new assignment $T_i\alpha$, where T_i is an operator acting on α. Assignment $T_i\alpha$ stems from α if we minimize $G(\alpha)$ with respect to pair $(f_{i,i-1}, f_{i,i+1})$ subject to $f_{i,i-1} + f_{i,i+1} = \hat{s}_{i,i+1}$, while keeping all other assignments fixed. Namely, $f_{k,j}$ for $k \in \hat{\mathcal{S}}$, $k \neq i$ and $j \in \Omega$ are fixed. Assignment $T_i\alpha$ affects cliques $i-1$ and $i+1$. The loads of these cliques are $Q_{i-1} = Q_{i-1}^0 + f_{i,i-1} + f_{i-2,i-1}$ and $Q_{i+1} = Q_{i+1}^0 + f_{i,i+1} + f_{i+2,i+1}$.

If $\Phi(x) = x^2$, the minimum value of $G(\alpha)$ with respect to $(f_{i,i-1}, f_{i,i+1})$ is achieved by

$$f_{i,i-1} = \frac{1}{2}\left\{\hat{s}_{i,i+1} - \left[\left(Q_{i-1}^0 + f_{i-2,i-1}\right) - \left(Q_{i+1}^0 + f_{i+2,i+1}\right)\right]\right\}$$

$$f_{i,i+1} = \frac{1}{2}\left\{\hat{s}_{i,i+1} - \left[\left(Q_{i+1}^0 + f_{i+2,i+1}\right) - \left(Q_{i-1}^0 + f_{i-2,i-1}\right)\right]\right\}.$$

In the above, $f_{i,i-1}$ and $f_{i,i+1}$ denote the amount of assigned fluid load from overlap area i to cliques $i-1$ and $i+1$ in a step of SCLB if the load difference of cliques $i+1$ and $i-1$ is less than $\hat{s}_{i,i+1}$. If the load difference of cliques $i-1$ and $i+1$ exceeds $\hat{s}_{i,i+1}$, then $f_{i,i-1} = 0$ and $f_{i,i+1} = \hat{s}_{i,i+1}$ or $f_{i,i+1} = 0$ and $f_{i,i-1} = \hat{s}_{i,i+1}$, depending on the largest clique load.

Let $\alpha^{(0)}$ be an initial assignment and let $\{i_k : k \geq 1\}$ be a sequence of indices of overlap areas with $i_k \in \hat{\mathcal{S}}$ for all k. We assume there exists an integer Δ such that for any integer ℓ and any $i' \in \hat{\mathcal{S}}$, it is $i_k = i'$ for some k with $\ell \leq k \leq \ell + \Delta$. In other words, the sequence should visit all overlap areas within a period of Δ iterations. Define the sequence of assignments $\{\alpha^{(k)}, k \geq 0\}$ recursively as $\alpha^{(k+1)} = T_{i_k}\alpha^{(k)}$. In our proofs we apply the rationale presented in [10] with all necessary modifications. We first prove the following lemma.

Lemma 1: For any admissible assignment $\alpha \in \mathcal{F}_R$ and $i \in \hat{\mathcal{S}}$, define $|\alpha - T_i\alpha| = \max_{j \in \Omega} |f_{i,j}(\alpha) - f_{i,j}(T_i\alpha)|$. Then, we have

$$|\alpha - T_i\alpha|^2 \leq G(\alpha) - G(T_i\alpha). \tag{9}$$

Proof: Let $(f_{i,i-1}, f_{i,i+1})$ and $(\tilde{f}_{i,i-1}, \tilde{f}_{i,i+1})$ be admissible assignments based on α and $T_i\alpha$, respectively. Let Q_j and $\tilde{Q}_j$, for $j \in \{i-1, i+1\}$ be the clique loads corresponding to these assignments. Namely

$$\begin{aligned} Q_{i-1} &= Q_{i-1}^0 + f_{i,i-1} + f_{i-2,i-1} \\ Q_{i+1} &= Q_{i+1}^0 + f_{i,i+1} + f_{i+2,i+1}, \\ \tilde{Q}_{i-1} &= Q_{i-1}^0 + \tilde{f}_{i,i-1} + f_{i-2,i-1} \\ \tilde{Q}_{i+1} &= Q_{i+1}^0 + \tilde{f}_{i,i+1} + f_{i+2,i+1}. \end{aligned}$$

Since $f_{i,i-1} + f_{i,i+1} = \tilde{f}_{i,i-1} + \tilde{f}_{i,i+1} = \hat{s}_{i,i+1}$, we obtain

$$(Q_{i-1} - \tilde{Q}_{i-1}) + (Q_{i+1} - \tilde{Q}_{i+1}) = 0. \tag{10}$$

Define $\xi = \min\{\tilde{Q}_{i-1}, \tilde{Q}_{i+1}\}$. Since the AP assignment for the $\hat{s}_{i,i+1}$ users affects only cliques $i-1$ and $i+1$ we get

$$\begin{aligned} G(\alpha) - G(T_i\alpha) &= \sum_{j\in\{i-1,i+1\}} \left(Q_j^2 - \tilde{Q}_j^2\right) \\ &\overset{(10)}{=} \sum_{j\in\{i-1,i+1\}} \left[Q_j^2 - \tilde{Q}_j^2 - 2\xi(Q_j - \tilde{Q}_j)\right] \\ &= \sum_{j\in\{i-1,i+1\}} \left[(Q_j-\xi)^2 - (\tilde{Q}_j - \xi)^2\right] \\ &\geq \sum_{j\in\{i-1,i+1\}} (Q_j - \tilde{Q}_j)^2 \\ &\geq \max_{j\in\{i-1,i+1\}} |Q_j - \tilde{Q}_j|^2 \\ &= \max_{j\in\{i-1,i+1\}} |f_{i,j} - \tilde{f}_{i,j}|^2 \\ &= |\alpha - T_i\alpha|^2. \end{aligned}$$

The first inequality in the derivation above follows from the fact that $-\xi \geq -\tilde{Q}_{i-1}$ and $-\xi \geq -\tilde{Q}_{i+1}$. ■

The following theorem shows the convergence of the iterative procedure.

Theorem 2: The quantity $G(\alpha^{(k)})$ is monotone nonincreasing with k and converges to the optimal solution of Problem (P1). Any limit point of the sequence $\alpha^{(k)}$ is a solution to P1.

Proof: The monotonicity of $G(\alpha^{(k)})$ is clear since $G(\alpha) - G(T_i\alpha) \geq 0$ from Lemma 1. Since $G(\cdot)$ is a nonnegative and continuous function, we have that $\lim_{k\to\infty}\left[G(\alpha^{(k+1)}) - G(\alpha^{(k)})\right] = 0$. In addition, we have $\lim_{k\to\infty} G(\alpha^{(k)}) = G(\alpha^*)$ for any limit point α^* of $\alpha^{(k)}$ and $k \geq 0$. From Lemma 1, it follows that $\lim_{k\to\infty}|\alpha^{(k+1)} - \alpha^{(k)}| = 0$. Since mapping T_i is continuous for $i \in \hat{\mathcal{S}}$, we have $T_i\alpha^* = \alpha^*$ for any i and any limit point α^*. Thus, α^* is a solution to Problem (P1). ■

We now need the following statement which is also included in [10, Corollary 4].

1) Fact: If α^* is a solution to Problem (P1) for a strictly convex function Φ, then it is a solution to (P1) for all convex functions Φ.

Next, we proceed to the main theorem for clique load balancing.

Theorem 3: If α^* is a solution to Problem (P1), then α^* minimizes the maximum clique load over all admissible continuous assignments. That is

$$Q_{\max}(\alpha^*) = \min_{\alpha\in\mathcal{F}_R} \max_{i\in\Omega} (Q_i^0 + Q_i(\alpha)). \tag{11}$$

Proof: Consider the convex function $\Phi(x) = (x-w)^+$, where w is a fixed real number and $y^+ = y$ if $y > 0$ and 0 otherwise. Let $G_w(\alpha)$ equal $G(\alpha)$ in (8) for that $\Phi(x)$. From fact 1, α^* also minimizes $G_w(\alpha)$ over all admissible AP assignments α. Since $G_w(\alpha) = 0$ if and only if $\max_{i\in\Omega}\left(Q_i^0 + Q_i(\alpha)\right) \leq w$, the result follows readily. ■

The resulting real-valued AP assignment satisfies the admissibility constraints. An integral assignment can be obtained by rounding up or down each AP assignment to the closest integer. Let α^* and Q_i denote the optimal solution of problem (P1) and the load of clique i at the end of the real-valued assignment algorithm. Let Q_i' and α' be the rounded values for the load of clique i and the corresponding solution. Then $Q_i' \in \{[Q_i]-1, [Q_i], [Q_i]+1\}$ and $|\alpha' - \alpha^*| \leq 1$, where $[\cdot]$ denotes integer part.

Alternatively, we can view our system as a consumer-resource network and use its correspondence to a flow network that was stated in [10]. Our system is depicted as a bipartite network graph $U \cup V$ as suggested by Fig. 4. Node set U consists of one node for each overlap area and node set V has one node for each clique. The flow corresponding to an assignment α is a vector $\mathbf{f}(\alpha) = (f_{i,j}(\alpha) : i \in \hat{\mathcal{S}}$ and $j \in \Omega)$. The capacity of link (i,j) equals the number of type-2 users at overlap area i. In [10, Thm.6] it is proved that the maximum flow algorithm provides a means for finding an integral solution to problem (P1). However, the proposed iterative SCLB algorithm involves low coordination overhead and is more amenable to a distributed implementation.

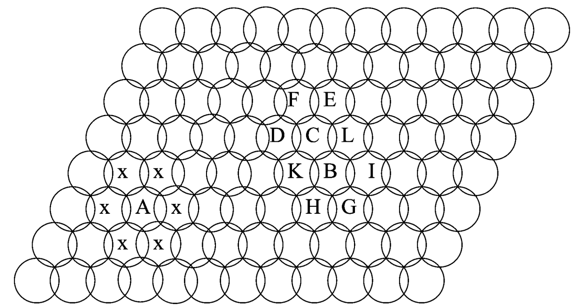

Fig. 5. 2-D cellular network with overlapping cells.

III. The Problem in 2-D Cellular Networks

In this section we study the joint AP selection and channel allocation problem in 2-D cellular networks such as the one depicted in Fig. 5. For easiness of exposition, we assume that cells are sufficiently far from each other and we only consider overlap regions between two APs and not among three APs. Each clique consists of several cells. For $R = 2$, the maximum clique consists of 7 cells, namely a central cell and 6 surrounding cells. Each clique is identified by the index of its central cell. Thus, clique A consists of cell A and 6 surrounding cells marked with "X." In the sequel, we use the term clique to refer to the maximum clique. The set of cells of clique j is denoted as $\mathcal{C}_j$ and the set of 12 cells that are neighbors to cells in $\mathcal{C}_j$ is denoted as $\mathcal{N}_j$. Let $\hat{\mathcal{S}}$ be the set of cell overlap regions, where overlap region i is identified by the pair (κ_i, ℓ_i) of overlapping cells. The occupancy of a system with N cells is described by the AP nonoverlap area load vector $\mathbf{s} = (s_1, \ldots, s_N)$, where s_i is the number of active users in the nonoverlap region of cell i, and the overlap area load vector $\hat{\mathbf{s}} = (\hat{s}_{\kappa,\ell} : \kappa, \ell \in \{1, \ldots, N\}$ and k, ℓ overlap).

For a 2-D cellular system with nonoverlapping cells, the problem of satisfying an AP load vector with the minimum number of channels cannot be solved optimally in polynomial time as in a linear cellular network. The minimum number of required channels equals again the maximum clique load. Although the MP policy can achieve this bound, the required rearrangements to accommodate a new user increase exponentially with N. Thus, MP is impractical and serves as a benchmark policy only. Clearly, the joint problem of AP selection and channel allocation that arises if cells overlap is not an easier problem.

Consider the overlap area between cells B and C in Fig. 5. The type-2 users of this overlap area that select AP C affect the load of the three cliques that contain cell C but not B, namely cliques D, E and F. The type-2 users that are assigned to AP B load the three cliques that contain cell B but not C, namely cliques G, H and I. Clearly, the load of cliques which contain both cells C and B (namely cliques C, B, K and L) is not affected by selection of AP B or C. An admissible AP allocation α is specified as a collection of AP allocations for type-2 users of overlap areas. The set of admissible integer-valued AP allocations $\mathcal{F}$ is the set of assignments α where $f_{i,j_1}(\alpha)$, $f_{i,j_2}(\alpha)$ are integers and overlap areas $i \equiv (\kappa_i, \ell_i) \in \hat{\mathcal{S}}$ and cliques $j_1, j_2 \in \Omega$ satisfy $\kappa_i \in \mathcal{C}_{j_1} \setminus \{j_1\}$, $\kappa_i \notin \mathcal{C}_{j_2}$, and $\ell_i \in \mathcal{C}_{j_2} \setminus \{j_2\}$, $\ell_i \notin \mathcal{C}_{j_1}$. Furthermore, it is $f_{i,j_1}(\alpha) + f_{i,j_2}(\alpha) = \hat{s}_{\kappa_i,\ell_i}$. The load of each clique i consists of two components:

- A fixed load Q_i^0 that includes all type-1 users at cells of the clique and type-2 users in the six overlap areas between the central cell i and each one of the six surrounding cells
$$Q_j^0 = \sum_{i \in \mathcal{C}_j} s_i + \sum_{\substack{i \in \mathcal{C}_j \\ i \neq j}} \hat{s}_{i,j}. \tag{12}$$
- A load $Q_j(\alpha)$ that depends on AP selection for type-2 users in the overlap areas between each one of the six outer cells of the clique and the 12 other cells that constitute set $\mathcal{N}_j$. There exist 18 such overlap areas and $Q_j(\alpha)$ is given by
$$Q_j(\alpha) = \sum_{\substack{i \equiv (\kappa_i, \ell_i) \in \hat{\mathcal{S}} \\ \kappa_i \in \mathcal{C}_j \setminus \{j\}, \ell_i \in \mathcal{N}_j}} f_{i,j}(\alpha). \tag{13}$$

The problem of AP selection and channel allocation for the case of 2-D cellular network is formulated as

$$\min_{\alpha \in \mathcal{F}} \max_{i \in \Omega} \left(Q_i^0 + Q_i(\alpha)\right) \tag{14}$$

subject to the constraints:

a): $f_{i,j_1}(\alpha) + f_{i,j_2}(\alpha) = \hat{s}_{\kappa_i,\ell_i}, \forall\, i \equiv (\kappa_i, \ell_i) \in \hat{\mathcal{S}}, j_1, j_2 \in \Omega$ with $\kappa_i \in \mathcal{C}_{j_1} \setminus \{j_1\},\ \kappa_i \notin \mathcal{C}_{j_2}$ and $\ell_i \in \mathcal{C}_{j_2} \setminus \{j_2\},\ \ell_i \notin \mathcal{C}_{j_1}$.

b): for $m_1, m_2, m_3 \in \Omega : \kappa_i \in \mathcal{C}_{m_\rho} \setminus \{m_\rho\},\ \ell_i \notin \mathcal{C}_{m_\rho}$ and $\rho = 1, 2, 3,$ it is $f_{i,m_1} = f_{i,m_2} = f_{i,m_3}$

c): for $n_1, n_2, n_3 \in \Omega : \ell_i \in \mathcal{C}_{n_\rho} \setminus \{n_\rho\},\ \kappa_i \notin \mathcal{C}_{n_\rho}$, $\rho = 1, 2, 3,$ it is $f_{i,n_1} = f_{i,n_2} = f_{i,n_3}$.

In the formulation above, constraint a) refers to admissibility of the AP assignment. Constraint b) states that if AP κ_i is selected, the three affected cliques are equally loaded. Indices m_1, m_2, m_3 refer to the three cliques in which cell κ_i but not cell ℓ_i is included. Constraint c) states that if AP ℓ_i is selected, cliques n_1, n_2, n_3 are equally loaded, where n_1, n_2, n_3 refer to the three cliques in which cell ℓ_i but not cell κ_i is included.

We would like to devise a suboptimal heuristic algorithm for the joint AP selection and channel assignment problem in 2-D cellular networks by using SCLB as a guideline. Define function $G(\alpha)$ as in (8) and consider the problem of minimizing $G(\alpha)$ over all admissible real-valued assignments $\alpha \in \mathcal{F}_R$. For a given assignment $\alpha \in \mathcal{F}_R$ and overlap area $i \equiv (\kappa_i, \ell_i) \in \hat{\mathcal{S}}$, we define the assignment $T_i\alpha$. This assignment stems from α by minimizing $G(\alpha)$ *only* with respect to $\{(f_{i,m_\rho}, f_{i,n_\rho}), \rho = 1, 2, 3\}$, while keeping all other assignments $f_{k,j}$ fixed, for $k \in \hat{\mathcal{S}}$, $k \neq i$ and $j \in \Omega$.

The constraints of AP selection for overlap area i are $f_{i,m_1} = f_{i,m_2} = f_{i,m_3}$ and $f_{i,n_1} = f_{i,n_2} = f_{i,n_3}$ and that $f_{i,m_\rho} + f_{i,n_\rho} = \hat{s}_{\kappa_i,\ell_i}$ for $\rho = 1$, 2,3. This equality constraint holds for each of the nine clique pairs (m_ρ, n_ρ). The load of each one of cliques $\{m_\rho : \rho = 1, 2, 3\}$ is $Q_{m_\rho} = Q_{m_\rho}^0 + f_{i,m_\rho} + X_{m_\rho}$, where $Q_{m_\rho}^0$ is the fixed load of clique m_ρ and X_{m_ρ} is the load that comes from overlap areas other than i. Similarly for each one of cliques $\{n_\rho : \rho = 1, 2, 3\}$, we have $Q_{n_\rho} = Q_{n_\rho}^0 + f_{i,n_\rho} + X_{n_\rho}$. For $\Phi(x) = x^2$, the problem of minimizing $G(\alpha)$ with respect to $\{(f_{i,m_\rho}, f_{i,n_\rho}), \rho = 1, 2, 3\}$ subject to the constraints above has solution

$$f_{i,m_\rho} = \frac{1}{2}\left\{\hat{s}_i - \frac{1}{3}\left[\sum_{\rho=1}^{3}(Q_{m_\rho} + X_{m_\rho}) - \sum_{\rho=1}^{3}(Q_{n_\rho} + X_{n_\rho})\right]\right\}$$
$$f_{i,n_\rho} = \frac{1}{2}\left\{\hat{s}_i - \frac{1}{3}\left[\sum_{\rho=1}^{3}(Q_{n_\rho} + X_{n_\rho}) - \sum_{\rho=1}^{3}(Q_{m_\rho} + X_{m_\rho})\right]\right\}$$

for $\rho = 1$, 2, 3, where we have set $\hat{s}_{\kappa_i,\ell_i} = \hat{s}_i$ for notational simplicity. Thus, the AP selection balances the *average* load of cliques m_1, m_2, m_3 with that of cliques n_1, n_2, n_3. In the equations above, the difference in average loads of the two clique groups was assumed to be less than $\hat{s}_i$. If this is not the case, we have $f_{i,m_\rho} = 0$ and $f_{i,n_\rho} = \hat{s}_i$ or $f_{i,m_\rho} = \hat{s}_i$ and $f_{i,n_\rho} = 0$, depending on the largest average load.

The iterative algorithm is as follows. We start with a random AP selection for each type-2 user. Overlap areas are parsed sequentially again with a PAC and a NA phase. In the PAC phase of the nth iteration, AP selections from the $(n-1)$th iteration are canceled. In the NA phase, APs are selected so that the average loads of the two affected clique groups are as balanced as possible. The proof of convergence of the algorithm is similar to that of SCLB.

An AP selection by a type-2 user affects two groups with three cliques each. The algorithm minimizes the maximum clique *group* load over all possible AP assignments. Since this is not equivalent to minimizing the maximum clique load, the algorithm does not lead to a feasible load vector for the channel assignment problem even if there exists one. Nevertheless, the algorithm described above constitutes a meaningful heuristic with several adopted characteristics from the optimal SCLB algorithm for linear cellular networks.

IV. The Algorithm in a Dynamic Scenario

In previous sections, we considered the static version of the problem, namely a snapshot model. Our objective was to satisfy a given load vector by using the minimum number of channels. In a dynamic scenario, the load vector $(\mathbf{s}(t), \hat{\mathbf{s}}(t))$ varies with time because of new appearing calls, call terminations and handoffs due to user mobility. The joint problem needs to be solved whenever the load of overlap and nonoverlap areas changes. Whenever a new user arises in cell i at time t, a channel allocation algorithm needs to determine whether load vector $\mathbf{s}(t)+\mathbf{e}_i$ is feasible so that the user can be admitted. If a connection at cell i is terminated at time t, the channel that was used is released and the load vector becomes $\mathbf{s}(t)-\mathbf{e}_i$. If a handoff for a user occurs at time t from cell i to j, the algorithm needs to decide whether $\mathbf{s}(t)-\mathbf{e}_i+\mathbf{e}_j$ is feasible. If the load vector is feasible, a channel allocation needs to be identified. In a dynamic system, a meaningful QoS measure is the probability of the event of blocking for new users, which occurs when all channels are occupied upon arrival of a user.

With MP, a new call in cell i at time t is accepted if $\mathbf{s}(t)+\mathbf{e}_i \in \mathcal{S}_{\rm MP}$. MP has the largest space of feasible load vectors over all channel allocation algorithms and hence it leads to minimum blocking probability. In a linear network with N nonoverlapping cells and reuse distance $R=2$, MP is implemented in polynomial time with the so-called ordered dynamic channel assignment algorithm with rearrangements [9]. When a new call arises in AP i, the AP attempts to find a channel that does not violate channel reuse constraints starting the check from channel 1 in increasing order if i is odd or from channel L in decreasing order if i is even. If a call terminates in an odd-indexed cell, the call that uses the highest-indexed channel in that cell is reassigned to the channel that was used by the terminated call if the index of this channel is smaller than that of the channel currently in use. If a call terminates in an even-indexed cell, the call that uses the lowest-indexed channel in that cell is reassigned to the channel that was used by the terminated call if the index of this channel is higher than that of the channel currently in use.

We now provide conditions for admission and blocking of a new user. Consider a linear network with overlapping cells and $R=2$. Let K_i be the event that the load of clique i exceeds the number of available channels L i.e., $Q_i \geq L$. The following conditions hold for blocking of a new request.

- A new type-1 user in cell i, $1<i<N$ unavoidably loads cliques $i-1$ and i. Thus, it gets blocked if one or of these cliques are full, namely when event $K_i \cup K_{i-1}$ is true. Blocking in cell 1 or N occurs only when event K_1 or K_{N-1}, respectively, is true.
- A new type-2 user in overlap region i, $1<i<N-1$ loads clique $i-1$ or $i+1$ if it selects AP i or $i+1$, respectively. Clique i is loaded anyway. If $Q_i<L$, there exists a channel assignment such that the user can be admitted, otherwise the user is blocked. However, if $Q_i<L$ and both cliques $i-1$ and $i+1$ are full, the call is blocked. Hence, blocking occurs when event $K_i \cup (K_{i-1} \cap K_{i+1})$ is true. A type-2 call in overlap region 1 or $N-1$ is blocked when K_1 or K_{N-1}, respectively, is true.

V. Numerical Results

The goal of simulations is to compare the performance of the proposed algorithm to other candidate ones and to capture the impact of different parameters on performance. We first consider a linear cellular network with $N=20$ cells. Type-1 and type-2 users arrive in independent Poisson streams and each user requires one channel. The arrival rate of type-1 user requests for each cell is λ_1 requests/min and the arrival rate of type-2 user requests for each overlap region is $\lambda_2=\lambda_1/4$ requests/min. This ratio was considered to be small enough to capture realistic scenarios of type-1 and type-2 user arrival rates and large enough to explicitly manifest the performance of our algorithm. Mobility is considered to be low enough so that load variations primarily occur due to new call arrivals or call terminations. Call duration is exponentially distributed with mean $\tau=1/\mu=90$ sec. Traffic load is measured in Erlangs as $\lambda_1\tau/60$. The convergence of the SCLB algorithm with integral assignments was verified in all conducted experiments. We compare the performance of the following five algorithms for joint AP selection and channel assignment.

- Algorithm 1: (SCLB). The algorithm is executed at regular time instants t_k, $k=1,2,\ldots$. New calls arrive or existing calls terminate at each time instant t_k and clique loads are updated. Existing type-2 calls that originated at time $t_j<t_k$ are counted as being already assigned to an AP with SCLB at t_j.
- Algorithm 2: SCLB with handoff option (SCLB-H/O). Although SCLB balances clique loads, an AP selection by a new call may cause an existing type-2 call to be blocked. With the handoff option, accommodation of existing users and the new one may be achieved if adjacent cliques are not heavily loaded. With algorithm 2, an ongoing call in overlap area i that is assigned to clique $i+1$ (AP $i+1$) may switch to AP i if $Q_{i+1}=L$.
- Algorithm 3: Least Loaded Clique Routing (LLCliqueR). Users select the AP once upon arrival. A type-2 user is assigned to the AP that corresponds to the least loaded clique.
- Algorithm 4: Least Loaded Cell Routing (LLCellR). A type-2 user selects the AP once and is assigned to the cell with the least load. Clique balancing is not used here.
- Algorithm 5: Random Routing (RR). Upon arrival, a type-2 user picks randomly one of the two neighboring APs with probability $1/2$.

Algorithm 2 is an enhanced version of SCLB, while algorithms 3 and 4 are noniterative balancing algorithms. The performance metric is the blocking probability P_b and the contribution from the first and last cell is not considered. Different traffic loads under DCA and FCA are considered. Results are averaged over several experiments.

In Fig. 6, the performance of the algorithms above is depicted as a function of available channel capacity for light (5 Erlangs) and moderate (30 Erlangs) traffic loads with DCA. The SCLB algorithm with the option of handoff provides the lowest blocking probability for a given number of channels. For low traffic load and moderate system capacity, blocking probability under SCLB is two or three orders of magnitude lower than that of the least loaded (LL) routing policies. Thus, for 50 available channels SCLB achieves $P_b=10^{-5}$ while other policies result in P_b of the order of 10^{-3}. For moderate traffic load of 30 Erlangs and moderate number of available channels, blocking probability of SCLB is lower than that of other policies by a factor of 5–10. When the number of channels increases, this performance difference is more evident.

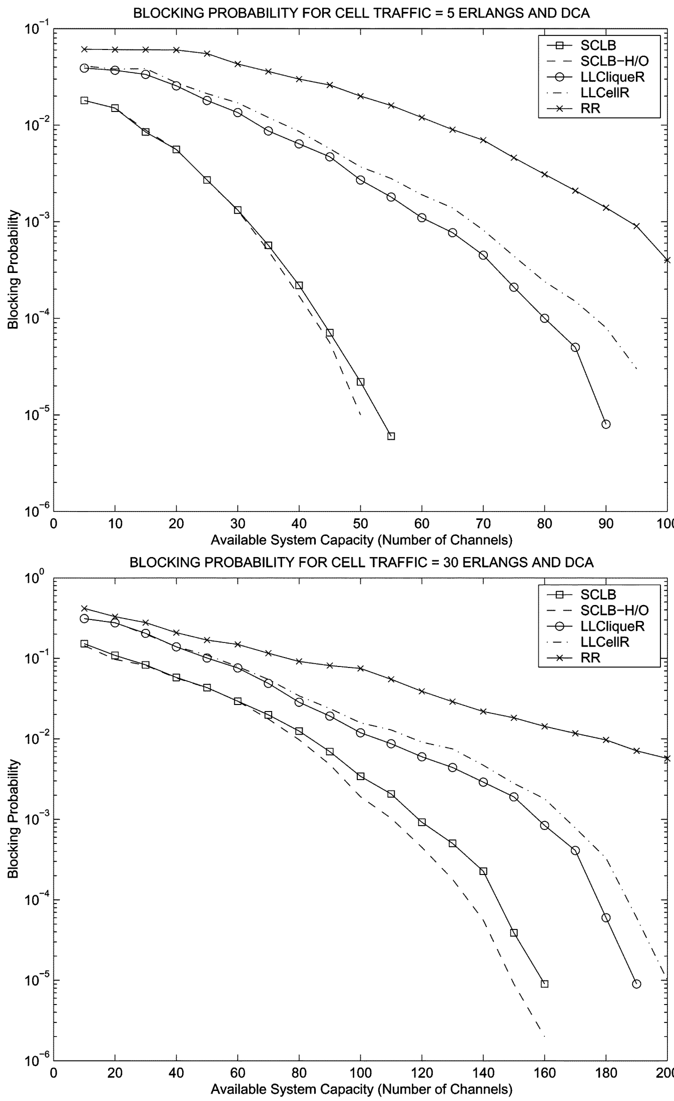

Fig. 6. Blocking probability versus number of available channels for DCA and traffic loads of 5 Erlangs or 30 Erlangs.

TABLE I
BLOCKING PROBABILITY FOR 150 CHANNELS AND DIFFERENT TRAFFIC LOADS UNDER DCA

Erlangs	SCLB	SCLB-H/O	LLCliqueR	LLCellR
10	0	0	$3.9 \cdot 10^{-6}$	$2.7 \cdot 10^{-5}$
20	$5.7 \cdot 10^{-7}$	$9.1 \cdot 10^{-6}$	$6.2 \cdot 10^{-4}$	$1.4 \cdot 10^{-3}$
40	$8.1 \cdot 10^{-4}$	$4.3 \cdot 10^{-4}$	$5.2 \cdot 10^{-2}$	$8.4 \cdot 10^{-2}$
60	$2.8 \cdot 10^{-1}$	$2.8 \cdot 10^{-1}$	$4.6 \cdot 10^{-1}$	$5.5 \cdot 10^{-1}$

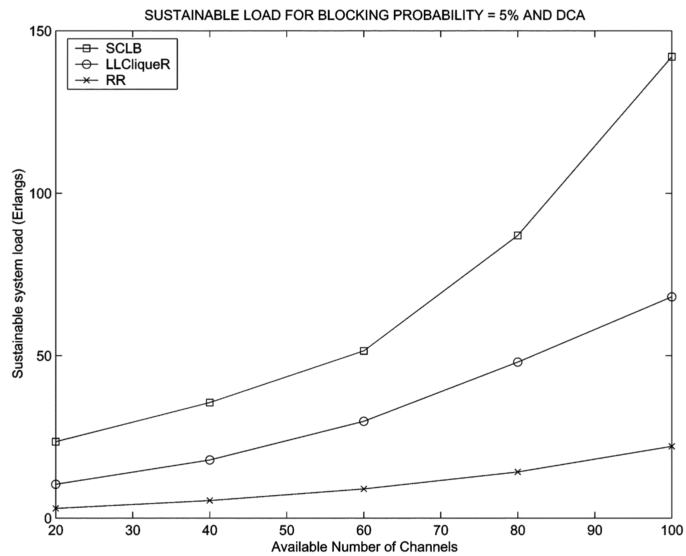

Fig. 7. Sustainable load (in Erlangs) as a function of available number of channels for $P_b = 5\%$ and DCA.

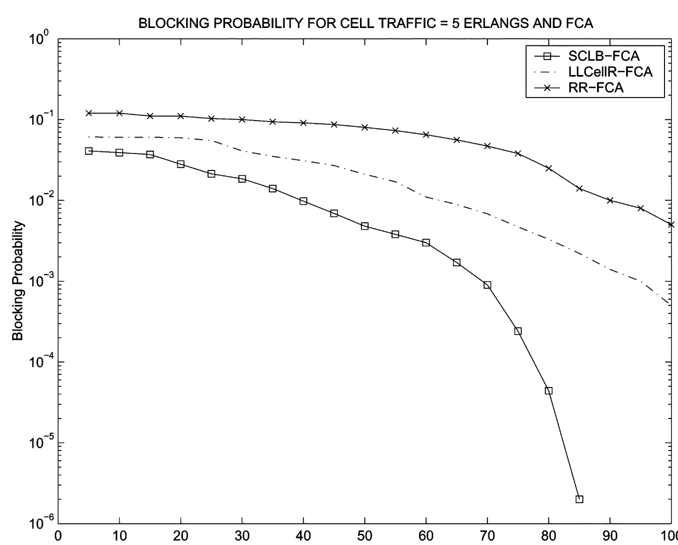

Fig. 8. Blocking probability vs. number of available channels for a traffic regime of 5 Erlangs and FCA.

For larger channel capacity, blocking under SCLB can even be eliminated. We observe that the handoff option in SCLB provides substantial improvement only for large number of channels. Therefore it should be used only in that case, since it involves significant message overhead. LL clique or cell routing results in tolerable blocking probability only for light traffic. LL clique routing always performs better that LL cell routing but this difference decreases as traffic load increases. As expected, random allocation has the lowest performance of all policies.

In Table I, we present performance results for different traffic loads of 10, 20, 40, and 60 Erlangs for a capacity of 150 available channels. When traffic load increases, SCLB reaches its performance limits and thus P_b under SCLB has the same order of magnitude as that of other policies.

A different way of interpreting the performance gain of SCLB is to note that fewer channels are required to maintain a given blocking probability for a certain traffic load. From Fig. 6 we deduce that for 5 Erlangs, SCLB needs only 35 channels to guarantee $P_b = 5 \cdot 10^{-3}$, while other policies require at least 80 channels. Equivalently, under SCLB the system can sustain larger traffic load with a certain number of channels. In Fig. 7, we plot the sustainable system load as a function of number of channels for $P_b = 5\%$ for SCLB, LLCliqueR, and RR DCA. Fig. 8 depicts the performance of algorithms under FCA for light traffic load. In FCA the available channels are evenly distributed across

TABLE II
BLOCKING PROBABILITY FOR 50 CHANNELS AND TRAFFIC LOAD OF 5 ERLANGS FOR DIFFERENT REUSE DISTANCES R UNDER DCA

R value	SCLB	SCLB-H/O	LLCliqueR	LLCellR
$R = 2$	$2.1 \cdot 10^{-5}$	$1.3 \cdot 10^{-5}$	$2.9 \cdot 10^{-3}$	$4.3 \cdot 10^{-3}$
$R = 3$	$6.8 \cdot 10^{-4}$	$6.5 \cdot 10^{-4}$	$7.5 \cdot 10^{-2}$	$9.8 \cdot 10^{-2}$
$R = 4$	$5.9 \cdot 10^{-3}$	$5.8 \cdot 10^{-3}$	$1.1 \cdot 10^{-1}$	$2.4 \cdot 10^{-1}$
$R = 5$	$7.7 \cdot 10^{-2}$	$7.7 \cdot 10^{-2}$	$4.7 \cdot 10^{-1}$	$4.7 \cdot 10^{-1}$

TABLE III
2-D NETWORK: BLOCKING PROBABILITY FOR 500 CHANNELS AND DIFFERENT TRAFFIC LOADS UNDER DCA

Traffic (Erl)	SCLB	LLCliqueR	Random
3	$2.6 \cdot 10^{-3}$	$3.8 \cdot 10^{-3}$	$7.1 \cdot 10^{-3}$
6	$1.4 \cdot 10^{-2}$	$5.7 \cdot 10^{-2}$	$1.2 \cdot 10^{-1}$
10	$8.9 \cdot 10^{-2}$	$1.3 \cdot 10^{-1}$	$3.3 \cdot 10^{-1}$

cells of a clique. The performance of SCLB is again better than that of noniterative schemes but this difference is not as notable as in the DCA case, especially for moderate number of channels. By comparing results for DCA and FCA in Figs. 6 and 8, we deduce that SCLB is more suitable for DCA.

The impact of reuse distance on performance under DCA is shown in Table II for 50 channels and traffic load of 5 Erlangs. A large reuse distance denotes more stringent reuse constraints. For $R = 1$, namely for the sum interference model, we observed zero blocking probability for all algorithms. For $R > 1$, results indicate that an increase in reuse distance by one yields an increase of approximately one order of magnitude in blocking probability for SCLB. The handoff option leads to slight improvement for small reuse distances. Non-iterative policies have lower performance than SCLB, but the difference is smaller for larger reuse distances. In order to see the performance for $R = 1$, we performed experiments for 20 channels and 5 Erlangs. For the SCLB, SCLB with handover, LLCliqueR, LLCellR, and random allocation, we measured blocking probabilities of 0.013, 0.01, 0.06, 0.06, and 0.092, respectively.

Experiments were also conducted for a 2-D cellular network with a 12×8 cellular grid. The ratio of arrival rates at overlap and nonoverlap areas is $1/12$. In Table III, results are provided for SCLB, LLCliqueR, and RR for a system with 500 channels and traffic loads of 3, 6, and 10 Erlangs. Arrival rates were kept sufficiently low so as to give meaningful blocking probabilities. SCLB achieves again the best performance in terms of blocking. This performance gain is expected to be larger for higher type-2 user arrival rates. It can be seen that the performance difference between SCLB and the other two algorithms is less than the difference for the linear network case. This is attributed to the fact that SCLB may not result in a feasible allocation even if such an allocation exists.

VI. DISCUSSION

We addressed the joint problem of AP selection and channel assignment for cellular networks with overlapping cells and a DCA regime. We reduced the joint problem to a plain channel allocation one by devising an iterative load balancing algorithm. The result is a procedure that minimizes the number of required channels to satisfy given user loads in a linear cellular network. We also extended the approach to a meaningful heuristic for 2-D cellular systems.

There exist several directions for future study. A first issue is that of implementing the algorithm in a distributed fashion. In a linear network with nonoverlapping cells, MP can be implemented in a distributed manner with the ordered DCA method. Each AP informs the neighboring APs about the highest or lowest-indexed channel in use and thus each AP can independently assign channels to users. In the joint AP selection and channel assignment problem, the SCLB algorithm can be amenable to distributed implementation. Each AP communicates only with adjacent APs and learns the number of type-1 users in these cells and the number of type-2 users in adjacent overlap areas. With appropriate coordination for message exchange between type-2 users and APs, these users can relay information to neighboring APs. This issue is more challenging in a 2-D network. An interesting problem would also be to study the convergence of asynchronous versions of the distributed algorithm that lack coordination.

In this work we adhered more to a circuit-switched philosophy, since we associated each user with one channel and considered blocking probability as QoS metric. Our approach is applicable to packet-based systems where each user may need different number of channels due to different rate requirements. A more detailed model should take into account different rate requirements, precise co-channel interference and effects of channel errors in terms of retransmissions and queueing in AP buffers.

Another interesting issue emerges in the case of cells of controllable size. Depending on traffic load variations or other parameters, the AP transmitter can vary the transmit power and thus affect the loads of nonoverlap and overlap areas. It would be interesting to study methods similar in flavor to SCLB which incorporate such enhanced adaptation mechanisms.

REFERENCES

[1] I. Koutsopoulos and L. Tassiulas, “Joint base station and channel allocation in mobile cellular networks,” in *Proc. IEEE Int. Conf. Communications*, 2000, pp. 1558–1562.

[2] I. Katzela and M. Naghshineh, “Channel assignment schemes for cellular mobile telecommunication systems: A comprehensive survey,” *IEEE Pers. Commun.*, vol. 3, no. 3, pp. 10–31, Jun. 1996.

[3] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*. New York: Freeman, 1979.

[4] K. N. Sivarajan, R. J. McEliece, and J. W. Ketchum, “Channel assignment in cellular radio,” in *Proc. IEEE Vehicular Technology Conf.*, 1989.

[5] C. W. Sung and W. S. Wong, “Sequential packing algorithm for channel assignment under cochannel and adjacent-channel interference constraint,” *IEEE Trans. Veh. Technol.*, vol. 46, no. 3, pp. 676–686, Aug. 1997.

[6] R. Mathar and J. Mattfeldt, “Channel assignment in cellular radio networks,” *IEEE Trans. Veh. Technol.*, vol. 42, no. 4, pp. 647–655, Nov. 1993.

[7] M. Frodigh, “Bounds on the performance of DCA algorithms in highway microcellular systems,” *IEEE Trans. Veh. Technol.*, vol. 43, no. 3, pp. 420–427, Aug. 1994.

[8] D. Everitt and D. Manfield, “Performance analysis of cellular mobile communication systems with dynamic channel assignment,” *IEEE J. Sel. Areas Commun.*, vol. 7, pp. 1172–1180, Oct. 1989.

[9] P. A. Raymond, “Performance analysis of cellular networks,” *IEEE Trans. Commun.*, vol. 39, no. 12, pp. 1787–1793, Dec. 1991.

[10] B. Hajek, “Performance of global load balancing by local adjustment,” *IEEE Trans. Inf. Theory*, vol. 36, no. 6, pp. 1398–1414, Nov. 1990.

[11] M. Alanyali and B. Hajek, "On simple algorithms for dynamic load balancing," in *Proc. IEEE INFOCOM*, 1995, pp. 230–238.
[12] R. Yates and C. Y. Huang, "Integrated power control and base station assignment," *IEEE Trans. Veh. Tech.*, vol. 44, no. 3, pp. 638–644, Aug. 1995.
[13] S. Papavassiliou and L. Tassiulas, "Joint optimal channel base station and power assignment for wireless access," *IEEE/ACM Trans. Networking*, vol. 4, no. 6, pp. 857–872, Dec. 1996.
[14] S. Papavassiliou and L. Tassiulas, "Improving the capacity in wireless networks through integrated channel base station and power assignment," *IEEE Trans. Veh. Technol.*, vol. 47, no. 2, pp. 417–427, May 1998.

Iordanis Koutsopoulos (S'99–M'03) was born in 1974 in Athens, Greece. He received the Diploma in electrical and computer engineering from the National Technical University of Athens, Athens, Greece, in 1997, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Maryland, College Park, in 1999 and 2002, respectively.

He is a Lecturer at the Department of Computer and Telecommunications Engineering, University of Thessaly, Greece. From 1997 to 2002 he was a Fulbright Fellow and a Research Assistant with the Institute for Systems Research (ISR) of UMCP. He has held internship positions with Hughes Network Systems, Germantown, MD, Hughes Research Laboratories LLC, Malibu, CA, and Aperto Networks Inc., Milpitas, CA, in 1998, 1999, and 2000, respectively. For the summer period of 2005, he was a visiting scholar with the University of Washington, Seattle. His research interests are in the field of networking with emphasis on wireless networks cross-layer design, sensor networks, smart antennas and wireless network security.

Leandros Tassiulas (S'89–M'91–SM'05) was born in 1965 in Katerini, Greece. He received the Diploma in electrical engineering from the Aristotelian University of Thessaloniki, Thessaloniki, Greece, in 1987, and the M.S. and Ph.D. degrees in electrical engineering from the University of Maryland, College Park, in 1989 and 1991, respectively.

He is a Professor in the Department of Computer and Telecommunications Engineering, University of Thessaly, Greece, and a Research Professor in the Department of Electrical and Computer Engineering and the Institute for Systems Research, University of Maryland, College Park, since 2001. He has held positions as Assistant Professor at Polytechnic University New York (1991-1995), Assistant and Associate Professor at University of Maryland, College Park (1995-2001), and Professor at University of Ioannina, Greece (1999-2001). His research interests are in the field of computer and communication networks with emphasis on fundamental mathematical models, architectures and protocols of wireless systems, sensor networks, high-speed internet and satellite communications.

Dr. Tassiulas received a National Science Foundation (NSF) Research Initiation Award in 1992, an NSF CAREER Award in 1995, an Office of Naval Research (ONR) Young Investigator Award in 1997, a Bodosaki Foundation Award in 1999, and the IEEE INFOCOM '94 Best Paper Award.