Cell Selection in 4G Cellular Networks

David Amzallag, Reuven Bar-Yehuda, Danny Raz, and Gabriel Scalosub

Abstract—Cell selection is the process of determining the cell(s) that provide service to each mobile station. Optimizing these processes is an important step towards maximizing the utilization of current and future cellular networks. We study the potential benefit of global cell selection versus the current local mobile SNR-based decision protocol. In particular, we study the new possibility available in OFDMA-based systems, such as IEEE 802.16m and LTE-Advanced, of satisfying the minimal demand of a mobile station simultaneously by more than one base station.

We formalize the problem as an optimization problem, and show that in the general case this problem is not only NP-hard but also cannot be approximated within any reasonable factor. In contrast, under the very practical assumption that the maximum required bandwidth of a single mobile station is at most an r-fraction of the capacity of a base station, we present two different algorithms for cell selection. The first algorithm produces a $(1-r)$-approximate solution, where a mobile station can be covered simultaneously by more than one base station. The second algorithm produces a $\frac{1-r}{2-r}$-approximate solution, while every mobile station is covered by at most one base station. We complete our study by an extensive simulation study demonstrating the benefits of using our algorithms in high-loaded capacity-constrained future 4G networks, compared to currently used methods. Specifically, our algorithms obtain up to 20% better usage of the network's capacity, in comparison with the current cell selection algorithms.

Index Terms—Cellular Networks, 4G, WiMAX, LTE-Advanced, Approximation Algorithms, Cell Selection, Association, Resource Allocation

1 Introduction

The ability to provide services in a cost effective manner is one of the most important building blocks of competitive modern cellular systems. Usually, an operator would like to have a maximal utilization of the installed equipment, that is, to maximize the number of satisfied customers at any given point in time. This paper addresses one of the basic problems in this domain, the cell selection mechanism. This mechanism determines the base station (or base stations) that provides the service to a mobile station - a process that is performed when a mobile station joins the network (called *cell selection*), or when a mobile station is on the move in idle mode (called *cell reselection*, or *cell change*, in HSPA).

In most current cellular systems the cell selection process is done by a local procedure initialized by a mobile device according to the best detected SNR. In this process the mobile device measures the SNR to several base stations that are within radio range, maintains a "priority queue" of those that are best detected (called an *active set*), and sends an official service subscription request to base stations by their order in that queue. The mobile station is connected to the first base station that positively confirmed its request. Reasons for rejecting service requests may be handovers or drop-calls areas, where the capacity of the base station is nearly exhausted. Such approaches usually result in significantly sub-optimal associations of mobile users to base stations. Optimizing cell-selection is crucial in future 4G networks due to both the restricted capacity available at the base stations (e.g., in the case of femtocells), combined with the increased demand by the mobile users. To the best of our knowledge, very little work has focused on optimizing these procedures taking into account the scarcity of spectrum expected in future 4G cellular networks.

Consider for example the settings depicted in Figure 1. Assume that the best SNR for Mobile Station 1 (MS1) is detected from microcell A, and thus MS1 is being served by this cell. When Mobile Station 2 (MS2) arrives, its best SNR is also from microcell A, who is the only cell able to cover MS2. However, after serving MS1, microcell A does not have enough capacity to satisfy the demand of MS2 who is a heavy data client. However, if MS1 could be served by picocell B then both MS1 and MS2 could be served. Note that MS1 and MS2 could represent a cluster of clients. The example

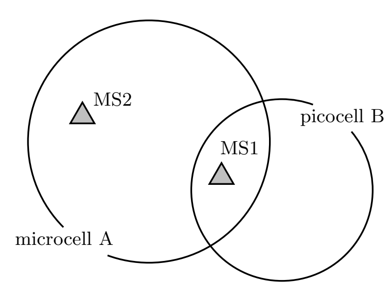


Fig. 1. Bad behavior of the *best detected SNR* algorithm in high-loaded capacitated network.

- *D. Amzallag is with British Telecommunications plc. E-mail:* dudu.amzallag@gmail.com. *R. Bar-Yehuda and D. Raz are with the Computer Science Department, Technion – I.I.T., Haifa 32000, Israel. E-mail:* {reuven,danny}@cs.technion.ac.il. *G. Scalosub is with the Department of Communication Systems Engineering, Ben Gurion University of the Negev, Beer-Sheva 84105 Israel. E-mail:* sgabriel@bgu.ac.il.
- *A Preliminary version of this work appeared in the Proceedings of IEEE INFOCOM 2008, The 27th Annual Joint Conference of the IEEE Computer and Communications Societies, pp. 700-708, 2008.*

shows that the best-detected-SNR algorithm can be a factor of $\max\{\tilde{d}\}/\min\{\tilde{d}\}$ from an optimal cell assignment, where $\tilde{d}$ is the demand of any mobile station in the coverage area. Theoretically speaking, this ratio can be arbitrarily large.

This simple example illustrates the need for a global, rather then a local, cell selection solution that tries to maximize the global utilization of the network, and not just the SNR of a single user. In voice only networks, where base station capacities are considered to be high, sessions have limited duration, and user demands are uniform, this may not be a big barrier. That is, the current base station selection process results, in most cases, in a reasonable utilization of the network. However, in the forthcoming 4G cellular networks this may not be the case.

Although the detailed structure of 4G systems is as of yet not well defined, there is a clear consensus regarding some of the important aspects of the technologies to be implemented in these systems[1]. Fourth generation systems are planned to provide even higher transmission rates and larger capacity than current 3G (IMT-2000 based) systems, both in terms of the number of users as well as in terms of traffic volume. Most likely, 4G systems will be designed to offer bit rates of 100 Mbit/s (peak rate in mobile environment) to 1 Gbit/s (fixed indoors) with a 5 MHz frequency bandwidth. The systems' capacities are expected to be at least 10 times larger than current 3G systems. In addition, these objectives should be met together with a drastic reduction in the cost (1/10 to 1/100 per bit) [19]. Such high frequencies yield a very strong signal degradation and suffer from significant diffraction resulting from small obstacles, hence forcing the reduction of cell size (in order to decrease the amount of degradation and to increase the degree of coverage), resulting in a significantly larger number of cells compared to previous generations. In addition, with the further deployment of technologies such as femtocells in providing broadband access, it is expected that users will remain connected for long periods of time. When considering the user demand to cell capacity ratio, for voice traffic in current cellular technologies, a single cell may support up to $\sim$100 concurrent voice connections [16]. As demands are expected to increase on the order of 2-3 folds, whereas capacities are expected to increase by a mere 1-2 folds [22], the resulting user demand to cell capacity ratio is expected to increase by at least one order of magnitude. These facts render the optimization of cell-selection and association extremely significant in optimizing spectrum and capacity utilization.

The increased number of base stations, and the variable bandwidth demand of mobile clients, will force operators to optimize the way the *capacity* of a base station is utilized. Unlike in previous generations, the ability of a base station to successfully satisfy the service demand of all its mobile clients would be highly limited and will mostly depend on its infrastructure restrictions, as well as on the service distribution among its mobile clients.

Another interesting aspect is the support for different QoS classes for the mobile stations, (e.g., *gold*, *silver*, or *bronze*). In such a case, the operator would like to have as many satisfied "gold" customers as possible, even if this means several unsatisfied "bronze" customers.

In this paper we study the potential benefit of a new global cell selection mechanism, which should be contrasted with the current local mobile SNR-based decision protocol. In particular, we rigourously study the problem of maximizing the number of mobile stations that can be serviced by a given set of base stations in such a way that each of the serviced mobile stations has its minimal demand fully satisfied. We differentiate between two coverage paradigms: The first is *cover-by-one* where a mobile station can receive service from at most one base station. The second is *cover-by-many*, where we allow a mobile station to be simultaneously satisfied by more than one base station. This means that when a mobile stations has a relatively high demand (e.g., video-on-demand) in a sparse area (e.g., sea-shore), several base stations from its active set can participate in its demand satisfaction. This option is not available in third-generation networks (and not even in HSPA networks) since these networks have universal frequency reuse and the quality of a service a mobile station receives will be severely damaged by the derived co-channel interference. However, OFDMA-based technology systems and their derivatives are considered to be among the prime candidates for future cellular communication networks. The ability to satisfy the demand of a mobile station by more than one member of its active set is *possible* in these systems, as defined by the IEEE 802.16m standard, using MIMO technology [2], and also as part of LTE-Advanced (e.g., in the form of cooperative multipoint transmission) [1], [15]. An important question in this context is whether cover-by-many is indeed more powerful than cover-by-one, in the sense that it improves the ability of the network to satisfy more clients. Simple examples can show that improvement can indeed be made, by considering, e.g., combining the leftover capacities of base stations in order to support additional users. However, it is not clear how much more powerful is the cover-by-many paradigm, when compared with the cover-by-one approach.

Our work focuses on the *offline* version of the problem, where the entire input is given in advance. We believe that from a theoretical point of view, a better understanding of the offline problem is extremely helpful if one is to design good algorithms for distributed or online settings, which are the natural settings in which our problem is cast. For example, we note that our offline hardness results guide us to focus on restricted settings. Furthermore, additional work that followed the preliminary version of our work has built on our offline results to design distributed algorithms for related models, essentially focusing on the cover-by-one paradigm [23].

Approximation algorithms and heuristics play a major role in our paper. A γ-approximation algorithm is a polynomial-time algorithm that always finds a feasible solution for which the value of the objective function is within a proved factor of γ of the optimal solution. Heuristics will be described in comparison with the worst-case behavior of approximation algorithms, in order to design a good practical solution to the problems in question.

1. See International Telecommunication Union (ITU) Web Site at `http://www.itu.int/home/index.html`.

Our Contribution

In this paper we present a new approach for cell selection that is derived from the anticipated 4G technologies. To the best of our knowledge, despite recent extensive research done on future cellular networks planning and coverage optimization (e.g., [4], [21]), there is no explicit study in the literature discussing the new IEEE 802.16m and LTE-Advanced possibility of simultaneous servicing of mobile clients by more than one base station.

We model, in Section 2, the cell selection problem as an optimization problem called *all-or-nothing demand maximization* (AoNDM). We show that the general version of AoNDM cannot be approximated within a factor better than $|J|^{1-\epsilon}$, unless NP = ZPP, for any $\epsilon > 0$, where J is the set of mobile stations. Motivated by this result, we address a special case of the problem. Following practical scenarios, we define a restrictive version of AoNDM, the r-AoNDM problem, for some $r < 1$, where the network satisfies the condition that the demand of every mobile station is at most an r fraction of the capacity of any base station that can potentially cover the mobile station. We show that even this special case of the problem is NP-hard. These results appear in Section 4.

We further present, in Section 4, two different algorithms for this problem. The first is a $\frac{1-r}{2-r}$-approximation algorithm, which uses the cover-by-one paradigm, i.e., every mobile station is covered by at most one base station. Note that this approximation guarantee is with regard to the optimal *cover-by-many* assignment. The second algorithm uses the cover-by-many paradigm, where a mobile station can be covered simultaneously by more than one base station. It is a careful refinement of the first algorithm, and we prove it guarantees at least a $1 - r$ fraction of the value of an optimal solution, at a price of increased running time. We note that better approximation guarantees were known for the cover-by-one assignment problem, however these algorithms were based on solving a linear programming relaxation of the problem and rounding the resulting solution in order to obtain an integral solution. Our algorithms, on the other hand, are based on the local-ratio method, and are thus strictly *combinatorial*. Furthermore, our results for the cover-by-many assignment problem provide the currently best-known guarantees for this problem.

In order to evaluate the practical differences between global and local mechanisms for cell selection in future networks we conducted an extensive simulation study (Section 5). We compare between global mechanisms that are based on our approximation algorithms and the current best-SNR greedy cell selection protocol. We study the relative performance of these three algorithms under different conditions. In particular, we show that in a high-load capacity-constrained 4G-like network, where clients' demands may be large with respect to cell capacity, global cell selection can achieve up to 20% better coverage than the current best-SNR greedy cell selection method.

2 Model and Definitions

Consider a bipartite graph $G = (I, J, E)$ where $I = \{1, 2, \ldots, m\}$ is the set of base stations and $J = \{1, 2, \ldots, n\}$ is the set of mobile stations (or *clients*). Every client $j \in J$ has a non-negative demand $d(j)$, and a non-negative profit $p(j)$, and every base station $i \in I$ has a non-negative capacity $c(i)$. In addition, for every base station $i \in I$, the coverage area of i is modeled by a subset $S_i \subseteq J$ of clients which can be serviced by i. The set of base stations $N(j) \subseteq I$ connected by edges to a client $j \in J$, represents the active set of this client. We further extend the above definitions to sets of nodes, such that for every $A \subseteq J$, $d(A) = \sum_{j\in A} d(j)$ and $p(A) = \sum_{j\in A} p(j)$, and for every $B \subseteq I$, $c(B) = \sum_{i\in B} c(i)$. Furthermore, given any $A \subseteq J$, we let $N(A) = \bigcup_{j\in A} N(j)$. Given a subset of clients $S \subseteq J$, a *cover plan* for S is a weight function $x : E \to \mathbb{R}^+$, such that for every $j \in S$, $\sum_{i\,:\,(i,j)\in E} x(i,j) \geq d(j)$, and for every $i \in I$, $\sum_{j\,:\,(i,j)\in E} x(i,j) \leq c(i)$. Notice that such a restriction of $\sum_{i\,:\,(i,j)\in E} x(i,j) \geq d(j)$, for every $j \in S$, is also known as *all-or-nothing*-type of coverage. This means that clients that are partially satisfied are not considered to be covered (such a model appears, for example, in OFDMA-based networks where mobile stations have their slot requirements over a frame and these are not useful if not fulfilled).

The *all-or-nothing demand maximization problem* (AoNDM) is to find a subset of clients $S \subseteq J$, and a cover plan x for S, such that $p(S)$ is maximized.

For $i \in I$, we use $x(i) = \sum_{j\,:\,(i,j)\in E} x(i,j)$, and for $j \in J$, we use $x(j) = \sum_{i\,:\,(i,j)\in E} x(i,j)$. As before, we extend these notations to sets of nodes, such that for every $A \subseteq I$, $x(A) = \sum_{i\in A} x(i)$, and for every $B \subseteq J$, $x(B) = \sum_{j\in B} x(j)$. We further extend this notation to subgraphs of G, such that given any $A \subseteq I$ and $B \subseteq J$, $x(A, B) = \sum_{(i,j)\in E\cap(A\times B)} x(i,j)$.

In addition, for every $v \in I \cup J$ we denote by $E(v)$ the set of edges with endpoint v, and for every $W \subseteq I \cup J$, let $E(W) = \bigcup_{v\in W} E(v)$. We further denote for every $A \subseteq I$ and $B \subseteq J$, $E(A, B) = \{(i,j) \in E \cap (A \times B)\}$.

Given any constant $r < 1$, we say an instance is *r-restricted* if for every $(i,j) \in E$, $d(j) \leq r \cdot c(i)$. We further define the problem of *r-AoNDM* as the AoNDM problem limited to r-restricted instances.

3 Related Work

Cell selection has received much attention in recent years (e.g., [13], [18], [24], [25]) where research focused mainly on multiple-access techniques, as well as on power control schemes and handoff protocols [13], [24], [25].

In [13] a cell selection algorithm is presented where the goal is to determine the power allocations to the various users, as well as a cover-by-one allocation, so as to satisfy per-user SINR constraints. An HSPA-based handoff/cell-site selection technique is presented in [24], [25], where the objective is to maximize the number of connected mobile stations (very similar to our objective), and reaching the optimality of this objective is done via a new scheduling algorithm for this cellular system. All the above results did not take into account variable base station capacities nor mobile station bandwidth demands. In the case of [24], [25], this enables the authors to reduce their corresponding optimization problem to a polynomial-time solvable matching problem. As shown in our paper, when base station capacities and/or mobile stations' demands are incorporated, this approach is no longer feasible.

An integrated model for optimal cell-site selection and frequency allocation is shown in [18], where the goal is to maximize the number of connected mobile stations, while maintaining quasi-independence of the radio based technology. The optimization problem in this model is shown to be NP-hard.

The case where we restrict the clients in AoNDM to be satisfied by a *single* base station belongs to the family of generalized assignment problems. Among this class of problems the most related problem to AoNDM is the *separable assignment problem* (SAP) [12].

In this problem we are given a set U of m bins, a set H of n items, and a profit, f_{ij}, for assigning item j to bin i. The assignment constraints are such that every $i \in U$ has a family $\mathcal{I}_i$ of feasible subsets that can be packed in bin i, such that $\mathcal{I}_i$ is closed under taking subsets, i.e., if $A \in \mathcal{I}_i$, then so is every subset of A. The goal is to find an assignment of items to bins with the maximum aggregate profit.

The suggested SAP solution presented in [12], depends on an algorithm which solves the single-bin subproblem in SAP. Given a β-approximation algorithm for finding the highest profit packing of a single bin, they present a polynomial-time LP-rounding based $((1-\frac{1}{e})\beta)$-approximation algorithm and a polynomial-time local search $(\frac{\beta}{\beta+1}-\epsilon)$- approximation algorithm, for any $\epsilon > 0$. Specifically, for all special cases of SAP that admit an approximation scheme for the single-bin problem, there exists an LP-based algorithm with a $(1-\frac{1}{e}-\epsilon)$-approximation guarantee, and a local search algorithm with a $(\frac{1}{2}-\epsilon)$)-approximation guarantee.

This problem is a generalization of several well known problems. Among these problems are the *maximum generalized assignment problem* (GAP), and the *multiple knapsack problem* (MKP). In GAP we are given a set of bins with capacity constraints and a set of items that have a possibly different size and profit for each bin. We wish to pack a maximum-profit subset of items into the bins. MKP is the special case of GAP where the size and the profit of each item are the same for all the bins.

Shmoys and Tardos [26] give an LP-rounding based 2-approximation algorithm for the minimization version of GAP. However, Chekuri and Khanna [8] observed that a $1/2$-approximation for standard GAP is implicit in [26]. In addition, Chekuri and Khanna [8] develop a PTAS for MKP and also classify the APX-hardness of GAP. Cast in our terminology, the algorithm of Shmoys and Tardos [26] for the GAP problem solves our problem using the cover-by-one paradigm. Their algorithm, which is based on solving a linear programming (LP) relaxation of the problem, and rounding the solution to obtain an integral solution, guarantees a 2-approximate solution with respect to the optimal *fractional* solution to the LP, which also serves as an upper bound on the optimal solution to our problem under the cover-by-many paradigm. This result holds for any $r \leq 1$. Furthermore, for r-restricted instances with $r < 1/2$, the observation made by Chekuri and Khanna [8] implies that the cover-by-one algorithm proposed by Shmoys and Tardos yields a $(1-2r)$-approximate solution. Our scheme provides a two-fold improvement upon these results; First, our algorithms are combinatorial, and do not resort to solving linear programming relaxations of the problem. Second, our ultimate result guarantees a $(1-r)$-approximate solution for any $r \leq 1$, which improves upon the best known results for any $r \leq 1/2$.

Fractional packings, where items are allowed to be partitioned among more than one bin, are studied in [20] and in [17]. The assumption in the models studied in these papers is that splitting an item is associated with overhead, thus the objective is to pack the items (possibly splitting some of the items) with maximal profit and minimum overhead. Our problem is different since in our case, while splitting is for free, not every item can be packed in every bin.

AoNDM is closely related to the problem of planning 4G cellular networks under budget limitation as described in [3], [5]. In this problem, in addition to the input of AoNDM, we are given a set I of possible configuration of base stations, as well as an opening cost $w(i)$ for every $i \in I$. When a client belongs to the coverage area of more than one base station, interference between the servicing stations may occur. These interferences are modeled by a penalty-based mechanism and may reduce the contribution of a base station to a client. The *budgeted cell planning problem* asks for a subset of base stations $I' \subseteq I$ whose cost does not exceed a given budget B, and the total number of fully satisfied clients is maximized. Notice that in these settings, by taking the set I of base stations with zero opening costs, without interferences, we get a special case of AoNDM where all clients have the same profit. It was shown [3] that this problem cannot be approximated, unless P=NP, and that a $\frac{e-1}{3e-1}$-approximation algorithm exists for a special case of the problem where every set of k open base stations can fully satisfy at least k clients, for every integral value of k.

Another closely related problem is the *all-or-nothing multicommodity flow problem* discussed in [9] and [10]. In this problem we are given a capacitated undirected graph $G = (V, E, u)$ (where u is the edge-capacity function) and set of k pairs $(s_1, t_1), \ldots, (s_k, t_k)$. Each pair has a unit demand. The objective is to find a largest subset S of $\{1, \ldots, k\}$ such that one can simultaneously route for every $i \in S$ one unit of flow between s_i and t_i. It is straightforward to verify that the unit profit version of AoNDM is a special case of this problem. It was shown that the all-or-nothing multicommodity flow problem can be approximated within an $O(\log^2 k)$ factor of the optimum [10]. On the other hand, for any $\epsilon > 0$, the problem cannot be approximated to within a factor of $O(\log^{\frac{1}{3}-\epsilon}|E|)$ of the optimum, unless $\mathrm{NP} \subseteq \mathrm{ZPTIME}(|V|^{\mathrm{poly}\log|V|})$ [6]. However, no special attention is given to specific network topologies (e.g., bipartite graphs, as in our case), and other special instances.

Following the preliminary version of our work, Patt-Shamir et al. have considered our problem in the distributed setting, focusing merely on the cover-by-one paradigm (essentially, the GAP problem). They build upon our results and show, for any

$\epsilon \in (0,1)$, a distributed $\frac{1-r}{2-r}(1-\epsilon)$-approximation (with high probability) algorithm which runs in $O(\epsilon^{-2}\log^3 n)$ communication rounds. This result uses a distributed emulation of our proposed cover-by-one algorithm, which iteratively finds maximal matchings to improve upon the solution.

4 Approximating the r-AoNDM problem

4.1 Lower Bounds and General Framework

The important goal of efficiently solving the AoNDM problem is beyond our reach since this problem is NP-hard, as we mentioned before. Moreover, as the following theorem shows, even obtaining a reasonable approximation algorithm for the problem is improbable under standard complexity assumptions.

Theorem 4.1: For any $\epsilon > 0$, AoNDM cannot be approximated to within a factor better than $|J|^{1-\epsilon}$, unless $\mathrm{NP} = \mathrm{ZPP}$.

Proof: We present a reduction from the Maximum-Size Independent Set (MIS) problem to AoNDM. Let $G = (V,E)$ be any input to MIS. Consider the bipartite graph $\tilde{G} = (I,J,\tilde{E})$, where $I = E$, $J = V$, and $\tilde{E} = \{(e,v) \in E \times V \mid v \text{ is an endpoint of } e\}$. For every $v \in V$, let $\delta(v)$ denote the degree of v in G, and let $M = \max_v \delta(v)$. For every $j \in J$ we set $d(j) = \delta(j)$ and set $p(j) = 1$. Finally, we define for every $i \in I$, $c(i) = 1$.

Since all clients have unit profit, our goal is to maximize the number of clients which can be covered. Let S be any subset of J, and let x be any cover plan for S. For any $j \in S$, the overall capacity of the base stations connected to j is

$$\sum_{i:(i,j)\in\tilde{E}} c(i) = \delta(j) = d(j).$$

It follows, that any client covered in S uses all the capacity of the base stations in its range. Hence, a base station may contribute to the covering of at most one client, and in particular, any $e \in E$ can contribute to covering at most one of its endpoints. It follows that for any $S \subseteq V$, S has a cover plan if and only if it is an independent set in G. Since for any $\epsilon > 0$, MIS cannot be approximated to within a factor better than $|V|^{1-\epsilon}$, unless $\mathrm{NP} = \mathrm{ZPP}$ [14], the same holds for AoNDM. □

Note that in this reduction the client's demand is $\delta(j)$ which is greater than the capacity of each of the base stations ($= 1$). In realistic cellular networks this is not the case. Motivated by this fact, and the above theorem, we focus on a special case of the problem. Namely, for any $r < 1$ we consider the r-AoNDM problem. The following theorem shows that even in such restrictive settings, the problem is still intractable.

Theorem 4.2: For any fixed $r < 1$, the r-AoNDM problem is NP-hard, even if there is only one base station.

Proof: We show a reduction from the Knapsack problem, which is known to be NP-hard. Let K be any instance to the Knapsack problem, which comprises of a set of elements A, and a knapsack of size B, such that every element $a \in A$ has a size $s(a)$, and a value $v(a)$. Let $S = \sum_{a\in A} s(a)$, and $P = \sum_{a\in A} v(a)$. Given any $r \in (0,1)$, we let $M \in \mathbb{N}$ such that $\frac{1}{M} \leq r < \frac{1}{M-1}$.

We construct a bipartite graph $G = (I,J,E)$, such that I consists of a single node i, and $J = J_A \cup J_B$ where $J_A = \{j_a \mid a \in A\}$, and $J_B = \{b_1,\ldots,b_M\}$. We let $E = I \times J$. For every $j_a \in J_A$ we let $d(j_a) = s(a)$, and $p(j_a) = v(a)$. For every $b_\ell \in J_B$ we let $d(b_\ell) = 2S$, and $p(b_\ell) = 2P$. We set $c(i) = 2MS + B$.

First note that the above instance to AoNDM is r-restricted. To see this note that for every $j \in J_A$, $d(j) \leq S < 2S + \frac{B}{M}$, and for every $j \in J_B$, $d(j) = 2S < 2S + \frac{B}{M}$. Since $2S + \frac{B}{M} = \frac{c(i)}{M} \leq r \cdot c(i)$ we have for every $(i,j) \in E$, $d(j) \leq r \cdot c(i)$.

In addition, note that for every optimal solution X to the above r-AoNDM instance, $J_B \subseteq X$. This follows from the fact that J_B is a feasible solution, and for every $j \in J_B$, the profit obtained by covering j is strictly greater than the profit obtained from covering all of J_A. It therefore follows that the cover plan for X uses exactly $2MS$ units of i's capacity, leaving a capacity of B to cover clients in J_A. Hence, the subset $X \cap J_A$ induces an optimal solution to the original knapsack problem. This completes the proof that r-AoNDM is NP-hard for any $r < 1$, even if there is only one base station. □

In the following sections we present two approximation algorithms for the r-AoNDM problem. The algorithms are local-ratio algorithms that are based on a decomposition of the profit obtainable from every client into two non-negative terms; One part is proportional to the demand of the client, while the other part is the remaining profit. We define a family of feasible solutions, which we dub "maximal" (see below for the formal definition), and prove that any such solution is an approximate solution when considering a profit function which is proportional to the demand. The algorithms we present generate such maximal solutions recursively. We then apply an inductive argument which proves that the solution generated by the algorithm is also an approximate solution w.r.t. the original profit function.

We first present an approximation algorithm that guarantees a solution whose value is within a factor of $\frac{1-r}{2-r}$ from the value of an optimal solution. This algorithm follows the cover-by-one paradigm, and thus every mobile station is covered by at most one base station. Our second algorithm is obtained by a careful refinement of this algorithm, and an appropriate change to the notion of maximality. This algorithm uses the cover-by-many paradigm, and is guaranteed to produce a solution whose value is within a factor of $(1-r)$ from the value of an optimal solution, while the complexity increases by a polynomial factor. Next we specify several definitions needed for the analysis of the proposed algorithms.

Given any instance of r-AoNDM over a graph $G = (I,J,E)$, and any two subsets $A \subseteq I$ and $B \subseteq J$, we define the A-B *flow-graph of* G, $G_f(A,B) = (V,F)$, such that $V = \{s\} \cup A \cup B \cup \{t\}$ for new vertices $s,t \notin I \cup J$, and $F = (\{s\} \times A) \cup E(A,B) \cup (B \times \{t\})$. We define a capacity function $\gamma : F \to \mathbb{R}^+$ as follows:

$$\gamma(u,v) = \begin{cases} c(v) & \text{if } u = s,\ v \in A \\ \infty & \text{if } u \in A,\ v \in B \\ d(u) & \text{if } u \in B,\ v = t. \end{cases}$$

For brevity of notation, we let $G_f = G_f(I, J)$. Given any two subsets $C, D \subseteq V$, we let $\gamma(C, D) = \sum_{u,v \in F \cap (C \times D)} \gamma(u, v)$.

A cover plan x for $S \subseteq J$ is said to be a *cover-by-one plan* if for every $j \in S$, there is exactly one $i \in I$ such that $x(i, j) > 0$. Given a cover-by-one plan x for $S \subseteq J$, a cover-by-one plan x' for $T \subseteq J$ is said to be a *T-extension of* x, if for any $j \in S$ and every $i \in I$, $x'(i, j) = x(i, j)$. Note that in such a case one is guaranteed to have $S \subseteq T$. Given a cover plan x for $S \subseteq J$, a cover plan x' for $T \subseteq J$ is said to be a *T-rearrangement of* x, if $S \subseteq T$.

Given any cover-by-one plan x for $S \subseteq J$, we say that x is *cover-by-one-maximal (CBO-maximal)* if for any $j \in J \setminus S$, no $S \cup \{j\}$-extension of x exists. We further say $S \subseteq J$ is CBO-maximal when it has a CBO-maximal cover plan which is clear from the context. For any $A \subseteq I$ and $B \subseteq J$, and any flow y in $G_f(A, B)$, we can denote the value of the flow by $y(s)$. Given any cover plan x for $S \subseteq J$, we say that x is *rearrangement-maximal* if for any $j \in J \setminus S$, no $S \cup \{j\}$-rearrangement of x exists. Given any set $S \subseteq J$, let $\overline{S} = J \setminus S$ and $Y_S = I \setminus N(\overline{S})$. We say a cover plan x for $S \subseteq J$ is *cover-by-many-maximal (CBM-maximal)* if x is rearrangement-maximal, and $x(Y_S, S)$ is a maximum flow in the flow graph $G_f(Y_S, S)$. As before, we further say $S \subseteq J$ is CBM-maximal when it has a CBM-maximal cover plan which is clear from the context.

The following lemma, appearing in [7], serves as a basic tool with which we analyze the approximation guarantee of the algorithms proposed in this section.

Lemma 4.3 (Local Ratio): Let $\mathcal{I}$ be an instance to r-AoNDM, over a graph $G = (I, J, E)$, with profit function p. Then, if $p = p_1 + p_2$, and x is a cover plan for some set $S \subseteq J$ which is c-approximate w.r.t. p_1, and also c-approximate w.r.t. p_2, then x is c-approximate w.r.t. p.

4.2 A cover-by-one $\frac{1-r}{2-r}$-approximation algorithm

We start with Algorithm CBO-MC. We note that there are at hand cover-by-one algorithms that ensure an approximation ratio better than our proposed algorithm [8], [26]. However, these algorithms are based on solving linear programming relaxations of the problem, and rounding the resulting solution in order to obtain an integral solution. Our algorithm, on the other hand, is strictly *combinatorial*. Furthermore, our cover-by-one algorithm forms the basis for the algorithm we describe in section 4.3, which uses the cover-by-many paradigm. In this sense, understanding the simpler settings set forth in this section also serves to clarify many of the tools and approaches used when describing and analyzing the more complex settings of the cover-by-many algorithm discussed in the sequel.

We now turn to describe our cover-by-one algorithm, Algorithm CBO-MC. Roughly speaking, under CBO-MC, given a specific ordering of the clients, and given an existing cover plan x, a client is added greedily by finding a CBO-extension of x, if such an extension exists. Otherwise, the client is discarded. See Algorithm 1 for the pseudocode of the algorithm.

Lemma 4.4: Consider any instance of the r-AoNDM problem such that for every client j, $p(j) = \epsilon \cdot d(j)$, for some constant ϵ. Any cover-by-one plan x for $S \subseteq J$ which is CBO-maximal is a $\frac{1-r}{2-r}$-approximate solution w.r.t. profit function p.

Algorithm 1 CBO-MC ($G = (I, J, E)$, demands d, profits p, capacities c)

if $J = \emptyset$ **then**
    return $x \equiv 0$
end if
if there exists a $j \in J$ such that $p(j) = 0$ **then**
    $x \leftarrow$ CBO-MC ($G' = (I, J \setminus \{j\}, E \setminus E(j))$, d, p, c)
    return x
else
    for every $j \in J$, set $\epsilon_j = \frac{p(j)}{d(j)}$
    set $\epsilon = \min_j \epsilon_j$
    for every $j \in J$, set $p_1(j) = \epsilon \cdot d(j)$
    set $p_2 = p - p_1$
    $x \leftarrow$ CBO-MC (G, d, p_2, c)
    for every j such that $p_2(j) = 0$ **do**
        if $\exists i \in N(j)$ such that $c(i) - x(i) \geq d(j)$ **then**
            set $x(i, j) = d(j)$
        else
            discard j
        end if
    end for
    return x
end if

Furthermore, this approximation is with regards to the optimal cover-by-many solution.

Proof: Let $\overline{S} = J \setminus S$. Without loss of generality, we can assume that no uncovered client receives any service, i.e., for every $j \in \overline{S}$, $x(j) = 0$.

If $S = J$, then x is an optimal cover plan, and therefore clearly a $\frac{1-r}{2-r}$ approximate solution. Assume therefore that $S \subsetneq J$. First note that for every $i \in N(S)$, one of the following holds:

- Either there are no edges between i and $\overline{S}$, or
- $x(i) = x(i, S) > (1 - r)c(i)$.

To see this, assume by contradiction that there exists an $i \in N(S)$ such that there are edges between i and $\overline{S}$, and $x(i) \leq (1 - r)c(i)$. By the assumption, there exists at least one client $j \in \overline{S}$ such that $(i, j) \in E$. Consider the function $x' : E \to \mathbb{R}^+$ defined by

$$x'(i', j') = \begin{cases} d(j') & \text{if } i' = i,\ j' = j \\ x(i', j') & \text{otherwise.} \end{cases}$$

Clearly, for every $i' \neq i$, x' does not violate the capacity constraint imposed by $c(i')$, since by the feasibility of x, for every such i', $x'(i) = x(i) \leq c(i)$. Furthermore, since x was a cover-by-one plan, then so is x'. Consider base station i. Since by the assumption $x(i) \leq (1-r)c(i)$, using the fact that the instance is r-restricted, we have $x'(i) = x(i) + d(j') \leq c(i)$, hence the capacity constraint is satisfied for i as well. Finally, note that all clients $j' \in S \cup \{j\}$ are satisfied by the cover plan x'. It follows that x' is an $S \cup \{j\}$-extension of x, contradicting the assumption that x is CBO-maximal. Using a similar argument one can show that $N(\overline{S}) \subseteq N(S)$, otherwise there is a base station in $N(\overline{S}) \setminus N(S)$ that can satisfy at least one client in $\overline{S}$, contradicting the maximality of S. It follows that for every $i \in N(\overline{S})$, $x(i) > (1-r)c(i)$.

Let $\mathrm{OPT} \subseteq J$ denote any optimal solution to the problem, *assuming the cover-by-many paradigm*. Note that

$$\begin{aligned} p(\mathrm{OPT}) &= p(\mathrm{OPT} \cap S) + p(\mathrm{OPT} \cap \overline{S}) \leq p(S) \\ &+ \epsilon \cdot \sum_{j \in \mathrm{OPT} \cap \overline{S}} d(j) \leq p(S) + \epsilon \cdot c(N(\overline{S})) \end{aligned}$$

where the last inequality follows from the feasibility of OPT.

On the other hand, by the maximality of S, we are guaranteed to have

$$\begin{aligned} d(S) &= \sum_{j \in S} d(j) = \sum_{i \in I} x(i) \geq \sum_{i \in N(\overline{S})} x(i) \\ &> \sum_{i \in N(\overline{S})} (1-r) \cdot c(i) = (1-r) \cdot c(N(\overline{S})), \end{aligned}$$

which in turn implies

$$p(S) = \epsilon \cdot d(S) > \epsilon(1-r) \cdot c(N(\overline{S})).$$

It follows that

$$p(\mathrm{OPT}) \leq p(S) + \frac{p(S)}{1-r} = p(S)\left(1 + \frac{1}{1-r}\right) = \frac{2-r}{1-r} \cdot p(S),$$

hence S is a $\frac{1-r}{2-r}$ approximate solution w.r.t the profit function p, when compared to the optimal cover-by-many solution. □

Theorem 4.5: Algorithm CBO-MC produces a $\frac{1-r}{2-r}$-approximate solution. Furthermore, this approximation is with regards to the optimal cover-by-many solution.

Proof: We prove by induction on the recursion that the cover plan returned from every call is a $\frac{1-r}{2-r}$-approximate solution. Note that the number of clients in every two consecutive recursive calls decreases by at least 1, thus the recursion will terminate.

For the base case, since $J = \emptyset$, there are no clients to cover, hence $x \equiv 0$ is an optimal cover, and therefore clearly a $\frac{1-r}{2-r}$-approximate solution. For the inductive step, we have two cases to consider. First, consider the cover plan x' for $B \subseteq J \setminus \{j\}$ returned in line 6. By the induction hypothesis, B is a $\frac{1-r}{2-r}$ approximate solution w.r.t. the graph $G' = (I, J \setminus \{j\}, E \setminus E(j))$ and profit function p. Since $p(j) = 0$, the optimal profit w.r.t the graph $G = (I, J, E)$ and profit function p cannot be greater than the optimal profit w.r.t the graph G' and profit function p. Hence, B is also a $\frac{1-r}{2-r}$ approximate solution w.r.t. the graph $G = (I, J, E)$ and profit function p. The second case to consider is the cover plan x' for B returned in line 20. By the induction hypothesis, B is a $\frac{1-r}{2-r}$ approximate solution w.r.t. the graph $G = (I, J, E)$ and profit function p_2. Since for every client j considered in lines 13–19, $p_2(j) = 0$, the optimal profit w.r.t the graph $G = (I, J, E)$ and profit function p_2 cannot be greater than the optimal profit attainable from the instance returned from the recursive call. Hence, the solution returned in line 20 is a $\frac{1-r}{2-r}$ approximate solution w.r.t. the graph $G = (I, J, E)$ and profit function p_2, and so is any extension of it using clients j such that $p_2(j) = 0$. Note that for every client j such that $p_2(j) = 0$, who has a neighbor with sufficient residual capacity, j is added to the cover, where exactly one base station is used to satisfy its demand. It follows that the solution returned in line 20 is a CBO-maximal solution. By Lemma 4.4 it follows that this solution is a $\frac{1-r}{2-r}$ approximate solution w.r.t. the graph $G = (I, J, E)$ and profit function p_1. Using Lemma 4.3 we conclude that the solution returned is a $\frac{1-r}{2-r}$ approximate solution w.r.t. the graph $G = (I, J, E)$ and profit function $p = p_1 + p_2$, which completes the proof. □

Note that the solution x produced by algorithm CBO-MC is a cover-by-one plan. It therefore follows that the ratio between the *optimal* cover-by-one solution and the optimal cover-by-many solution is at most $\frac{1-r}{2-r}$ as well.

The running time of CBO-MC is governed by two components: (i) in each recursive call we sort the current set of items (to either find the minimum profit or minimum ratio, in lines 4 and 9, respectively), and (ii) finding the candidate base station for coverage (in line 14). This gives an overall running time of $O(n^2 \log n + nm)$.

4.3 A cover-by-many $(1-r)$-approximation algorithm

We now turn to describe our second algorithm, called CBM-MC, which achieves an approximation ratio of $(1-r)$ using the cover-by-many paradigm. Under CBM-MC, a client is added by first trying to exhaust the capacities of base stations which cannot contribute to uncovered clients, and then using the capacity of the remaining base stations in order to complete the cover. If such a cover cannot be produced, then the client is discarded. The pseudocode of the algorithm is given in Algorithm 2, where we use the subroutine EK-MaxFlow $(G_f(A, B))$ to denote the computation of the maximum s-t flow in the flow graph $G_f(A, B)$ using the Edmonds-Karp algorithm [11]. Our choice of the Edmonds-Karp algorithm is motivated by two of its properties, namely, the fact that it converges from any feasible flow, and the fact that it uses augmentation paths. This choice can be substituted by any algorithm for computing maximum flow, which satisfies these properties. The reason we require such an algorithm is due to the fact that our algorithm must ensure the property of CBM-maximality is maintained whenever a client is added to the cover (recall the definition of CBM-maximality appearing at the end of Section 4.1). Note that by duality, given any s-t flow in a flow graph $G_f(A, B)$, it is easy to verify if a cut is a minimum cut by checking that all the edges are saturated.

Given a cover plan x for $S \subseteq J$, let $\overline{S} = J \setminus S$, and consider I as partitioned into two sets: $N_{\overline{S}} = N(\overline{S})$, and $Y_S = I \setminus N_{\overline{S}}$. Note that by definition, for every $j \in \overline{S}$ and $i \in Y_S$, $(i, j) \notin E$. The following lemma provides a necessary and sufficient condition for covering a set of clients.

Lemma 4.6: For any instance of r-AoNDM over a graph $G = (I, J, E)$, and any $A \subseteq I$ and $B \subseteq J$, $\{t\}$ is a minimum s-t cut in the flow-graph $G_f(A, B)$ if and only if A can cover all clients in B.

Proof: Let x be a maximum s-t flow in $G_f(A, B)$. By duality, the value of the maximum s-t flow is the same as the capacity of the minimum s-t cut, and the edges of any such cut are all saturated by any maximum flow. Assume $\{t\}$ is a minimum s-t cut in $G_f(A, B)$, and assume by contradiction that

Algorithm 2 CBM-MC ($G = (I, J, E)$, demands d, profits p, capacities c)

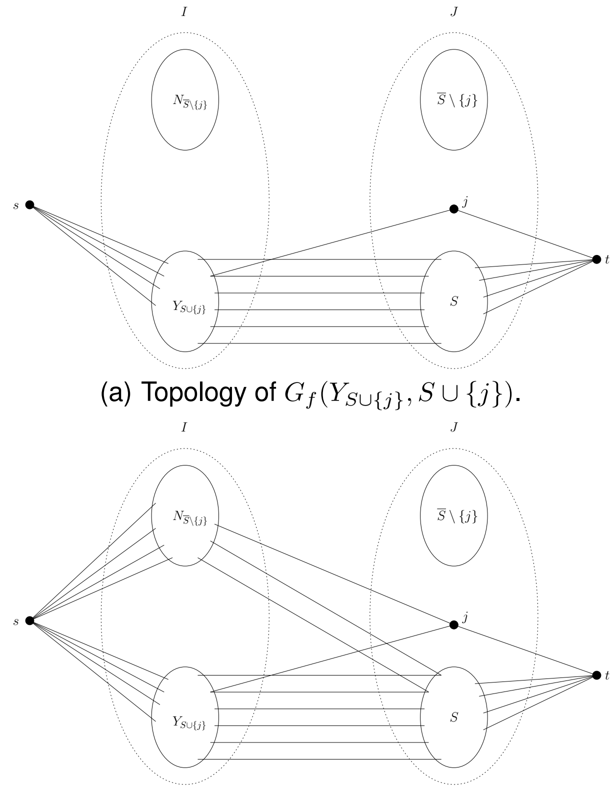
$x \leftarrow$ EK-MaxFlow (G_f)
if $\{t\}$ is a MinCut in G_f **then**
  return x
end if
if there exists a $j \in J$ such that $p(j) = 0$ **then**
  $x \leftarrow$ CBM-MC ($G' = (I, J \setminus \{j\}, E \setminus E(j))$, d, p, c)
  return x
else
  for every $j \in J$, set $\epsilon_j = \frac{p(j)}{d(j)}$
  set $\epsilon = \min_j \epsilon_j$
  for every $j \in J$, set $p_1(j) = \epsilon \cdot d(j)$
  set $p_2 = p - p_1$
  $x \leftarrow$ CBM-MC (G, d, p_2, c)
  for every j such that $p_2(j) = 0$ **do**
    $S \leftarrow \{j' \in J \mid x(j') = d(j')\}$
    set $N_{\overline{S}\setminus\{j\}} = N(J \setminus (S \cup \{j\}))$
    set $Y_{S\cup\{j\}} = I \setminus N_{\overline{S}\setminus\{j\}}$
    $y \leftarrow$ EK-MaxFlow($G_f(Y_{S\cup\{j\}}, S \cup \{j\})$) ▷ See Figure 2(a)
    $z \leftarrow$ EK-MaxFlow($G_f(I, S \cup \{j\})$), starting from the initial feasible flow y. ▷ See Figure 2(b)
    if $\{t\}$ is a MinCut in $G_f(I, S \cup \{j\})$ **then**
      $x \leftarrow z$
    end if
  end for
  return x
end if

there is client $j \in B$ which is not fully covered. It follows that $\sum_{i\,:\,(i,j)\in E} x(i,j) < d(j)$. By flow conservation, we have $x(j,t) = \sum_{i\,:\,(i,j)\in E} x(i,j)$, which implies $x(j,t) < d(j)$. Since (j,t) is an edge in the minimum s-t cut $\{t\}$, and its capacity is $d(j)$, this contradicts the fact that all such edges are saturated by the maximum flow x. Assume now $\{t\}$ is not a minimum s-t cut, and assume by contradiction that all clients can be covered. By flow conservation it follows that for every $j \in J$, $\sum_{i\,:\,(i,j)\in E} x(i,j) = d(j) = x(j,t)$, which by summing over all $j \in J$ implies that the value of the flow equals the capacity of the cut $\{t\}$. By duality, this implies that $\{t\}$ is a minimum s-t cut, contradicting our assumption. □

Lemma 4.6 admits a method for finding a rearrangement-maximal cover plan, as shown in the following lemma:

Lemma 4.7: Given any instance of r-AoNDM over a graph $G = (I, J, E)$, any cover plan x for $S \subseteq J$, and a client $j \in J \setminus S$, the task of finding an $S \cup \{j\}$-rearrangement of x, if one exists, can be done in polynomial time.

Proof: In order to find a rearrangement of x, consider the flow graph $G' = G_f(I, S \cup \{j\})$, and let y be a maximum flow in this graph. If $\{t\}$ is a minimum cut in G', then by Lemma 4.6, y is a cover plan for $S \cup \{j\}$, and hence it is an $S \cup \{j\}$-rearrangement. If $\{t\}$ is not a minimum cut in G'_f, then its capacity is strictly greater than the maximum flow, and therefore not all clients in $S \cup \{j\}$ can be satisfied. Verifying whether or not $\gamma(t) = y(t)$ can clearly be done in polynomial time. □



(a) Topology of $G_f(Y_{S\cup\{j\}}, S \cup \{j\})$.

(b) Topology of $G_f(I, S \cup \{j\})$.

Fig. 2. Outline of the flow graphs used in CBM-MC.

The following is an immediate corollary of Lemma 4.7:

Corollary 4.8: Given any instance to r-AoNDM over a graph $G = (I, J, E)$, any cover plan x for $S \subseteq J$, and a client $j \in J \setminus S$, the task of finding a rearrangement of x which is rearrangement-maximal can be done in polynomial time.

Proof: By iteratively applying Lemma 4.7, we are guaranteed to obtain a rearrangement-maximal cover plan. □

The following lemmas describe the correlation between the maximum flow in G_f, and the maximum flow in flow graphs of the form $G_f(Y_S, S)$, for sets S which have a cover plan.

Lemma 4.9: Assume $S \subseteq J$ has some cover plan. Then, there exists a maximum flow x in G_f such that $x(Y_S, S) =$ MaxFlow$(G_f(Y_S, S))$. Furthermore, such a flow can be found in polynomial time.

Proof: Let $y =$ MaxFlow$(G_f(Y_S, S))$. Clearly y is a feasible flow in G_f as well. Consider the Edmonds-Karp Algorithm (EK-MaxFlow, see [11] for details) for finding a maximum flow, executed on graph G_f, starting from the initial feasible flow y. We show that for every augmentation path found by EK-MaxFlow, after increasing the flow along this path and obtaining some flow y', $y'(s, Y_S) \geq y(s, Y_S)$.

First note that we can assume that all the augmentation paths used by the EK-MaxFlow algorithm are simple paths. Furthermore, note that by the fact that any augmentation path is simple, we obtain that for every flow y' obtained during executing the EK-MaxFlow algorithm, and for every $i \in Y_S$,

$y(s,i) \leq y'(s,i)$, since such flow can only decrease if the algorithm uses a path p such that $(i,s) \in p$, which implies that p is not a simple path.

Since for every feasible flow z we have $z(Y_S, S) = z(s, Y_S)$ (by flow conservation, and using the fact that there are no edges between Y_S and $\overline{S}$), we can conclude that during the entire execution of the EK-MAXFLOW algorithm, the flow y' resulting in augmenting any path p satisfies $y'(s, Y_S) \geq y(s, Y_S)$. On the other hand, note that given any maximum flow in G_f, if we consider its flow path-decomposition, then the set of paths using edges between Y_S and S also constitutes a flow in H_S (due to the unidirectionality of edges between $N_{\overline{S}}$ and S in G_f). Hence these paths cannot support a flow whose value is greater than $\text{MAXFLOW}(G_f(Y_S, S))$.

Finally note that EK-MAXFLOW produces a maximum flow in G_f in polynomial time, which completes the proof of the lemma. □

The above lemma gives rise to the following corollary:

Corollary 4.10: If there exists a rearrangement-maximal cover plan y for $S \subseteq J$, then there exists a CBM-maximal cover plan x for S. Furthermore, such a cover plan can be found in polynomial time.

Proof: Using a similar argument as the one used in Lemma 4.9, by running EK-MAXFLOW with an initial feasible flow $z = \text{MAXFLOW}(G_f(Y_S, S))$, we are guaranteed to produce a cover plan for S (by the existence of y, S can be covered by I). Furthermore, this cover must also be CBM-maximal. Note that such a cover plan can be found in polynomial time by the same arguments as the ones used in Lemma 4.9. □

The following lemma shows a bound on the value of any maximum flow in G_f.

Lemma 4.11: Given any $S \subseteq J$, if S has a CBM-maximal cover plan, then $\text{MAXFLOW}(G_f) \leq \text{MAXFLOW}(G_f(Y_S, S)) + c(N_{\overline{S}})$.

Proof: Let y be a CBM-maximal cover plan for S, and consider a partition of y into two types of flow paths, each consisting of 3 edges:

- $T_1 = \{p = (s, i, j, t) \mid \text{such that } i \in Y_S\}$.
- $T_2 = \{p = (s, i, j, t) \mid \text{such that } i \in N_{\overline{S}}\}$.

Note that such a packing exists, by the directionality of the edges in G_f.[2] If we denote the flow along a flow path p by $x(p)$, then clearly

$$\sum_{p \in T_1} x(p) \leq \text{MAXFLOW}(G_f(Y_S, S))$$

since all paths in T_1 are paths in $G_f(Y_S, S)$, and therefore cannot support a flow greater than $\text{MAXFLOW}(G_f(Y_S, S))$. On the other hand,

$$\sum_{p \in T_2} x(p) \leq c(s, N_{\overline{S}}) = c(N_{\overline{S}})$$

since all these paths use edges in the cut $(s, N_{\overline{S}})$. It therefore follows that

$$\text{MAXFLOW}(G_f) \leq \text{MAXFLOW}(G_f(Y_S, S)) + c(N_{\overline{S}}).$$

□

We can now continue in the same way as we did with the simpler algorithm, where CBM-maximality replaces CBO-maximality.

Lemma 4.12: Consider any instance of the r-AoNDM problem such that for every client j, $p(j) = \epsilon \cdot d(j)$, for some constant ϵ. Any cover plan x for $S \subseteq J$ which is CBM-maximal is a $(1-r)$-approximate solution w.r.t. profit function p.

Proof: Let x be any cover plan for $S \subseteq J$ which is CBM-maximal. If $S = J$, then x is an optimal cover plan, and therefore clearly a $(1-r)$ approximate solution. Assume $S \subsetneq J$. Note that by maximality of x, $x(Y_S, S) = \text{MAXFLOW}(G_f(Y_S, S))$, and since $S \subsetneq J$, $x(N_{\overline{S}}, S) > (1-r)c(N_{\overline{S}})$, i.e., $c(N_{\overline{S}}) < \frac{x(N_{\overline{S}}, S)}{1-r}$. By the fact that x is a cover plan for S, we have $p(S) = \epsilon d(S) = \epsilon(x(N_{\overline{S}}, S) + x(Y_S, S))$, since $N_{\overline{S}}, Y_S$ are a partition of I.

Let $\text{OPT} \subseteq J$ denote any optimal solution to the problem. We wish to bound the value of $p(\text{OPT})$. Clearly, for any maximum s-t flow y in G_f, $d(\text{OPT}) \leq y(s)$, since any cover plan for OPT induces a feasible flow in G_f. Combining the above with Lemma 4.11 we obtain that for any maximum s-t flow y in G_f,

$$\begin{aligned} d(\text{OPT}) &\leq y(s) \\ &\leq \text{MAXFLOW}(G_f(Y_S, S)) + c(N_{\overline{S}}) \\ &< x(Y_S, S) + \frac{x(N_{\overline{S}}, S)}{1-r} \\ &= \frac{1}{1-r}\left((1-r) \cdot x(Y_S, S) + x(N_{\overline{S}}, S)\right) \\ &\leq \frac{1}{1-r}\left(x(Y_S, S) + x(N_{\overline{S}}, S)\right) \\ &= \frac{1}{1-r} d(S). \end{aligned}$$

By the definition of p we obtain that $p(S) > (1-r) \cdot p(\text{OPT})$, which completes the proof. □

Theorem 4.13: Algorithm CBM-MC produces a $(1-r)$-approximate solution.

Proof: We prove by induction on the recursion that the cover plan returned from every call is a $(1-r)$-approximate solution, similarly to the proof of Theorem 4.5.

For the base case, if $\{t\}$ is minimum cut, then by Lemma 4.6 all the clients can be covered, hence x is an optimal cover plan, and therefore clearly a $(1-r)$-approximate solution. For the inductive step, we have two cases to consider. First, consider the cover plan x' for $B \subseteq J \setminus \{j\}$ returned in line 7. By the induction hypothesis, B is a $(1-r)$-approximate solution w.r.t. the graph $G' = (I, J \setminus \{j\}, E \setminus E(j))$ and profit function p. Since $p(j) = 0$, the optimal profit w.r.t the graph $G = (I, J, E)$ and profit function p cannot be greater than the optimal profit w.r.t the graph G' and profit function p. Hence, B is also a $(1-r)$-approximate solution w.r.t. the graph $G = (I, J, E)$ and profit function p. The second case to consider is the cover plan x' for B returned in line 24. By the induction hypothesis, B is a $(1-r)$ approximate solution w.r.t. the graph $G = (I, J, E)$ and profit function p_2. Since for every client j considered in lines 13–19, $p_2(j) = 0$, the optimal profit w.r.t the graph $G = (I, J, E)$ and profit function p_2 cannot be greater than the optimal profit attainable from the instance returned from the recursive call. Hence, the solution returned in line 20 is a $(1-r)$-approximate solution w.r.t. the graph $G = (I, J, E)$ and profit function p_2, and so is any superset of this solution produced by adding any of the clients for which $p_2(j) = 0$.

2. Note that these are not augmentation paths used in computing the maximum flow by EK-MAXFLOW. These paths are part of an actual path decomposition of the maximum flow.

Note that by Lemma 4.7, for every client $j \notin S$ such that $p_2(j) = 0$, lines 19 and 20 compute an $S \cup \{j\}$-rearrangement of the current cover plan, if such a rearrangement exists. Hence the resulting solution returned in line 24 is a rearrangement-maximal solution. In addition, by lines 18– 19, the cover plan x' computed in every iteration also satisfies that $x(Y_S, S)$ is a maximum flow in the flow graph $G_f(Y_S, S)$. It therefore follows that the cover plan returned in line 24 is also CBM-maximal. By Lemma 4.12 it follows that this solution is a $(1 - r)$-approximate solution w.r.t. the graph $G = (I, J, E)$ and profit function p_1. Using Lemma 4.3 we conclude that the solution returned is a $(1 - r)$ approximate solution w.r.t. the graph $G = (I, J, E)$ and profit function $p = p_1 + p_2$, which completes the proof. □

The running time of CBM-MC is governed by two components: (i) in each recursive call we sort the current set of items (to either find the minimum profit or minimum ratio, in lines 4 and 9, respectively), and (ii) computing the max flow on lines 18 and 19. If we denote by F the running time of the augmenting-paths max-flow algorithm used in lines 18 and 19, we obtain an overall running time of $O(n^2 \log n + nF)$. In case we use EK-MaxFlow, which finds a maximum flow in a graph $G = (V, E)$ in time $O(|V|\,|E|^2)$, we obtain an overall running time of $O((n + m)n^3 m^2)$.

5 Simulation results

In the previous sections we proposed two different algorithms for a new global mechanism for cell selection in 4G cellular networks. The main difference between these two algorithms is the way the demand of a mobile client is satisfied. In the CBO-MC Algorithm (Section 4.2) at most one base station satisfies the demand of any given mobile station while the CBM-MC Algorithm (Section 4.3) allows satisfaction of the demand simultaneously by more than one base station.

In order to study the expected performance of the proposed global cell selection algorithms with respect to the current local mobile SNR-based protocol we conducted several simulations over high-loaded, capacity constrained, 4G-like networks. A secondary goal of these simulations was to study the "benefit" of using the new ability, as made possible by the evolving standards of IEEE 802.16m and LTE-Advanced, of a mobile station to be serviced simultaneously by more than one base station.

5.1 Methodology

We considered a network consisting of an $n \times n$-grid of clients' locations (demand points, each considered as a single client, or *bin*). Each client has a service request for either voice or data service. The demand of a voice and data client is defined as 1 and 25, respectively[3]. Under this ratio between the demand of data and voice clients, the number of the data clients was chosen so that the overall voice volume is 20% of the network's traffic[4]. The locations for each type of client was uniformly and randomly selected over the grid. The profit for satisfying the demand of a voice client was defined as 1, while satisfaction of a data client is credited with a profit that is proportional to its demand (i.e., 25 units of profit).

We maintain microcells and picocells in our network. Since we implemented the restricted version of AoNDM, the demand of every client must be less than or equal to an r-fraction of the capacity of any base station service this client. Therefore, the capacity of a picocell was taken to be about $25/r$, for any given value of $0 < r < 1$. To simulate high-loaded networks we assumed that the total sum of (client) demands equals the sum of (base station) capacities in the network. The ratio between the number of picocells and microcells was defined to be λ while this factor was also selected as the ratio between the corresponding radiuses and capacities of microcells and picocells. By taking $\lambda = 5$, we can now derive the appropriate number of microcells and picocells. The locations for each type of base station was uniformly and randomly selected over the grid and clients were associated with (omnidirectional) base stations according to their distance from each of the centers.

In each of the following three sets of simulations we measured the ratio between the total profit achieved by each of the three algorithms and the total profit of all connected clients, i.e., clients that are within service range of some base station. As AoNDM is NP-hard, the maximum possible profit is hard to calculate, and we consider the total profit of all connected clients as an upper bound on the optimal solution.

5.2 Results

In the first set of simulations we study the performance of the three algorithms over various network sizes (10K to 40K) and different values of r (0.05 to 0.3). Typical results are shown in figures 3-5, where the upper, middle and the lower curves correspond to the cover-by-many algorithm, cover-by-one algorithm, and the greedy-best detected-SNR algorithm, respectively. In each of the three scenarios, our results show that the cover-by-many algorithm is better than the cover-by-one algorithm by 5% (for $r = 0.05$) to 11% (for $r = 0.3$). An improvement of at least 10% (and up to 20%) was achieved by the cover-by-many algorithm in comparison with the greedy-best detected-SNR algorithm. The results show that the performances of all three algorithm are nearly independent of the size of the network. Moreover, due to the existence of the simultaneous coverage in the third algorithm, when r increases the "distance" between the performance of the cover-by-many algorithm and the other two algorithms also increases in a significant fashion. This shows that when there exist mobile clients with demands that are relatively close to the capacity of the servicing cell (e.g., in case of picocells) allowing satisfaction of a client by more than one base station

3. The bit rate for voice applications is 64Kbps and the downlink rate for data application is approximately 2Mbps in HSDPA. This gives a ratio of 25-30 between the demand of voice and data clients.

4. To be precise, if n_v and n_d are the number of voice and data clients, respectively, and d_v and d_d are the corresponding demands, then the following are satisfies for an overall voice volume of γ of the network's traffic: $\frac{d_v \cdot n_v}{d_v \cdot n_v + d_d \cdot n_d} = \gamma$, $n_d = n^2 - n_v$, and $n_v = \left\lfloor \frac{\gamma \cdot d_d \cdot n^2}{(d_d - d_v) \cdot \gamma + 1} \right\rfloor$. In our case $\gamma = 0.2$.

is crucial in order to maintain high utilization of the network capacities.

The second set of simulations investigates the level of profit achieved by the three algorithms when the value of r varies (from $r = 0.01$ to $r = 0.5$). We fixed a network of 15129 clients (i.e., a grid of 123×123) with a number of picocells and microcells as explained above. Focusing on the relative fraction of the demand of a client with respect to the capacity of any serviced base station, the results show (Figure 6) that when this fraction increases the ability to reach a higher percentage of the total possible profit decreases. As shown in Figure 6, all three algorithms exhibit the same behavior. The performance of the cover-by-many algorithm (upper curve) decreases from 100% to 89% when r increases from 0.01 to 0.5. The cover-by-one algorithm decreases by 21% (from 100% in $r = 0.01$ to 79.5% in $r = 0.5$), and the greedy-best detected-SNR algorithm (lower curve) exhibited a decrease of 30% (from 89% to 59%). The third set of simulations examines the level of profit obtained by the three algorithms when the available capacity increases. We fixed a network of 15129 clients, where each client has a demand (of any service) that is at most a fraction of 1/4 ($r = 0.25$) of the capacity of each of the servicing base stations. In this study, the number of picocells as well as microcells was increased by j times their basic number, $j = 1, 1.5, 2, \ldots, 5$, where the basic numbers are the same as the ones computed in the first set of simulations (65 microcells and 327 picocells). Note that for $j > 1$, the total capacity is higher than the total demand of clients. As one might expect (see Figure 7), when there is a larger number of base stations the performance of the three algorithms can only improve. The greedy-best detected-SNR algorithm (lower curve) achieve an improvement of up to 8% (from 79% to 87%) when the number of base station grows from 392 to 1960. The cover-by-one algorithm (in the middle) achieves an improvement of up to 8% (from 89% to 97%), and the cover-by-many algorithm (upper curve) is nearly constant (around 99%) in its ability to satisfy clients.

Finally, the worst-case running time of each of the algorithms, for all cases, was approximately 4 minutes for the case of $n = 40000, r = 0.25$, on a Pentium M machine, 1.4 GHz, and 256 Mb of RAM.

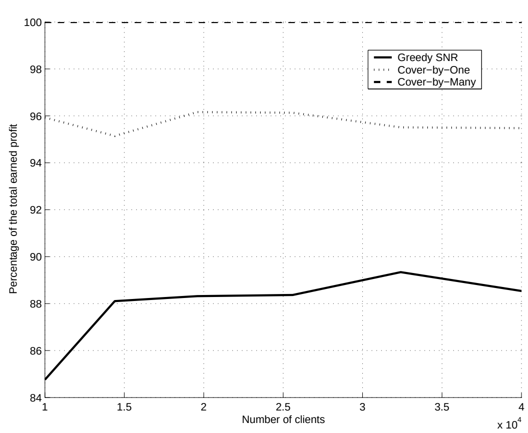


Fig. 3. Expected profit as a function of the number of clients, $r = 0.05$

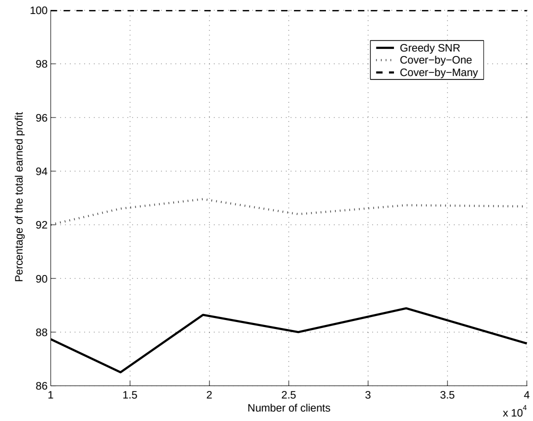


Fig. 4. Expected profit as a function of the number of clients, $r = 0.1$

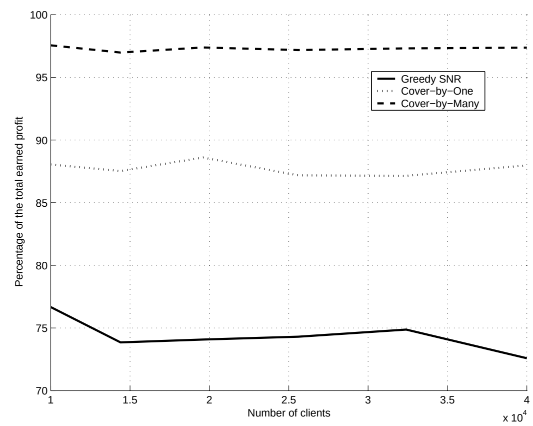


Fig. 5. Expected profit as a function of the number of clients, $r = 0.3$

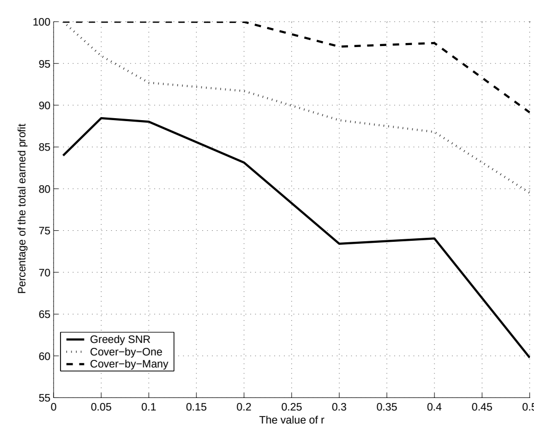


Fig. 6. Expected profit as a function of r ($n = 15129$)

6 Conclusions

In this paper we present a rigorous study of a new approach for cell selection in fourth generation cellular networks. Unlike the current cell selection protocol, our proposed mechanism is global, has a performance guarantee, and addresses many of

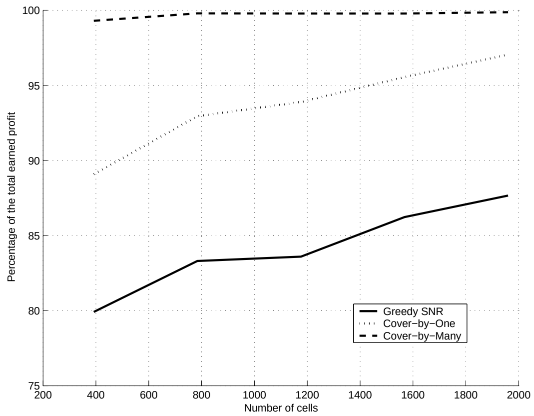


Fig. 7. Expected profit as a function of available capacity ($r = 0.25, n = 15129$)

the anticipated 4G technologies. We show that even though AoNDM is hard to approximate to within a reasonable factor, we can still cover all practical scenarios by adopting the assumption that every mobile station has a traffic demand that is relatively smaller than the capacity of any base station that is able to participate in its coverage. We give two approximation algorithms for this problem. The first is a $\frac{1-r}{2-r}$-approximation algorithm for the case where each mobile station can be covered by exactly one base station (*cover-by-one*). The second is a slower, delicate refinement of the first algorithm, guaranteeing a $(1-r)$-approximate solution, that adopt the new IEEE 802.16m and LTE-Advanced possibility of simultaneously servicing a mobile clients by more than one base station (*cover-by-many*). We compare between global mechanisms that are based on our approximation algorithms and a local procedure performed by the current best-SNR greedy cell selection protocol. We show that when clients of very high bandwidth demand, relatively to the base station's capacity, exist, the use of multiple base station to satisfy the demand of a mobile station can maintain a level of at least 97% of the possible coverage - 20% better coverage than the current best-SNR greedy cell selection method. In addition to 4G networks, such relevant scenarios may be found in spread areas where there are several very small populated areas and 'standard' infrastructure is not cost-effective. In these areas, coverage can be achieved using several WiMAX-cells and situations where such cells are over-loaded may be common. Our scheme for cell selection can be used in order to allow a better utilization of these coverage solutions.

There are several interesting problems that arise from this work. The first and most important aspect is to devise an online, distributed, algorithm for doing cell selection. Such algorithms will undoubtedly make extensive use of the network initiated handoffs, which are available, e.g., under the specification of IEEE 802.16m and LTE-Advanced. Another interesting question is whether or not one can devise a constant-factor approximation algorithm to the r-AoNDM *that is independent of* r. In this respect it seems that a primary starting point for answering such a question is determining the complexity of r-AoNDM in the case where $r = 1$. A more general question is whether or not there exists a PTAS for the problem, and under which conditions. We note that by our reduction, the general case is proven to be hard to approximate for instances in which the demand of every client is strictly greater than the capacity of the base stations which can contribute to its coverage. Abusing our notation, it is unclear whether such a phase transition occurs in $r = 1$, or is there some $r > 1$ for which the r-AoNDM problem still adheres to good approximation algorithms.

Acknowledgements

The authors would like to thank the anonymous referees for their very useful comments that helped improve the manuscript. This research was partially supported by REMON - Israel 4G Mobile Consortium, sponsored by Magnet Program of the Chief Scientist Office in the Ministry of Industry, Trade and Labor of Israel.

References

[1] 3GPP TR 36.913: Requirements for further advancements for Evolved Universal Terrestrial Radio Access (E-UTRA) (LTE-Advanced), 2009.

[2] IEEE 802.16m/D4, Section 16.5, 2010.

[3] David Amzallag, Roee Engelberg, Joseph Naor, and Danny Raz. Cell planning of 4G cellular networks. Technical Report CS-2008-04, Computer Science Department, Technion - Israel Institute of Technology, 2008.

[4] David Amzallag, Michael Livschitz, Joseph Naor, and Danny Raz. Cell planning of 4G cellular networks: Algorithmic techniques, and results. In *Proceedings of the 6th IEE International Conference on 3G & Beyond*, pages 501–506, 2005.

[5] David Amzallag, Joseph Naor, and Danny Raz. Coping with interference: From maximum coverage to planning cellular networks. In *Proceedings of the 4th Workshop on Approximation and Online Algorithms (WAOA)*, pages 29–42, 2006.

[6] Matthew Andrews and Lisa Zhang. Logarithmic hardness of the undirected edge-disjoint paths problem. *Journal of the ACM*, 53(5):745–761, 2006.

[7] Reuven Bar-Yehuda. One for the price of two: A unified approach for approximating covering problems. *Algorithmica*, 27(2):131–144, 2000.

[8] Chandra Chekuri and Sanjeev Khanna. A polynomial time approximation scheme for the multiple knapsack problem. *SIAM Journal on Computing*, 35(3):713–728, 2005.

[9] Chandra Chekuri, Sanjeev Khanna, and F. Bruce Shepherd. The all-or-nothing multicommodity flow problem. In *Proceedings of the 36th Annual ACM Symposium on Theory of Computing (STOC)*, pages 156–165, 2004.

[10] Chandra Chekuri, Sanjeev Khanna, and F. Bruce Shepherd. Multicommodity flow, well-linked terminals, and routing problems. In *Proceedings of the 37th Annual ACM Symposium on Theory of Computing (STOC)*, pages 183–192, 2005.

[11] Jack Edmonds and Richard M. Karp. Theoretical improvements in algorithmic efficiency for network flow problems. *Journal of the ACM*, 19:248–264, 1972.

[12] Lisa Fleischer, Michel X. Goemans, Vahab S. Mirrokni, and Maxim Sviridenko. Tight approximation algorithms for maximum general assignment problems. In *Proceedings of the 17th ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 611–620, 2006.

[13] Stephen V. Hanly. An algorithm for combined cell-site selection and power control to maximize cellular spread spectrum capacity. *IEEE Journal on Selected Areas in Communications*, 13(7):1332–1340, 1995.

[14] Johan Håstad. Clique is hard to approximate within $n^{1-\epsilon}$. *Acta Mathematica*, 182:105–142, 1999.

[15] Ralf Irmer, Hans-Peter Mayer, Andreas Weber, Volker Braun, Michael Schmidt, Michael Ohm, Norbert Ahr, André Zoch, Carsten Jandura, Patrick Marsch, and Gerhard Fettweis. Multisite field trial for LTE and advanced concepts. *IEEE Communications Magazine*, 47(2):92–98, 2009.

[16] Bong-Ho Kim, Jungnam Yun, Yerang Hur, Chakchai So-In, Raj Jain, and Abdel-Karim Al Tamimi. Capacity estimation and TCP performance enhancement over mobile WiMAX network. *IEEE Communications Magazine*, 47(6):132–141, 2009.

[17] C.A. Mandal, P.P. Chakrabarti, and S. Ghose. Complexity of fragmentable object bin packing and an application. *Computers and Mathematics with Applications*, 35:91–97(7), June 1998.

[18] Rudolf Mathar and Michael Schmeink. Integrated optimzal cell site selection and frequency allocation for cellular radio networks. *Telecommunication Systems*, 21:339–347, 2002.

[19] Toshio Miki, Tomoyuki Ohya, Hitoshi Yoshino, and Narumi Umeda. The overview of the 4th generation mobile communication system. In *Proceedings of the 5th International Conference on Information, Communications and Signal Processing (ICICS)*, pages 1551–1555, 2005.

[20] Nir Naaman and Raphael Rom. Packet scheduling with fragmentation. In *Proceedings IEEE INFOCOM 2002, the 21st Annual Joint Conference of the IEEE Computer and Communications Societies*, pages 427–436, 2002.

[21] Maciej J. Nawrocki, Mischa Dohler, and A. Hamid Aghvami, editors. *Understanding UMTS Radio Network: Modelling, Planning and Automated Optimisation*. John Wiley & Sons, Ltd., 2006.

[22] Nokia Siemens Networks. Beyond 4G: Radio evolution for the gigabit experience (white paper), 2011. [online] http://www.nokiasiemensnetworks.com/portfolio/products/mobile-broadband.

[23] Boaz Patt-Shamir, Dror Rawitz, and Gabriel Scalosub. Distributed approximation of cellular coverage. In *Proceedings of the 12th International Conference on Principles of Distributed Systems (OPODIS)*, pages 331–345, 2008.

[24] A. Sang, X. Wang, M. Madihian, and R. D. Gitlin. A Load-aware handoff and cell-site selection scheme in multi-cell packet data systems. In *Proceedings of the IEEE 47th Global Telecommunications Conference (GLOBECOM).*, volume 6, pages 3931–3936, 2004.

[25] Aimin Sang, Xiaodong Wang, Mohammad Madihian, and Richard D. Gitlin. Coordinated load balancing, handoff/cell-site selection, and scheduling in multi-cell packet data systems. *Wireless Networks*, 14(1):103–120, 2008.

[26] David Shmoys and Eva Tardos. An approximation algorithm for the generalized assignment problem. *Mathematical Programming*, 62(3):461–474, 1993.