Integrated Optimization of Heterogeneous-Network Management and the Elusive Role of Macrocells

RAPHAEL M. GUEDES, JOSÉ F. DE REZENDE, AND VALMIR C. BARBOSA, (Member, IEEE)
Systems Engineering and Computer Science Program, Federal University of Rio de Janeiro, Rio de Janeiro-RJ 21941-914, Brazil
Corresponding author: Valmir C. Barbosa (valmir@cos.ufrj.br)

This work was supported in part by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), and in part by the Cientista do Nosso Estado (CNE) Grant from the Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ).

ABSTRACT We consider heterogeneous wireless networks in the physical interference model and introduce a new formulation of the mixed-integer nonlinear programming (MINLP) problem that addresses base-station activation and many-to-many associations while minimizing power consumption. This formulation is complete, in the sense not only of tackling activations and associations in an integrated manner, but also in that the decoding possibilities allowed in the physical interference model are fully taken into account. The formulation’s integralities and resulting non-differentiabilities are all left in place as well, so solving the MINLP problem calls for somewhat unorthodox methods. Thus, we also introduce HetNetGA, a genetic algorithm that can tackle the problem as it comes, without any approximations. Though unsuitable for practical deployment, HetNetGA enables the investigation of such networks’ true possibilities, allowing for insights that more practical approaches cannot offer. In fact, the results we give, for scenarios involving both macrocells and picocells, though often aligned with what is expected, are sometimes unexpected and essentially point to the need to better understand the role of macrocells in helping provide capacity while remaining energetically advantageous. This contrasts sharply with previous approaches, which typically always spot situations in which macrocells can be accommodated. We expect HetNetGA, with its inherent ability to lead to results such as these, to be useful in helping identify and mitigate similar issues.

INDEX TERMS Capacity allocation, heterogeneous wireless networks, physical interference model, power optimization.

I. INTRODUCTION
In order to meet the rapidly increasing demand for wireless capacity, a current tenet is that future-generation networks will have to rely on heterogeneity to grow. That is, every large base station (macrocell) deployed will have to be accompanied by a number of small base stations (picocells) spread amid users to improve capacity. The resulting network density will make the problems of managing interference, meeting capacity demands, and saving power not only more pressing but also more tightly coupled with one another and consequently harder to solve. Because of the many trade-offs involved in all decisions related to tackling these problems, at least some of their key aspects will likely be handled as a single entity. These include deciding which base stations to be turned on and which associations to establish between base stations and users. Of course, such decisions will always have to take into account the way the available spectrum is handled so that, in the end, the resulting contributions to network capacity are as strong as possible.

In the last few years, a number of works have addressed, with varying degrees of detail and success, the formulation and solution of optimization problems targeting more than one of those aspects concomitantly. The resulting literature is vast, but in our view a representative set of examples can be identified after careful examination, based not only on perceived quality of both paper and the vehicle carrying it but also on key aspects for comparison with our own work. The set we have selected comprises references [1]–[6], all of which have considered the same class of
problems as we have. While most of them differ significantly from one another, they all share some important features, including optimization problems of a mixed nature (integer and continuous) that are handled only after undergoing simplifications or having their feasible sets substantially reduced through the adoption of a limited set of so-called "patterns." Moreover, it is generally unclear how the signal-to-interference-plus-noise ratio (SINR) threshold that is typical of the physical interference model is handled. In some cases, results are downright irreproducible. We revisit these works in Section II.

In this paper we address the problem of minimizing a network's power consumption while handling the SINR threshold appropriately, determining which base stations to be turned on as well as associations between base stations and users, and meeting user demands for capacity. Each base station can be associated with multiple users, and conversely each user with multiple base stations. To the best of our knowledge, ours is a complete formulation that stands apart from previous ones in at least one feature. In particular, our approach is unique in that we handle the resulting optimization problem, with all its integralities and resulting non-differentiabilities, as it comes. Instead of bending its intrinsic combinatorial nature to fit some optimization method of choice, we leverage the inherently stochastic, parallel nature of evolutionary methods and use a genetic algorithm within a simple exploratory methodology. The resulting framework requires substantial computational resources and is not, as such, suitable for practical deployment. Instead, its value lies in the possibility it offers for insights to be gained into the true nature of the MINLP problem's feasibility landscape.

We present results, all reproducible, in some scenarios. Some of these results fall smoothly in line with what has become expected of heterogeneous wireless networks. This includes the effect of increased bit-rate demands on power consumption as well as on feasibility. Others have been unexpected, suggesting that the role of macrocells in such networks may be less clear than generally assumed thus far. In particular, for the network model and parameters used, we have been unable to unambiguously pinpoint a situation in which the combined use of macrocells and picocells would achieve feasibility while using picocells alone would not. That is, we have found no situation in which the use of macrocells would be energetically advantageous.

II. CONTRIBUTIONS

Minimizing power consumption while at the same time deciding which base stations to be turned on, deciding which associations between base stations and users to establish, and ensuring that user demands for capacity are all met usually amounts to a daunting problem, full of integralities and hence non-differentiabilities as well. Solving this problem lies at the heart of heterogeneous-network management, so for practical deployment both network models and the algorithms to be used must be simplified in order for efficiency and scalability to be achieved. The downside of such simplifications is that the operational decisions they lead to may fail to save as much power as possible or to meet user demands when they could be met. Thus, while reconciling efficiency and scalability with solution quality is unavoidably fraught with difficult trade-offs, the need remains for approaches that do not target practical deployment but rather the detailed study of a network’s true properties. In this paper we contribute one such approach by introducing a complete model (in Section III) and a formulation of the associated optimization problem that makes no simplifications (in Section IV, with the appropriate solution methodology given in Section V).

Our model is complete in the sense that it incorporates crucial elements omitted from previous models. Most notable of all is a clear treatment of how SINR thresholds are handled for proper decoding. This is lacking in previous models [1]–[6], which may result in poor interference coordination. Our model also provides for the determination of base-station activation (unlike [4]) and which users to associate with which base stations in a many-to-many fashion (unlike [4], where associations are not considered at all, and unlike [3], [5], [6], where associations are not many-to-many). As for simplifications to the optimization problem, our approach improves on previous ones by considering the model’s complete domain (unlike [1], [2], [5], where restrictions specified by the “patterns” mentioned in Section I or similar combinatorial structures are imposed), and by completely shunning any form of smoothing (unlike [1], [6], where integralities are relaxed, and unlike [4], where the functions involved are approximated) and any form of problem breakup (unlike [3]). Importantly, we have taken every possible precaution to make sure the experimental setup laid down in Section V is fully reproducible. This, too, is pointedly unlike some of the previous works (most notably the one in [1], whose results are hardly reproducible even at the level of how base stations and users are deployed).

These contributions have led to the one we consider to be most important, viz., a clear demonstration that by looking into the model’s unaltered characteristics it is possible to glean some properties that thus far have remained unobserved (or at least unreported). Specifically, for a reasonable set of parameter choices the results we give in Section VI call for a better look into the role of macrocells in heterogeneous wireless networks. This is discussed in Section VII, in which we also conclude.

III. NETWORK MODEL

We consider a set $B$ of base stations and a set $K$ of receivers. For $P_b$, the power with which base station $b \in B$ transmits, and assuming that all transmissions take place outdoors, the power $R_{bk}$ that reaches receiver $k \in K$ is

$$R_{bk} = P_b L_h^{-1} d_{bk}^{-ts},$$

(1)

where $L_h$ accounts for antenna- and frequency-related losses (as well as for frequency- or distance-unit conversion), $d_{bk}$ is the Euclidean distance between $b$ and $k$, and...
Moreover, the maximum number \( n_{\text{max}} \) of concomitant associations for a given receiver is related to \( G \) and \( \beta \) as in

\[
n_{\text{max}} < 1 + \frac{G}{\beta},
\]

so \( n_{\text{max}} > 1 \) requires \( \beta < G \). A derivation of (5) can be found on page 142 of [7].

At any given time, the total capacity provided by the network depends on how much each of the base stations in \( B_{\text{on}} \) can provide individually. It also depends on the current associations in the network. Clearly, any base station \( b \in B_{\text{on}} \) can always transmit as many bits per second as given by the greatest \( C_{bk} \) over any subset of \( K \) (since the expression in (4) comes from \( b \) and \( k \) being associated with each other). That is, letting \( K_k \subseteq K \) be the set of receivers currently associated with base station \( b \in B_{\text{on}} \), the transmission capacity of base station \( b \) is always at least \( \max_{k \in K_k} a_{bk} C_{bk} \), where \( a_{bk} \) is the fraction of time base station \( b \) spends transmitting to receiver \( k \in K_k \). On the receivers’ side, the total capacity available to receiver \( k \) is \( \sum_{b \in B_k} a_{bk} C_{bk} \), where \( B_k \subseteq B_{\text{on}} \) is the set of base stations currently associated with \( k \). Naturally,

\[
\sum_{k \in K_k} a_{bk} \leq 1
\]

and \( |B_k| \leq n_{\text{max}} \) hold at all times, respectively for every \( b \in B_{\text{on}} \) and every \( k \in K \). Additionally, meeting some demand \( d_k \) at receiver \( k \) requires

\[
\sum_{b \in B_k} a_{bk} C_{bk} \geq d_k.
\]
As a consequence, $\phi_{bk}^{tx} < 1$ is necessary for $P_b > 0$.

Given the sets $B$ and $K$ and their members’ locations in Euclidean space, as well as the values of $\gamma_0$, $\beta$, $N$, and $G$; of $L_b$, $\tau_b$, $p_{bk}^{sf}$, $\Phi_{bk}^{tx}$, and $\alpha_{bk}^{tx}$ for every $b \in B$; and of $d_k$ for every $k \in K$, the optimization problem to be solved to determine all $\alpha_{bk}$’s and all $\alpha_{bk}$’s is the following mixed-integer nonlinear programming (MINLP) problem.

\[
\text{minimize } P = p_{bk}^{sf} + p_{bk}^{tx} \tag{14}
\]

subject to
\[
\begin{align*}
& a_{bk} \in [0, 1], \quad \forall b \in B, k \in K \tag{15} \\
& \alpha_{bk} \in [0, 1], \quad \forall b \in B, k \in K \tag{16} \\
& [\text{SINR}_{bk} < \beta] a_{bk} = 0, \quad \forall b \in B, k \in K \tag{17} \\
& \sum_{b \in B} a_{bk} < 1 + G/\beta, \quad \forall k \in K \tag{18} \\
& \sum_{k \in K} \alpha_{bk} a_{bk} \leq 1, \quad \forall b \in B \tag{19} \\
& \sum_{b \in B} a_{bk} \alpha_{bk} C_{bk} \geq d_k, \quad \forall k \in K \tag{20}
\end{align*}
\]

Clearly, (17)–(20) are straightforward rewrites of the constraints given in (3) and (5)–(7), respectively, now making use of all the problem’s variables. By (10), the constraint in (19) is equivalent to $\rho_{bk} \leq 1$. It is important to note that, in reference to our discussion in Section II, the constraints in (17) and (18) are instrumental in making our formulation much more aligned with previous ones with the requirements of the physical interference model, and therefore more complete as well. The one in (17) can be read as forbidding an association between base station $b$ and receiver $k$ ($a_{bk} = 1$) when decoding is not possible ($\text{SINR}_{bk} < \beta$ or, equivalently, $[\text{SINR}_{bk} < \beta] = 1$), that is, as requiring that SINR thresholds be handled correctly. As for the constraint in (18), its role is to ensure that no receiver is to be associated with more base stations than the physical interference model allows when multi-packet reception is in effect.

We refer to any assignment of values to the $a_{bk}$’s and $\alpha_{bk}$’s satisfying the constraints in (17)–(20) as being feasible. While normally this denomination as feasible would require including the constraints in (15) and (16) as well, we note that these are always automatically satisfied in the experimental setup given in Section V. The basis for this is that the framework within which our genetic algorithm is built, detailed in that section, operates exclusively on variables in the interval $(0, 1]$. So, even though these two constraints must appear in the formal definition of the MINLP problem, omitting them when calling a value assignment feasible simplifies treatment by helping us focus on what the real difficulties are when seeking to ensure feasibility during optimization.

V. EXPERIMENTAL SETUP

We consider a two-dimensional circular region of radius $R$ in Euclidean space and place all base stations and receivers inside this region. The set $B$ of base stations has three macrocells and $|B| = 3$ picocells. The macrocells are placed at the circle’s center, each capable of transmitting only within an exclusive $120^\circ$–sector. The picocells can transmit in all directions and are placed in the circle in as uniform a manner as possible. This is achieved by mimicking the geometry of the sunflower head, followed by a rotation to ensure all sectors contain the same number of picocells (provided $|B|$ is a multiple of 3). For $n_{loc}$ the desired number of locations, this geometry is obtained by specifying the polar coordinates $r_i$ and $\theta_i$ of the $i$th location for each $i \in \{1, \ldots, n_{loc}\}$, given by $r_i = \sqrt{i/n_{loc}}R$ and $\theta_i = i\delta$, where $\delta = (\sqrt{5} - 1)2\pi/2 \approx 137.5^\circ$ is the golden (or Fibonacci) angle [9]. Receivers are placed in the same manner as picocells and are therefore to be thought of more as test points than as users. An illustration is given in Fig. 1.

All our computational results are based on using a genetic algorithm (GA) to solve the MINLP problem. We use brkgaAPI [10], an open-source, state-of-the-art framework for efficient GA implementations, and refer to the resulting GA as HetNetGA. The framework assumes all variables are continuous in the interval $(0, 1]$, which is consistent with the problem’s $a_{bk}$’s (since these can still be arbitrarily close to 0) but not with the $\alpha_{bk}$’s (since these must be either 0 or 1). In HetNetGA we circumvent this by substituting a proxy $d_{bk} \in (0, 1]$ for each $a_{bk}$ and letting $a_{bk} = [d_{bk} > 0.5]$. Moreover, all constraints must be implemented as penalties added to the objective function. We do this by keeping a count $\nu$ of constraint violations and re-expressing (14) as

\[
P = p_{bk}^{sf} + p_{bk}^{tx} + \nu p_{viol}, \tag{21}
\]

where $p_{viol}$ is the penalty to be incurred per violation. We use

\[
p_{viol} = \sum_{b \in B} (p_{bk}^{sf} + p_{bk}^{tx}), \tag{22}
\]

that is, the maximum possible value of $P$ in (14).
As a meta-heuristic, brkgaAPI can sometimes be nudged into better convergence to feasibility and subsequent optimization by tuning its behavior to problem-specific characteristics. We have found one such intervention to be particularly useful when designing HetNetGA. It consists in adding a further type of constraint to the MINLP problem in order to prevent the combined capacity available to receiver $k$ from surpassing $d_k$ by too wide a margin. For $\eta \in (0, 1)$, the further constraint for each $k \in K$ is
\[
\sum_{b \in B} a_{bk} (\alpha_{bk} - \eta) C_{bk} \leq d_k,
\] (23)
so the capacity available to \( k \), \( \sum_{b \in B} \alpha_{bk} C_{bk} \), must not exceed \( d_k \) by more than a fraction \( \eta \) of that capacity’s maximum possible value, obtained by setting every \( \alpha_{bk} \) to 1. Violating this constraint does not alter the feasibility status of any given assignment of values to the problem’s variables but does contribute to the count \( v \) affecting (21).

Tables 1 and 2 contain all parameter values used to obtain the results given in Section VI. Table 1 refers to the formulation of the MINLP problem, including the additional constraint in (23). The values for \( L_b \) (for a center frequency of 2 GHz and \( d_{bk} \) in kilometers) and \( t_b \) are from [11]. The values for \( P_{b}^k \), \( \phi_{b}^k \), \( P_{b}^x \), and \( \phi_{b}^x \) are loosely based on the discussion in [8]. By (13), we get \( P_{b} = 39.75 \) W if \( b \) is a macrocell, \( P_{b} = 1 \) W if \( b \) is a picocell. Likewise, by (22) we have \( P^{\text{viol}} = 1500 + (|B| - 3)33 \) W. The values for narrowband \( N \) and gain \( G \) imply a wideband \( W = 1.28 \) GHz.

Table 2 refers to the inner operation of brkgaAPI and its use by HetNetGA: \( p \) is the number of individuals (or chromosomes, in GA parlance) a population has, given in proportion to the number \( n_{\text{var}} = 2|B||K| \) of variables (or alleles); \( p_e \) is the fraction of \( p \) to be the elite set; \( p_m \) is the fraction of \( p \) to be replaced by mutants; \( p_a \) is the probability of inheriting each allele from the elite parent; \( n_{\text{pop}} \) is the number of independent populations; and \( n_{\text{gen}} \) is the number of generations allowed to elapse before termination. Each individual is an assignment of values to the problem’s variables. Owing to the strictly positive value of \( p_e \), HetNetGA gives rise to an individual that minimizes \( P \) in (21) globally with as high a probability as one wishes, provided the allotted \( n_{\text{gen}} \) is sufficiently large [12]. Of course, we have no efficient means of verifying whether this occurs for any given \( n_{\text{gen}} \), only of checking individuals for feasibility.

As mentioned in Section I, the setup detailed in this section allows for full reproducibility of the results we present next, and for further experimentation as well. Below are the key steps to be followed.

1) Choose the radius \( R \) of the two-dimensional circular region. We used \( R = 0.3 \) km.

2) Place the three macrocells at the region’s center, their associated 120°-sectors as in Fig. 1. For \( n_{\text{loc}} = |B| - 3 \) and \( i \in \{1, \ldots, n_{\text{loc}}\} \), let the polar coordinates of the \( i \)th picocell be \( r_i = \sqrt{|B|/n_{\text{loc}}}R \) and \( \theta_i = i(\sqrt{3} - 1)^2\pi/2 \). We used \( |B| = 15 \). (If \( |B| \) is a multiple of 3, a simple program in C is available from the authors upon request to optimally rotate the picocells so that each 120°-sector contains exactly \( n_{\text{loc}}/3 \) picocells.)

3) Place the \( |K| \) receivers in the same manner as the picocells, using \( n_{\text{loc}} = |K| \). We used \( |K| = 51 \).

4) Download brkgaAPI [13] and express the MINLP problem’s objective function and constraints in C++ as exemplified in the samples that come with the download, following (14), (17)–(20), and (23).

5) Select values for both the MINLP problem’s parameters and those related to brkgaAPI. We used the ones given in Tables 1 and 2.

VI. RESULTS

All our results refer to the setting depicted in Fig. 1. Within this setting we investigate five distinct scenarios, each forbidding certain base stations to ever be turned on. This can be enforced for any \( b \in B \) by overriding Table 1 and setting \( P_{b}^x = 0 \), thus leading to \( \text{SINR}_{bk} = 0 \) for every \( k \in K \), and consequently to \( \alpha_{bk} = 0 \) (so \( \phi_{b} = 0 \) = 0) whenever feasibility holds; cf. (17). We refer to the first scenario as 0m12p (no macrocells can ever be turned on, all picocells can), and similarly for the other four scenarios: 1m12p, 2m12p, 3m12p, and 3m0p. In relation to Fig. 1, the shorthand 1m refers to the upper right sector, 2m to the two upper sectors. Notably, when no picocells are allowed to be turned on (scenario 3m0p), by (2) it follows that \( \text{SINR}_{bk} \) does not depend on the value of \( G \) for any macrocell \( b \) or any \( k \in K \).
In what follows, every value we report for $P$, the total power consumed, is for feasible individuals and given by (14). That is, having $v > 0$ in (21) for such an individual implies violated constraints only of the type given in (23). Our results are summarized in Figs. 2 and 3, where information related to the output of HetNetGA is given for all five scenarios and five values of demand $d_k$, the same for every $k \in K$. For the reader’s benefit, these figures are complemented by Fig. 4, which relates our results to both the layout in Fig. 1 and the statistics in Fig. 2.

VII. DISCUSSION AND CONCLUSION

As expected, increasing $d_k$ increases total power consumption as well (Fig. 2), and moreover makes feasibility ever harder to attain (Fig. 3). Additionally, a larger $d_{bh}$ tends to be necessary as base station $b$ and receiver $k$ are placed farther apart from each other (Fig. 4). Unexpectedly, though, feasibility seems to become impossible for scenario 3m0p somewhere between $d_k = 6$ and 9 Mbps, and for the other scenarios involving one or more macrocells (1m12p, 2m12p, and 3m12p) somewhere between $d_k = 12$ and 15 Mbps (Fig. 2). This rules out the use of macrocells for higher bit-rate demands, those for which picocells alone will not do (this holds already for $d_k = 15.5$ Mbps; data not shown). Perhaps some sweet spot exists at which this happens, but locating it has proven elusive. As far as we have been able to observe, any assistance a macrocell might provide in meeting a certain bit-rate demand is offset by the interference it causes, and then the whole setting becomes energetically disadvantageous.

Naturally, the flip side of this conclusion is that HetNetGA, in spite of its properties of convergence to a global minimum, is after all the one to blame. Some support against this possibility is given in Fig. 3, which shows how early feasibility is first attained during the $n_{gen}$ generations. This happens ever later as $d_k$ increases, and also with confidence intervals much greater than those of Fig. 2. The suggestion here is that, notwithstanding all the variation in the number of generations to hit first feasibility, by the $n_{gen}th$ generation the solutions HetNetGA outputs are approximately equivalent to one another. This hardly rules out the above-mentioned sweet spot, but does make it hard enough to find to cast doubt on the practicality of looking for it.

Thus, insofar as the model outlined in Section III can be said to describe the system under study faithfully, a role is yet to be found for macrocells. Further research should concentrate on variations of this model, and also of its characteristics as summarized in Table 1, aiming to better delimit what can be expected of macrocells. HetNetGA, which inherently preserves the mathematical description of the associated MINLP problem, is expected to remain a useful tool. For example, HetNetGA seems particularly well equipped to help identify and mitigate fundamental infeasibilities at the earliest stages of a new idea’s development. In particular, techniques such as coordinated multi-point transmission and its variations [14]–[16], all of which target problematic issues similar to the ones we have highlighted in relation to macrocells, have complex optimization problems associated with them. Knowledge of these problems’ feasibility landscapes could improve significantly with the detailed scrutiny that HetNetGA is capable of affording.

To sum up, the results we have obtained are fundamentally distinct from those of similar previous studies of heterogeneous-network management, that is, those that aim to select base stations for activation and to decide how base stations and users should be associated with one another while meeting user demands and using energy as sparingly as possible. Such distinction stems from disparate objectives: while others have resorted to simplifications in problem formulation and to approximations during solution, aiming for the resulting approaches to be of practical use, we have tackled a complete variation of the problem as it comes and used a solution method that is theoretically capable of reaching optimality even so. Our results are negative and run contrary to the common wisdom in this type of study, which is that conditions exist for macrocells to step in when picocells alone are not sufficient. This is seen in the previous works discussed in Section II, which along with many others provide a broad baseline against which to compare our results. That we found no such conditions in a simple yet reasonable setting is indication that prospective studies of a problem’s feasibility landscape have an important place in the area and should be pursued.

REFERENCES


R. M. Guedes et al.: Integrated Optimization of Heterogeneous-Network Management and Elusive Role of Macrocells


RAPHAEL M. GUEDES received the degree in telecommunications engineering from the State University of Rio de Janeiro, in 2006, and the M.Sc. and D.Sc. degrees in electrical engineering from the Federal University of Rio de Janeiro (UFRJ), in 2009 and 2015, respectively. He then became a Postdoctoral Researcher at UFRJ’s Systems Engineering and Computer Science Program, where he continues to collaborate actively. His research interests include wireless networks, networks, cognitive radios, and heterogeneous networks.

JOSÉ F. DE REZENDE received the Ph.D. degree in computer science from the Université Pierre et Marie Curie, in 1997. He was an Associate Researcher at the Université Pierre et Marie Curie, in 1997. Since 1998, he has been an Associate Professor at the Federal University of Rio de Janeiro. His research interests include distributed multimedia applications, QoS in the internet, mobile networks, wireless communication, and experimental platforms.

VALMIR C. BARBOSA (Member, IEEE) received the Ph.D. degree in computer science from the University of California at Los Angeles, in 1986. He is currently a Professor at the Federal University of Rio de Janeiro. He is the author of An Introduction to Distributed Algorithms (The MIT Press, 1996) and two other books. His research interests include the locality and complexity aspects of natural and computational distributed systems. He is a member of the Brazilian Academy of Sciences. He has served in the editorial boards for Parallel Computing, the IEEE Transactions on Computers, and the Journal of the Brazilian Computer Society.

 delay/disruption-tolerant networks.