A Sharing- and Competition-Aware Framework for Cellular Network Evolution Planning

(Article begins on next page)

A Sharing- and Competition-Aware Framework for Cellular Network Evolution Planning

Paolo Di Francesco, *Student Member, IEEE,* Francesco Malandrino, Tim K. Forde,
Luiz A. DaSilva, *Senior Member, IEEE*
(Invited Paper)

Abstract—Mobile network operators are facing the difficult task of significantly increasing capacity to meet projected demand while keeping CAPEX and OPEX down. We argue that infrastructure sharing is a key consideration in operators' planning of the evolution of their networks, and that such planning can be viewed as a stage in the cognitive cycle. In this paper, we present a framework to model this planning process while taking into account both the ability to share resources and the constraints imposed by competition regulation (the latter quantified using the Herfindahl index). Using real-world demand and deployment data, we find that the ability to share infrastructure essentially moves capacity from rural, sparsely populated areas (where some of the current infrastructure can be decommissioned) to urban ones (where most of the next-generation base stations would be deployed), with significant increases in resource efficiency. Tight competition regulation somewhat limits the ability to share but does not entirely jeopardize those gains, while having the secondary effect of encouraging the wider deployment of next-generation technologies.

Index Terms—Cellular networks, Network sharing, Network planning, Real-data

1 Introduction

With cellular coverage reaching virtually every human being on the planet, the challenge now lies in *evolving* the existing infrastructure to cope with the foreseen explosion in the demand for capacity [1]. Evolution will mean different things at different locations: some parts of the infrastructure will be replaced with new-generation equipment, e.g., LTE and its successors; others will be upgraded for the purpose of enhancing capacity, e.g., by increasing sectorization; finally, underutilized base stations will be decommissioned, possibly permanently, as part of the network consolidation process.

The yellow, solid curve in Fig. 1 represents the network load and its familiar predicted almost-exponential growth from the current level in A to a future one in Ω [1]. Dashed lines represent possible evolutions of the network capacity: there is no question that capacity has to increase from its current level at point B to Ω, so as to serve all the demand; in this paper we investigate *how*

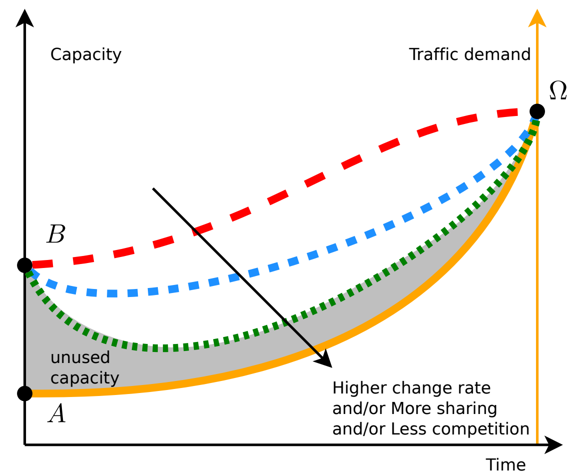


Fig. 1. The present and future evolution of cellular infrastructures. The solid yellow line represents the traffic demand, growing from its present value in A to a much higher one in Ω. Dashed lines represent different evolutions of network capacity. Its present value (in B) is higher than the current demand A, but lower than the future demand Ω. The area between the capacity and demand curves corresponds to unused capacity (shaded in gray). Network capacity therefore has to grow from B to Ω in the long term, possibly decreasing in the short term in order to reduce the amount of unused capacity.

and *when* the required changes to the network shall be performed, so as to efficiently match available capacity to demand, minimizing over-provisioning.

The reason why this matters is the gray area in Fig. 1, representing unused network capacity. Providing capacity that nobody uses is a waste of bandwidth, resources and, ultimately, money; therefore, it is of paramount importance for operators to keep overprovisioning as low as possible. It is possible to exploit this overprovisioning through, for example, 3G onloading [2], whereby wired connectivity is augmented by cellular links. Nevertheless, it is in operators' interest to minimize costs, and therefore, deploy wireless capacity only when and where it is needed. Furthermore, Fig. 1 refers to the network as a whole; the relative positions of points A, B, Ω can be

• *All authors are with CONNECT, University of Dublin, Trinity College, Ireland.*

This publication has emanated from research supported in part by research grants from Science Foundation Ireland (SFI) under Grant Numbers 10/IN.1/I3007 and 13/RC/2077.

different in different parts of the topology [3]. Extreme cases include some sparsely populated rural areas, where the current capacity may exceed not only the current but also the future demand, i.e., $B > \Omega$, and very dense urban areas, which may have $A \approx B$.

Ideally, operators would like their capacity to instantly fall from B to A at all locations, and would achieve this by decommissioning as many base stations as possible. Then, they would follow the demand curve all the way to Ω, by updating their infrastructure as the load increases, always keeping the unutilized capacity (i.e., the gray area in Fig. 1) to zero.

Such an idealized view conflicts with the reality that making any change to a network, be it deploying new base stations, updating or decommissioning existing ones, requires equipment, work-power, and funds – all resources that are scarce, and whose usage must be carefully planned. The number of such changes operators can perform in a given time, e.g., a month, is typically limited, and such a limit directly impacts the speed at which network capacity can go down or up, hence the gray area in Fig. 1.

Our first goal is then to study the efficient evolution of the cellular infrastructure in light of the limited budget of possible network changes. The input to our problem consists of the current set of base stations, the (projected, possibly based on real-world measurements and topologies) future demand and a limited *change rate* at which an operator can deploy, update, replace or decommission base stations. This *change rate* reflects the operators' limitations in terms of how they are able to reshape their own network. The output we seek is a list of the changes to perform to the network, and the time at which to enact each of them. The overall objective is to keep unused capacity, i.e., the gray area in Fig. 1, at a minimum.

In studying this problem we account for two important real-world issues: network *sharing* and competition *regulation*. Both are widely studied in the literature, but their impact on the evolution of cellular networks has received relatively little attention so far.

Active network sharing [4], [5] refers to roaming-like agreements between mobile operators, where users of each operator are served through both networks indifferently. Each network operator retains ownership and control over its own spectrum. Active sharing is emerging as a promising way to achieve cost savings and enhanced performance; indeed, running their networks in such a shared fashion makes it easier for operators to identify underutilized base stations to decommission, as well as making the most out of updated, more highly performing infrastructure.

Network sharing agreements, in which operators actually behave as one, decrease the level of market competition as, intuitively, users have less choice. To offset this effect, regulators often require operators to leave some spare capacity, so as to allow new market players (typically, virtual MNOs) to enter the market, as recently happened when O2 and Three merged their Irish branches [6].

Studying how sharing and competition regulation shape the evolution of networks is thus an important contribution of this paper.

The algorithms we present are most readily separated into three phases: meeting demand, regulation compliance, and cost reduction. Each of the phases occurs in series to update the network of an operator in order to provide service to increasing demand while minimising over-provisioning.

In our view, each individual phase conforms to an instantiation of the cognition loop. Specifically, each phase of the operation has observation, decision, and action steps. During observation the current situation is assessed in terms of the current network, the current demand, the already planned updates, and the expected demand. Decision involves the application of this situational awareness to some optimization. Actions take the form of changes to the schedule of network updates. Note that none of the phases explicitly involves learning. Rather, each phase implements cognition as optimization based on situational awareness of network and subscriber state.

Furthermore, the collection of all the phases together provides a more nuanced form of cognition. This cognition uses understanding of current network infrastructure to plan future deployments based on the input of expected demand and regulatory policy. As a unit the three phases of our algorithms periodically receive an observation of projected demand, whereupon a plan for network updates is constructed. This plan is then used to decide which base station updates should be applied to the current infrastructure and the action of making these adjustments is taken.

In this paper we propose a framework to study planning decisions for mobile network operators while taking into account how different aspects such as the ability to share network resources, and the constraints imposed by competition regulation impact the overall process. This work makes the following contributions:

(i) We propose a flexible framework that takes into account several actions such as deploy, update, replace or decommission mobile network infrastructures.
(ii) We include in our study the impact of competition regulations as expressed by the Herfindahl–Hirschman Index (HHI).
(iii) We propose a family of algorithms to schedule the changes to a network in a cost efficient way while satisfying the demand and complying with regulatory constraints.
(iv) We introduce a new way to exploit operators' deployment and traffic data in conjunction with demographic data.

The remainder of this paper is organized as follows. After a review of related work in Sec. 2, we present our system model in Sec. 3. In Sec. 4, we state and solve the problem of scheduling the changes to our network, and discuss its computational complexity and performance in Sec. 5. Sec. 6 presents our reference scenario, which we use to obtain the results we present in Sec. 7. Finally, Sec. 8 concludes the paper.

2 RELATED WORK

Our work studies the effects of infrastructure sharing and competition regulation on cellular network planning.

The classic network planning problem usually considers a single operator that aims at minimizing operational costs for a specific technology (e.g. 3G, LTE) while maintaining acceptable subscribers' satisfaction both in terms of coverage and capacity. The research following this line is vast and includes various aspects. For example, in [7], [8], [9], [10], [11], [12], [13] the authors studied the optimization of base station location in an area of interest. Some of these works [7], [8], [9], [10] dealt with 3G systems and were based on meta-heuristics aiming at minimizing the number of base stations to be deployed. Other more recent works [11], [12], [13] have focused on the same objective using similar approaches but on LTE networks. The work in [12] in particular used a model based on stochastic geometry, and the coverage probability as the metric to optimize. The problem addressed in our paper is fundamentally different from the ones addressed by the aforementioned works since it includes sharing of existing as well as to-be-deployed infrastructure by more than one operator and considers the impact of competition regulation, modeled similarly as imposed by the Irish regulator recently [6], on network planning.

Resource sharing among multiple operators in cellular networks, in fact, is another important aspect of our work and it has been studied in [3], [14], [15], [16]. In [14] the authors analyzed feasible sharing options in the near-term in LTE using co-located and non co-located base stations. Authors in [15] assessed the benefit of sharing both infrastructure and spectrum, using real base station deployment data. In [3] the authors have studied sharing opportunities between two operators, by looking at spatial variation in demand peaks. All the results obtained in these works suggested that resource sharing, whether spectrum or infrastructure, increases the network capacity and the ability to satisfy users' requirements. However none of them addressed the problem of *how* to efficiently plan a shared network comprising resources already deployed by existing operators. In our previous work [16] we investigated the *coverage efficiency* obtained by combining existing cellular networks considering the coverage redundancy in real deployments in Poland but, unlike this paper, our earlier work did not address capacity requirements nor regulatory constraints.

Demographic data are an important source of information for operators to make planning decisions. Combining such data with the traffic demand information at their base stations, operators can have a clearer picture of the needed planning interventions. Researchers rarely use real topologies and demographic data; they typically rely on simplified synthetic topologies often featuring regular, lattice-like deployments that do not reflect the complexity and heterogeneity of cellular networks. Among the studies that used real data that are related to our work are [16] and [15], which however did not use traffic demand data.

Other works used traffic demand, either concentrating on profiling the *users* [17], [18], [19], or focusing more on the *network* behaviour [19], [20], [21], [22] but none of them investigated long term planning decisions. The work in [20] is of interest to us because it had an objective orthogonal to ours; it was focused on energy savings and green networking and envisioned dynamically switching off base stations at off-peak times in different parts of the topology. Our work instead aims at reconfiguring the whole network, operating long-term (possibly, permanent) changes in its infrastructure. It is important to stress that the two approaches can, and indeed should, coexist: once we reconfigure the infrastructure through our network sharing scheme, we can manage it in an energy-efficient fashion.

Works such as [23] focused on economic aspects of resource sharing from a network virtualization perspective. It described the incentives operators have to pool their resources together. From an implementation perspective, the analysis of cooperative sharing arrangements presented in [24] highlights the diversity of approaches currently being used in existing networks, their successes and failures. It pointed to a process of learning within industry as to which sharing modes allow for both competition and cooperative sharing to thrive. A study of Pakistan's experience of network sharing indicated the varying economic gains made in a still-developing market by adopting different sharing strategies [25].

However, while these papers addressed some of the economic effects of sharing, none of these has dealt with regulatory concerns regarding the ensuing market concentration which occurs through network sharing in mature mobile markets. Our study is unique because it combines all the aforementioned aspects (real data analysis, spatial distribution of traffic, network sharing, and network planning), and it considers the impact of the limitations imposed by the regulators on the savings when managing two networks in a shared fashion in a mature market.

3 SYSTEM MODEL

Our system model revolves around two main elements: *base stations* and *subscriber clusters*.

Model elements Base stations $b \in \mathcal{B}$ are elements of the infrastructure with a certain position, capacity, and coverage area. Subscriber clusters $c \in \mathcal{C}$ can correspond to one or more actual users, which can be viewed as co-located. They have a known position and traffic demand.

We also have operators $o \in \mathcal{O}$, and time periods $k \in \mathcal{K} = \{1 \dots K\}$.

Base station type Base stations are not all equal. They can differ in technology, e.g., GSM, 3G, or LTE; furthermore, even base stations with the same technology differ in such aspects as frequency of operation and sectorization. Indeed, upgrading a cellular network essentially means changing their type, e.g., from 3G to LTE. Even decommissioning a base station can be seen as changing its type to "off".

In our model, possible base station types are collected in set $\mathcal{T}$, and every base station $b \in \mathcal{B}$ has a type $T(b) \in \mathcal{T}$. Decommissioned base stations have the special type $t_{\emptyset}$. The type of a base station determines its coverage and performance, as shown next, as well as its associated cost.

Coverage and demand Our coverage information comes in the form of binary flags $\gamma(b, c, t) \in \{0, 1\}$, expressing whether base station b covers subscriber cluster c if b is of type t, i.e., if $T(b) = t$. Notice how these values do not depend upon time. We indicate with $\delta(b, c) \in \mathbb{R}$ the received power (RSSI) from base station b at subscriber cluster c.

Requested and served traffic For each subscriber cluster c, operator o and time period k, we know the traffic demand $\tau(c, k, o)$ from users of operator o in cluster c at period k. To streamline the notation, we will often write $\tau(c, k) = \sum_{o \in \mathcal{O}} \tau(c, k, o)$, indicating the combined traffic demand of the multiple operators expressed in Mbit.

We also indicate with $\sigma(b, c, k, o, t)$ the traffic demand that can be met by base station b, of type t, when serving users of operator o in cluster c at period k. Similarly to τ, we will often drop indices to streamline the notation, and write, e.g., $\sigma(c, k) = \sum_{o \in \mathcal{O}} \sum_{b \in \mathcal{B}} \sum_{t \in \mathcal{T}} \sigma(b, c, k, o, t)$. In the case of a sharing agreement, the combined traffic demand $\tau(c, k)$ can be served simultaneously by base stations belonging to different operators ($\sigma(c, k)$), while if no-sharing agreements are in place, each operator only serves its traffic demand using its own base stations.

Cost Base stations also have an *operational cost* $\kappa(b, T(b))$. Such a cost is base station- and type-dependent, and models such aspects as maintenance, site rental, and energy consumption. The cost associated with decommissioned base stations is zero, i.e., $\kappa(b, t_{\emptyset}) = 0, \forall b \in \mathcal{B}$.

Network changes All our decisions concern network changes. At each time period k, we may decide to *change* the type of base station b, either to a better-performing type t_{dest}, in order to increase its capacity, or to $t_{\emptyset}$ to save on costs, as shown in Fig. 2. We track type changes through binary variables:

$$x(b, k, t_{dest}) \in \{0, 1\}.$$

Setting $x(b, k, t_{dest}) = 1$ means that, at time k, we change the type of base station $b \in \mathcal{B}$ to $t_{dest} \in \mathcal{T}$. Doing nothing, i.e., never changing b's type, is represented by having $x(b, k, t) = 0, \forall k, t$.

Base station	$k=1$	$k=2$	$k=3$	$k=4$
b_1		$x(b_1, 2, t_2) = 1$		$x(b_1, 4, t_3) = 1$
b_2	$x(b_2, 1, t_2) = 1$			
b_3			$x(b_3, 3, t_{\emptyset})$=1	

Fig. 2. An example schedule, with $|\mathcal{B}| = 3$ base stations and $|\mathcal{K}| = 4$ time periods. The maximum change rate is $N = 1$, i.e., we can make at most one change (updating or decommissioning a base station) per time period. Let us assume $\mathcal{T} = \{t_{\emptyset}, t_1, t_2, t_3\}$, with $t_1 \dots t_3$ having increasing capacity and the same coverage. All base stations start with type t_1. Green arrows mark the periods at which base stations need to be updated due to an increased load, i.e., because $\tau > \sigma$; some stations, such as b_1, may need more than one update. We update b_1 twice, from t_1 to t_2 and then to t_3, setting $x(b_1, 2, t_2) = x(b_1, 4, t_3) = 1$. Base station b_2 needs an update within $k = 2$; However, we cannot set $x(b_2, 2, t_2) = 1$, because doing so would violate constraint Eq. (1). This forces us to anticipate the update to $k = 1$, i.e., set $x(b_2, 1, t_2) = 1$. Similarly, the red arrow tells us that we would be able to decommission b_3 as soon as $k = 1$, but the updates we already scheduled force us to delay until $k = 3$, and set $x(b_3, 3, t_{\emptyset}) = 1$.

Also notice that we can change a base station's type multiple times, i.e., it can be that $\sum_{k \in \mathcal{K}, t \in \mathcal{T}} x(b, k, t) > 1$. We do not explicitly forbid changing the type to $t_{\emptyset}$ and then to some other type, i.e., first decommissioning and then re-enabling a base station, although it is unlikely to make sense in practice (and we never observe that behavior in our performance evaluation).

Changing the type of a base station to $t_{\emptyset}$ means cost savings while reducing network capacity, i.e., going down in Fig. 1. On the contrary, moving to a better-performing type means being able to serve more traffic, hence going up in Fig. 1.

In both cases, as discussed in Sec. 1, setting more x-values to 1 is linked to going up or down in Fig. 1 with a higher slope, i.e., being more effective in reducing the gap between requested and provided capacity.

What limits us is the maximum change rate N, defined as the maximum number of changes we can make to our network at each time period k. The following constraint must hold:

$$\sum_{b \in \mathcal{B}, t \in \mathcal{T}} x(b, k, t) \leq N, \forall k \in \mathcal{K}. \quad (1)$$

Having Eq. (1) in place means two things. First and most obviously, we can take fewer actions, e.g., decommission fewer base stations. Furthermore, we may have to move some actions in time, e.g., decommission a base station later than we would like to. Both make us less effective in tracking the demand, i.e., imply a larger gray area in Fig. 1.

Time scale It is important to understand the time scale at which our model, and the algorithms described later, work. We are modeling network *planning*, and we are concerned with the evolution of our network over a time span of months or years. Each time period $k \in \mathcal{K}$ may correspond to several weeks, and the decisions $x(b,k,t)$ can be mapped, e.g., to equipment orders or to the schedule of infrastructure deployment teams. Decommissioning or updating a base station is substantially different from turning it on and off in order to follow daily traffic fluctuation, as envisioned in “green networking” solutions [20], [26], [27]. Indeed, as discussed later, the two solutions are orthogonal and altogether compatible.

Since networks have to be provisioned for peak loads and not average ones, the $\tau(c,k)$ values express the *worst-case* amount of traffic requested by subscriber cluster c during the whole duration of time period k – e.g., the amount of traffic (in Mbit) that users in c will need served during the busiest hour of time period k. Such values typically come from forecasts and projections; from the viewpoint of our model, they are an input.

It is also worthwhile to observe that our network must be able to operate even if all the “worst hours” of all subscriber clusters take place at the same time. In other words, while it is possible, and indeed advisable, to *operate* the network so as to take advantage of the low space and traffic correlation in traffic demand [3], such an effect cannot be depended upon in the *planning* phase.

Assessing network performance It is important to remark that our model does not explicitly include a representation of how the traffic demand that can be met σ depends on the other parameters and variables, e.g., our decisions x. As we see in Fig. 3, the σ-values are obtained through an external performance assessment block. In addition to keeping our model simple, this choice affords us a higher degree of flexibility: we can interface our model with a simulation tool, or leverage any real-world data available to us, as discussed in Sec. 6.

Competition Healthy competition within the mobile market is of constant concern to regulators, as dominance on the part of large operators can lead to market abuses. A common regulatory tool to measure the level of competition and market concentration is the Herfindahl–Hirschman Index (HHI) [28], [29]. It is given by the sum of the squares of shares held by each operator in the market, and takes values between 0 (a multitude of operators with a zero-share) and 1 (a monopolist with a 100% share). The HHI can be used to assess concentration in different aspects of the market, e.g., overall market share, concentration in ownership of spectrum and concentration in ownership of network infrastructure.

In our scenario, we need to define a *local* version of HHI, specific to each subscriber cluster (as well as to each time period). Furthermore, we have to account not

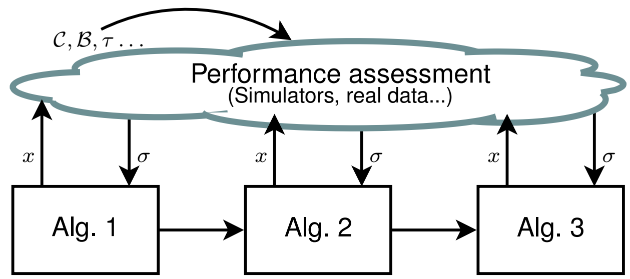


Fig. 3. Assessing the performance within our model and solution concept. Network performance depends on topology, traffic demand and decisions; however, our model does not explicitly represent such a dependency. Our algorithms rely instead on an external block, represented by the cloud in the figure. Given as an input the network topology (base stations $\mathcal{B}$, subscriber clusters $\mathcal{C}$), demand τ, and decisions x, it returns the traffic that can be served by the network, σ, as an output. The most straightforward way of implementing such a block is network simulation; however, other approaches are possible. As discussed in Sec. 6, we implement the performance assessment block leveraging real-world traces.

only for the operators currently in $\mathcal{O}$, but also for new operators that may enter the market if the conditions are favorable – typically mobile virtual operators (MVNOs). Our version of the HHI is thus given by:

$$H(c,k) = \left(\frac{\sigma(c,k) - \tau(c,k)}{\sigma(c,k)}\right)^2 + \sum_{o \in \mathcal{O}} \left(\frac{\tau(c,k,o)}{\sigma(c,k)}\right)^2 \quad (2)$$

In the denominator of Eq. (2) we always find the total capacity $\sigma(c,k)$ available to subscriber cluster c at time period k (recall our conventions about dropping indices). In the numerator we have the traffic demand faced by current operators in $\mathcal{O}$ in the summation, and the spare capacity, i.e., the traffic of potential new operators, in the other term.

When two operators deploy and manage their networks in a shared fashion, they behave as one from the competition viewpoint. Therefore, the set $\mathcal{O}$ shrinks, and the HHI in Eq. (2) increases. We apply the HHI methodology only to a dominant player, created by merging two mobile operators assuming that the average subscriber of the merged operator generates the same traffic. In our model, regulators require that the HHI not exceed a value $H_{\max}$ in at least a significant portion of the topology.

4 Problem formulation and solution

In this section, we address the following problem. Given the future demand $\tau(c,k,o)$, and the maximum change rate N, how should each operator schedule the network changes, i.e., set the x-variables? Operators have three goals:

Goal 1 is meeting the *traffic demand*, i.e., having:

$$\sigma(c,k,o) \geq \tau(c,k,o), \forall c \in \mathcal{C}, k \in \mathcal{K}, o \in \mathcal{O}. \quad (3)$$

Eq. (3) says that for all subscriber clusters c, time periods $k \in \mathcal{K}$ and operators $o \in \mathcal{O}$, the provided capacity σ must equal (or exceed) the demand τ.

Goal 2 is complying with existing regulation:

$$\sum_{c \in \mathcal{C}} \mathbb{1}_{[H(c,k) \leq H_{\max}]} \geq \phi \cdot |\mathcal{C}|, \forall k \in \mathcal{K}. \tag{4}$$

Eq. (4) imposes that, for each time period, at least a fraction ϕ of demand clusters – enough for a new operator to start building its network [6], [25] – have an HHI (as defined in Eq. (2)) not exceeding the limit $H_{\max}$.

Goal 3 is to minimize costs:

$$\min \sum_{b \in \mathcal{B}, k \in \mathcal{K}} \kappa(b, T(b)). \tag{5}$$

An obvious way to decrease the quantity in Eq. (5) is setting the type of some base stations to $t_{\emptyset}$, whose associated cost is 0, i.e., disabling them.

Multi-objective problems, where some kind of trade-off between different goals is sought, are in general hard to formulate and harder to solve. Thankfully, in our case goals have a clear hierarchy: the first two goals must be met through as few changes to the network as possible; any remaining change can be used to pursue the third goal. Indeed, the first two goals can be treated as constraints, and the third one is the objective we seek to optimize.

4.1 Solution concept

Our aim is to exploit the hierarchy of the goals stated above, as well as their features, to devise a solution concept that addresses them in sequence.

We begin by defining a class of network changes, that we call *capacity-preserving changes*, as follows:

Definition 1. *Changing the type of a base station b from t_{orig} to t_{dest} is* capacity-preserving *if the capacity available to each subscriber cluster does not decrease, i.e.,*

$$\pi(b, t_{orig}, t_{dest}) = 1 \Leftrightarrow \sigma(b, c, t_{dest}) \geq \sigma(b, c, t_{orig}), \forall c \in \mathcal{C}. \tag{6}$$

Intuitively, capacity-preserving network changes increase the capacity available to certain subscriber clusters, without hurting others. Notice that capacity-preserving changes are also coverage-preserving. Increasing the number of sectors of a base station is a capacity-preserving change, as is replacing a GSM base station with an LTE-800 one, having the same coverage and a higher capacity. Replacing the same GSM base station with an LTE-2600 one is not capacity-preserving, as the new base station will have smaller coverage and some subscriber clusters, namely the ones covered by the old base station but not by the new one, will suffer a decrease in their available capacity. Similarly, changing any base station's type to $t_{\emptyset}$ is not capacity-preserving.

We are now in the position of proving the following useful properties:

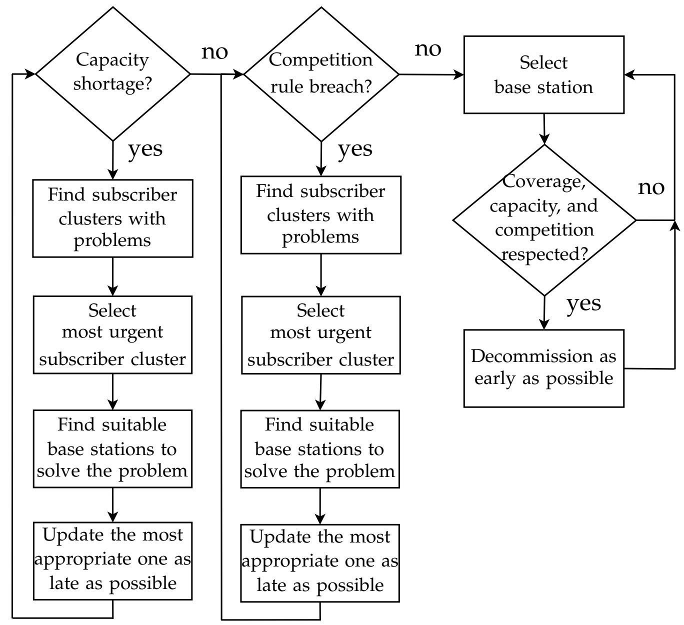


Fig. 4. Our solution concept. The three phases correspond to our three objectives – meeting the traffic demand, complying with competition regulation, and reducing costs. Within each phase, we proceed in a similar way: identify the subscriber clusters that call for action, e.g., whose demand is not satisfied; identify the base stations that can be acted upon to solve the problem, i.e., whose type shall be changed, and schedule the needed action at the most appropriate time. Notice that once we are done with a phase we never come back to it, and subsequent decisions are guaranteed not to jeopardize its objective.

Property 1. *Both goal 1 and goal 2 can be reached through capacity-preserving changes alone, i.e., changes that comply with Eq. (6).*

Proof: See the Appendix. □

Property 2. *If the initial configuration satisfies goal 1, then pursuing goal 2 by scheduling further capacity-preserving changes does not compromise goal 1.*

Proof: See the Appendix. □

Exploiting these properties, we propose the solution concept shown in Fig. 4, where objectives are addressed in sequence. Specifically, we first address goal 1, and do so by scheduling capacity-preserving network changes (Property 1 guarantees that it is sufficient). Then, we schedule further capacity-preserving changes in order to reach goal 2: Property 1 again guarantees that it is possible, and Property 2 makes sure that doing so will not jeopardize goal 1. Finally, we use any remaining changes we can make to the network to pursue goal 3, as long as doing so does not conflict with goals 1 and 2. Notice that in this last step we are not restricted to capacity-preserving changes, e.g., we can decommission base stations by changing their type to $t_{\emptyset}$.

With reference to Fig. 4, we can clearly see how the

Algorithm 1 Phase 1: ensuring that traffic demand is met.

Require: $\mathcal{B}, \mathcal{C}, \mathcal{K}, \tau$

1: **while true do**
2: $\quad \sigma \leftarrow \texttt{assess}(x, \delta, \gamma, \tau)$
3: $\quad \texttt{problems} \leftarrow \{(c,k) \in \mathcal{C} \times \mathcal{K} : \sigma(c,k) < \tau(c,k)\}$
4: $\quad$ **if** $\texttt{problems} = \emptyset$ **then**
5: $\quad\quad$ **break**
6: $\quad c^\star, k^\star \leftarrow \arg\min_{(c,k) \in \texttt{problems}} U(c,k)$
7: $\quad \texttt{actions} \leftarrow \{(b,t) \in \mathcal{B} \times \mathcal{T} : \gamma(b, c^\star, t) = 1 \wedge \pi(b, T(b), t) = 1\}$
8: $\quad b^\star \leftarrow \arg\max_{(b,t) \in \texttt{actions}} \delta(b, c^\star)$
9: $\quad t^\star \leftarrow \arg\min_{(b,t) \in \texttt{actions}:\, \sigma(c,k,t) \geq \tau(c,k)} \kappa(b,t)$
10: $\quad \hat{k} \leftarrow \arg\max_{h=0}^{k^\star} \{h : \sum_{b \in \mathcal{B}, t \in \mathcal{T}} x(b,k,t) < N\}$
11: $\quad x(b^\star, \hat{k}, t^\star) \leftarrow 1$

return x

three goals stated above correspond to three phases in the algorithm. Within each phase, we proceed in a similar, greedy way: identify the problems and find the most urgent one to fix; identify the possible actions and find the most appropriate one; and schedule said action at the most appropriate time.

It is worth stressing that from our viewpoint the future demand, i.e., the τ-values, is but an input to our problem. In practical settings, such demand will not be known with precision, and this will call for appropriate action, such as considering a safety margin. Our approach, however, remains unchanged.

4.2 Individual phases

Alg. 1 summarizes the steps we take in the first phase, where our objective is making sure that the traffic demand is met at all times and for all subscriber clusters. It works unmodified with and without network sharing: if there is no sharing, each operator will run Alg. 1 independently, feeding it its own network and its own load. If operators are performing their updates in a shared fashion, then Alg. 1 will be run only once, on the joint network and the total load.

The first thing we do is, in Line 2, to assess the performance we obtain from currently-scheduled actions, hence obtain the σ-values representing the traffic that can be served for each subscriber cluster. With reference to Fig. 3, calling function `assess` corresponds to entering the "performance assessment" cloud.

In Line 3, we look for *struggling* subscriber clusters, i.e., (c,k) pairs for which Eq. (3) does not hold. If there is no such pair (Line 4), then we are done and can move to phase 2. Otherwise, we proceed to Line 6, where we identify the $(c^\star, k^\star)$ pair that needs our attention next. The selection of the (c,k) pair to prioritize depends on the metric we decide to consider, expressed as a generic function $U(c,k)$ in our algorithm. For example, $U(c,k)$ can represent the degree of outage created by the network's problems, in which case $U(c,k) = \sigma(c,k) - \tau(c,k)$. In our case, we tackle the issue happening first, i.e., the one with the lowest k. In our case then $U(c,k) = k$.

So far, we have decided to perform a network change to tackle the capacity shortage affecting subscriber cluster $c^\star$ at time period $k^\star$. The set of base stations that we could decide to upgrade is identified in Line 7, and corresponds to the set of (b,t) pairs of base stations b such that (i) b would cover $c^\star$ if its technology were set to t, and (ii) the change would be capacity-preserving. Recall that we are relying on Property 1 and Property 2 to design the first two steps of our solution concept, and those properties only hold for capacity-preserving changes.

Among the base stations we may change, we have to identify the most appropriate one $b^\star$; in Line 8, we simply select the one that provides the highest RSSI to $c^\star$. In Line 9, we select the type $t^\star$ to update base station $b^\star$ to. We select, among the types that would restore the capacity constraint Eq. (3), the one with minimum cost. This also implies the mild assumption that the rate with which the traffic demand increases is never so high that we cannot restore Eq. (3), planning one action per time slot. Both the forecast in [1] and the fact that in our performance evaluation very few base stations are updated more than once throughout the whole simulation time, are consistent with such an assumption.

Last, we need to schedule the actual upgrade. We want to do so as late as possible, but no later than period $k^\star$. Therefore, in Line 10, we select the latest period between 0 and $k^\star$, in which we can still do something, i.e., for which we have scheduled to change the type of no more than $N-1$ base stations. Identified such a period $\hat{k}$, we proceed with scheduling the upgrade in Line 11, by setting the appropriate x-value to 1, and move to the next iteration. In the choice of $\hat{k}$, we can clearly see the relationship between the change rate N and our ability to keep unused capacity (i.e., the gray area in Fig. 1) to a minimum. Setting $\hat{k} = k^\star$ would mean making the change when needed, hence deploying no unused capacity; being forced to have $\hat{k} < k^\star$ means adding some network capacity that will be unused until time $k^\star$. As we clearly see from Line 10, the likelihood that we have to do so increases as N gets smaller.

It is also possible that the set in Line 10 is empty, i.e., there is no time at which we can schedule our action. This means that the network demand is growing too fast, that the change rate N is insufficient, and that network outages are unavoidable. Notice however that, as per Line 6, actions are decided in such a way to correct the earlier problems first: this means that even when outages do happen, Alg. 1 ensures they happen as late as possible.

Alg. 2 ensures that the competition constraint Eq. (4) is met. It has the same structure as Alg. 1, with some differences worth highlighting. The problematic subscriber clusters, identified in Line 3, are the ones where the HHI exceeds the value $H_{\max}$. The base station type $t^\star$

Algorithm 2 Phase 2: enforcing competition constraints.

Require: $\mathcal{B}, \mathcal{C}, \mathcal{K}, \tau$

1: **while true do**
2: $\quad \sigma \leftarrow \texttt{assess}(x, \delta, \gamma, \tau)$
3: $\quad \texttt{problems} \leftarrow \{(c,k) \in \mathcal{C} \times \mathcal{K} : H(c,k) < H_{\max}\}$
4: $\quad$ **if** $|\texttt{problems}| \leq (1-\phi) \cdot |\mathcal{C}|$ **then**
5: $\qquad$ **break**
6: $\quad c^\star, k^\star \leftarrow \arg\min_{(c,k)\in\texttt{problems}} k$
7: $\quad \texttt{actions} \leftarrow \{(b,t) \in \mathcal{B} \times \mathcal{T} : \gamma(b, c^\star, t) = 1\}$
8: $\quad b^\star \leftarrow \arg\max_{(b,t)\in\texttt{actions}} \delta(b, c^\star)$
9: $\quad t^\star \leftarrow \arg\max_{(b,t)\in\texttt{actions}:\, \pi(b,T(b),t)=1} \sigma(b, c^\star, t)$
10: $\quad \hat{k} \leftarrow \arg\max_{h=0}^{k^\star}\{h : \sum_{b\in\mathcal{B},t\in\mathcal{T}} x(b,k,t) < N\}$
11: $\quad x(b^\star, \hat{k}, t^\star) \leftarrow 1$
return x

Algorithm 3 Phase 3: reducing costs.

Require: $\mathcal{B}, \mathcal{C}, \mathcal{K}, \tau$

1: **for all** $b \in \mathcal{B}$ **do**
2: $\quad$ **for all** $t \in \mathcal{T} \setminus \{T(b)\}$ **do**
3: $\qquad \texttt{save}(b,t) \leftarrow \max(0, \kappa(b, T(b)) - \kappa(b,t))$
4: **sort** `save` **DESC**
5: **for all** $(b,t) \in \texttt{save}$ **do**
6: $\quad \hat{k} \leftarrow \arg\min_{h=0}^{K}\{h : \sum_{b\in\mathcal{B},t\in\mathcal{T}} x(b,k,t) < N\}$
7: $\quad x(b, \hat{k}, t) \leftarrow 1$
8: $\quad \sigma \leftarrow \texttt{assess}(x, \delta, \gamma, \tau)$
9: $\quad$ **if** Eq. (3) or Eq. (4) do not hold **then**
10: $\qquad x(b, \hat{k}, t) \leftarrow 0$
return x

to switch to is selected as the one that offers the highest capacity, so as to meet the constraint Eq. (4) with the smallest number of changes. Finally, the termination condition (Line 4) is triggered if at least a fraction ϕ of subscriber clusters have a sufficiently low HHI.

After Alg. 1 and Alg. 2, we are left with a set of capacity-preserving changes to the network that ensure that goals 1 and 2 (Eq. (3) and Eq. (4) respectively) are met. We now seek to schedule further changes, with the objective of minimizing the cost as defined in Eq. (5), i.e., attaining goal 3. Notice that we are not relying on Property 1 and Property 2 anymore, and are thus free to schedule non-capacity-preserving changes if need be.

We proceed as shown in Alg. 3, and begin by assessing, for each base station b (Line 1) and new type t (Line 2), how much we could save by switching b's type from $T(b)$ to t (Line 3). Then we examine the potential changes, starting from the one yielding the most savings (Line 4), and simply try them out (Line 7), scheduling them at the earliest possible time period, as shown in Line 6. In Line 8 we assess the impact of the newly-scheduled change on the capacity, i.e., the σ-values: if either Eq. (3) or Eq. (4) does not hold, i.e., if the new change impairs goal 1 or goal 2, we revert it in Line 10; otherwise, the change is confirmed.

The overall effect of Alg. 3 is scheduling further network changes, in addition to the ones decided in Alg. 1 and Alg. 2, with the purpose of saving money, i.e., reducing the operational costs as defined in Eq. (5). These changes are guaranteed not to jeopardize goal 1 and goal 2, thanks to the explicit check in Line 9. It is worth noting that changes are scheduled, in Line 6, as *early* as possible – conversely, the equivalent lines in Alg. 1 and Alg. 2 seek to schedule changes as late as possible. Notice that the check in Line 9 also implies that all subscriber clusters must be covered by at least one base station, i.e., no cost-saving action can be taken if that implies shrinking network coverage. As we will see in Sec. 7, this implicit constraint has an impact on the amount of savings we can achieve.

5 Solution properties

In this section, we examine the success conditions, computational complexity and optimality of our algorithms. We prove the properties for Alg. 1, but the same holds for Alg. 2 and Alg. 3 as well, which have the same structure.

5.1 Computational complexity

Our algorithms have been designed with scalability in mind, and exhibit low complexity, linear in the number of base stations. More formally:

Property 3. *The worst-case,* combined *time complexity of all our algorithms is $O(|\mathcal{B}|log|\mathcal{B}|)$.*

Proof: See the Appendix. □

This result allows us to efficiently tackle large-scale, real-world topologies, as we see in Sec. 7. Also notice that Property 3 refers to the combined complexity of our solution concept, i.e., all the algorithms described in Sec. 4, and to the worst case; real-world cases such as the one we consider in Sec. 7 show a substantially lower complexity.

5.2 Optimality

In the following, we assess how close our algorithms perform with respect to the optimum. Our algorithms make two kinds of decisions: *scheduling*, i.e., deciding when to perform network changes, and *choosing* the changes to make. The optimality of these decisions is discussed separately.

We state and prove our properties with reference to Alg. 1, i.e., the first step in Fig. 4; however, since the following steps have the same structure, similar arguments hold.

5.2.1 Scheduling

The question we look at is the following: given the times $k^\star$ by which changes need to be applied, how good are we at picking the time $\hat{k}$ at which changes are actually performed?

We indicate with $\mathbf{x} = (x_k)$ the vector of changes we have to schedule, i.e., x_k is the number of base stations whose type has to be changed within time period k. We begin by proving the following lemma, stating a necessary condition under which it is possible to schedule a set of changes:

Lemma 1. *The following condition is necessary for a set of updates to be schedulable:*

$$\sum_{h=1}^{k} x_h \leq kN, \forall k \in \mathcal{K}. \tag{7}$$

Lemma 1 can be verified by inspection of Eq. (7), as the total number of changes made is upper bounded by the change rate multiplied by the time in which the changes must be made. Lemma 1 says that any change set not satisfying Eq. (7) is impossible to schedule. Notice that we have not proven that condition Eq. (7) is sufficient, i.e., that if a set of changes does satisfy it then it is possible to schedule it, nor we know how to actually perform the scheduling. Thankfully, we can prove that Alg. 1 does the job:

Property 4. *If a set of changes satisfies Eq. (7), then Alg. 1 is able to schedule it.*

Proof: See the Appendix. □

Property 4 says that Alg. 1 can schedule all sets of changes satisfying Eq. (7), and Lemma 1 says that all other sets of changes are impossible to schedule, no matter the algorithm. It follows that if it is possible to schedule a given set of changes, then Alg. 1 will do it. This is important, but tells us nothing about how good Alg. 1 is at minimizing the unused capacity, i.e., the gray area in Fig. 1. Specifically, if $k_b^\star$ is the time period at which the type of base station b needs to be changed and $\hat{k}_b$ is the time at which the change is performed, we would like to minimize the quantity:

$$\sum_{b\in\mathcal{B}} \left(k_b^\star - \hat{k}_b\right). \tag{8}$$

Again, Alg. 1 happens to be as effective as it gets:

Property 5. *The schedule returned by Alg. 1 minimizes the quantity in Eq. (8).*

Proof: See the Appendix. □

5.2.2 Choice

After proving Property 4 and Property 5, we may be tempted to conclude that our approach is altogether optimal, i.e., shrinks the gray area in Fig. 1 to the absolute minimum. Regrettably, this is not the case: while the scheduling, i.e., deciding *when* to make changes to the network, is optimal, we cannot make the same claim about the *choice* of the changes to make, e.g., the base stations whose type is to be changed.

Indeed, optimally choosing the base stations to change is an NP-hard problem. (We skip the proof, which is based on reduction from the set-covering problem.) Greedy heuristics such as the one employed in Alg. 1 are widely adopted when dealing with NP-hard problems; indeed, inapproximability results show [30] that no better solutions than the ones provided by greedy algorithms exist unless $P = NP$. In other words, the best possible polynomial-time approximation for our problem yields a solution that is no closer to the optimum (except for a constant factor) than the one of our algorithms.

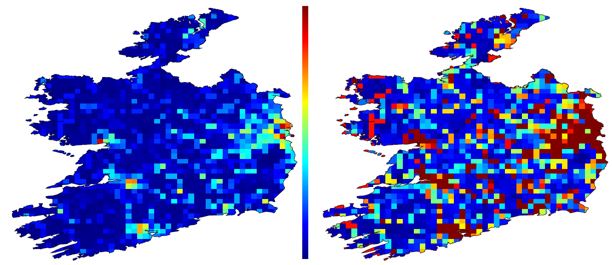


Fig. 5. Reference scenario. Blue areas correspond to low demand, red ones to high demand. The left plot refers to the present time, i.e., $k = 1$, the right plot to $k = K = 60$ months.

5.3 Summary

From our discussion, we can conclude that our algorithms exhibit a low level of complexity, and can schedule network changes in an optimal way, i.e., keeping the gap between the time when a change is needed and when it is applied to the minimum.

The choice of such changes is, in general, not optimal. On the other hand, greedy approaches similar to the one we adopt are commonly used in the literature [30], and have been shown to perform remarkably well in practice.

6 Reference Scenario

We study the performance of our algorithms and the factors affecting it in a large-scale, real-world scenario. In this section we describe our reference topology and traffic demand, as well as the simulator we employ.

Topology We leverage two demand and deployment traces, provided by two Irish operators. They consist of two weeks of anonymized call-detail records (CDR) information for both voice and data traffic, collected over the whole Republic of Ireland. They include position, (approximate) coverage, and sectorization information for over 6,000 base stations, which constitute our set $\mathcal{B}$. For each base station b, the corresponding type $\mathcal{T}(b)$ (e.g., GSM, 3G, LTE) is also given.

Traffic demand We generate the subscriber clusters using the demographic information publicly available from the Irish Central Statistics Office [31]. The demographic information is available in shapefile, dividing the surface of the Republic of Ireland into polygons, and a database file, containing for each polygon information such as the population, the area, general socio-economic information, and the classification of the area type (i.e., urban,

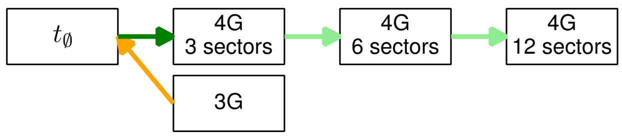


Fig. 6. The base station types in $\mathcal{T}$ and the possible changes. 3G base stations can be decommissioned (orange arrow), i.e., have their type set to $t_\emptyset$. New LTE base stations can be created (dark green arrow), possibly in the same location as existing base stations, or have their capacity enhanced through sectorization (light green arrows).

suburban, rural)[1]. We populate the set $\mathcal{C}$ by uniformly randomly distributing within the boundary of each polygon the minimum necessary number of points in such a way that each of them accounts for (i) at most 500 people and (ii) at most $5\ km^2$. By doing this we place around 31,000 subscriber clusters in our topology. We experimented with different population and area limits, obtaining essentially the same results we present later. Traffic demand $\tau(c, o, 1)$ at the present time slot $k = 1$ is obtained by preprocessing the traces and aggregating the traffic demand over time for each one-hour period. Since the nature of our problem is network planning, we retain for each base station the demand of its own busiest hour, even if such hours are not the same for all base stations. Then, by combining the preprocessed traces with the demographic data described earlier, we split the demand for each base station among all the subscriber clusters it covers according to the population each subscriber cluster covered represents. The macroscopic distribution of the traffic demand is summarized in the left-hand side plot of Fig. 5. Future demand is projected according to the Cisco forecast [1]. Our time horizon is $|\mathcal{K}| = 60$ time periods, with each period representing one month. The total demand for $k = K = 60$, represented on the right-hand side of Fig. 5, is six times the initial one.

Simulation and updates The sheer scale of our reference topology rules out network simulators such as ns-2 and OMNeT++; rather, we resort to a custom simulator written in Python. We estimate the received power strength (RSSI) at each subscriber cluster using the COST-231 Hata model [33], using the area type information to reduce inaccuracies due to overestimation and underestimation of the coverage in urban, suburban, and rural areas. We then compute the SINR and throughput between any (base station, subscriber cluster) pair through the OFCOM-vetted methodology in [34], under the assumption of a reuse factor of 1. The minimum fraction ϕ of clusters that must enjoy the target competition level is set to 0.7, unless otherwise specified.

We assume that the set of base station types $\mathcal{T}$ contains the following elements:

- the *decommissioned* type $t_\emptyset$;
- a type for 3G base stations;
- three types for LTE base stations, with different sectorizations.

The changes we can apply to the network are summarized in Fig. 6: we can *decommission* a 3G base station, or *create* a new LTE base station (possibly in the same location of an existing one), or *enhance* the capacity of an existing LTE base station by increasing the sectorization thereof. In the following, we will collectively refer to the last two operations as *updates*. It is worth stressing that these limitations are not inherent to our model, which is able to account for any kind of network update and to interface with any simulator (see Fig. 3), but merely a way to simplify (and speed up) our performance evaluation. Also notice that our model is able to account for the deployment of new base stations, which corresponds to an update from $t_\emptyset$ to any other type. Similarly, additional radio access technologies such as “small cells” would correspond to extra values in $\mathcal{T}$, and new branches in Fig. 6.

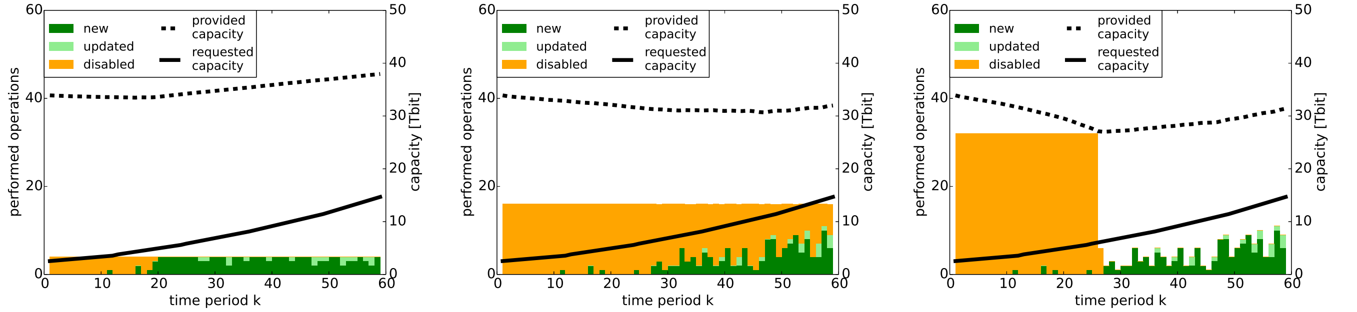
7 Results

We begin by looking at which network changes are performed at each time period k, and how they impact network capacity. In Fig. 7, solid and dotted lines correspond to requested traffic τ and provided capacity σ respectively; bars represent the number of created, enhanced and decommissioned base stations. We are setting in the most favorable case: networks can be operated jointly and there is no competition constraint, i.e., $H_{\max} = 1$.

Fig. 7(a) represents the case for $N = 4$, the minimum possible value of N in our scenario, as given by Eq. (7). It is easy to see that the value of N directly maps to the maximum height of the bars. Very low values of N, as in Fig. 7(a), imply that most of the changes operators are able to perform are updates (i.e., create or enhance base stations), so as to meet the demand goal Eq. (3). As N increases, as in Fig. 7(b), we are able to decommission more base stations, and to push forward in time all the updates. For even larger values of N, we see that there are some time periods when we perform fewer operations than we could, i.e., $\sum_{b\in\mathcal{B},t\in\mathcal{T}} x(b,k,t) < N$, as it happens for $k > 25$ in Fig. 7(c). This is because we scheduled to decommission all possible base stations at earlier times, as mandated by Line 6 in Alg. 3, and schedule all needed updates later in time, as in Line 10 of Alg. 1.

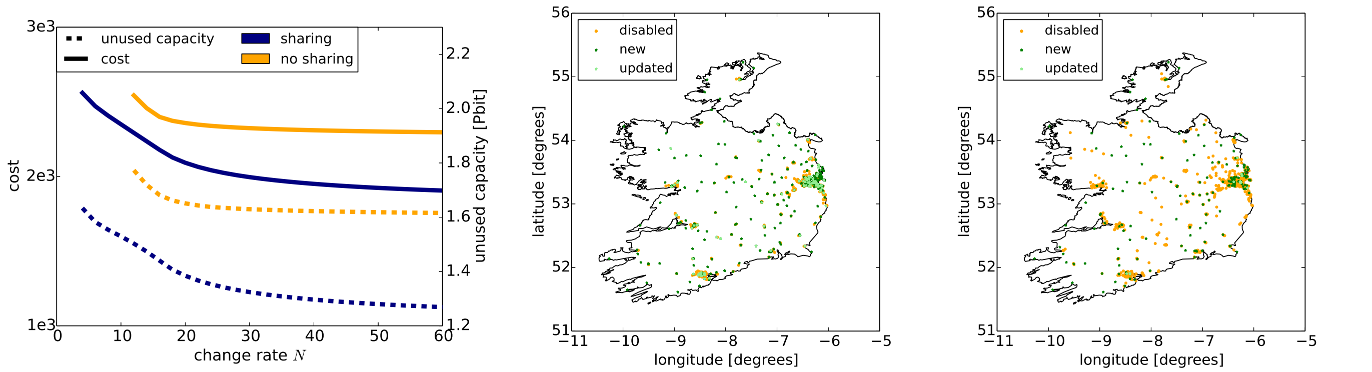
Looking at provided and requested capacity, we can notice that the provided capacity is always substantially higher than the demand. This is because we have to preserve the coverage in the entire topology, i.e., all subscriber clusters $c \in \mathcal{C}$. In sparsely populated areas, this inevitably translates into underutilized base stations that operators have no way to decommission. It is also interesting to see how N influences the evolution of provided

1. In some cases this information is explicit. In some other cases it can be inferred by looking at the density of the population per km^2 as it is done in [32].



(a) (b) (c)

Fig. 7. Changes applied to the network and requested and provided capacity for each time period k, when (a) $N = 4\ (= N_{\min})$, (b) $N = 16$, (c) $N = 32$.



(a) (b) (c)

Fig. 8. Unused capacity and cost savings (as defined in Eq. (5)) as a function of N with and without sharing (a); location of updated and decommissioned base stations without (b) and with (c) sharing.

capacity: low values of N imply that the capacity slowly increases as updates are performed (Fig. 7(a)). Higher values of N, as in Fig. 7(b), mean that we can observe the behavior we were expecting in Fig. 1, with network capacity first slowly decreasing due to decommissioning base stations and then leveling up due to the concurrent scheduling of updates and decommissions. As we can see from Fig. 7(c), further increasing N implies that network capacity decreases more swiftly (as operators can be quicker at decommissioning base stations) and increases more quickly afterwards, as most updates take place. Both effects are consistent with our intuition and expectations (Fig. 1): being able to perform more changes to the network means being more effective in tracking the traffic demand.

In Fig. 8 we look at the benefits of sharing, i.e., what savings operators can obtain by operating and updating their networks in a shared fashion. Fig. 8(a) is fairly clear – the benefits of sharing are very significant. The main effect of allowing sharing is that operators can save substantially more on operational costs, i.e., decommissioning more base stations. Sharing also reduces the unused capacity, which however remains quite significant, due to coverage requirements, as we already observed from Fig. 7.

The maps in Fig. 8(b) and Fig. 8(c) show *where* base stations are updated and decommissioned. Focusing on Fig. 8(b), which refers to the case where no sharing is allowed, we can observe that most updates are concentrated in densely populated areas (e.g., Dublin in the East), but some take place also in rural areas. On the other hand, virtually all decommissioned base stations are located in rural and suburban areas. Allowing sharing and moving to Fig. 8(c), we see a very different

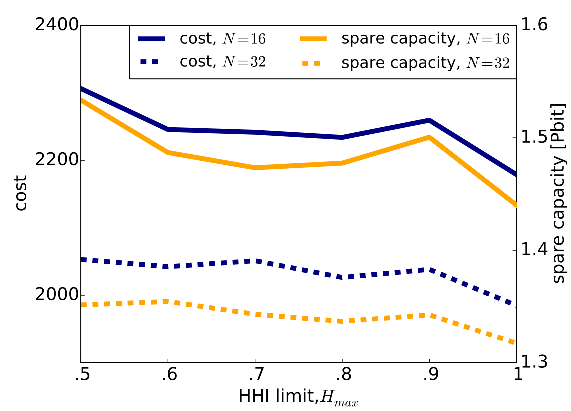
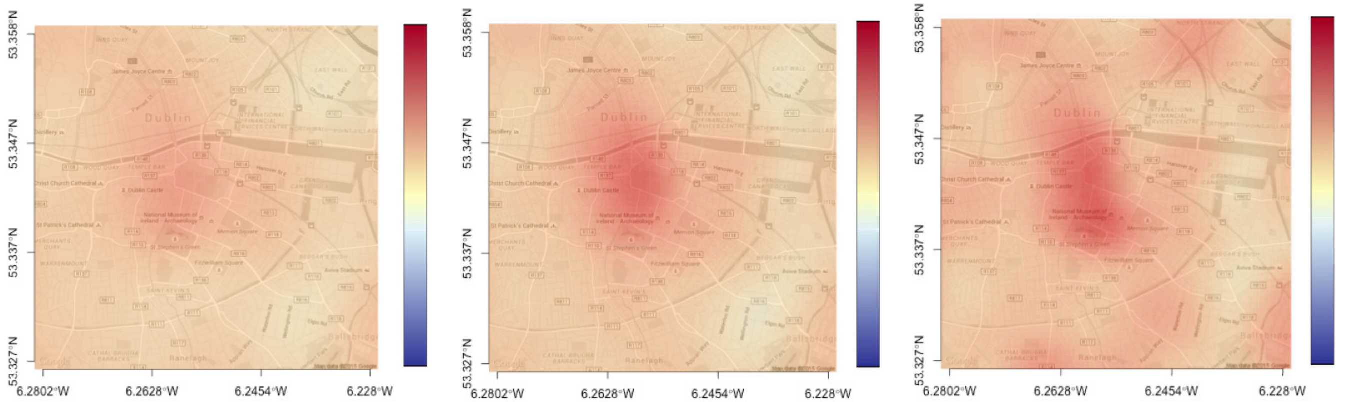
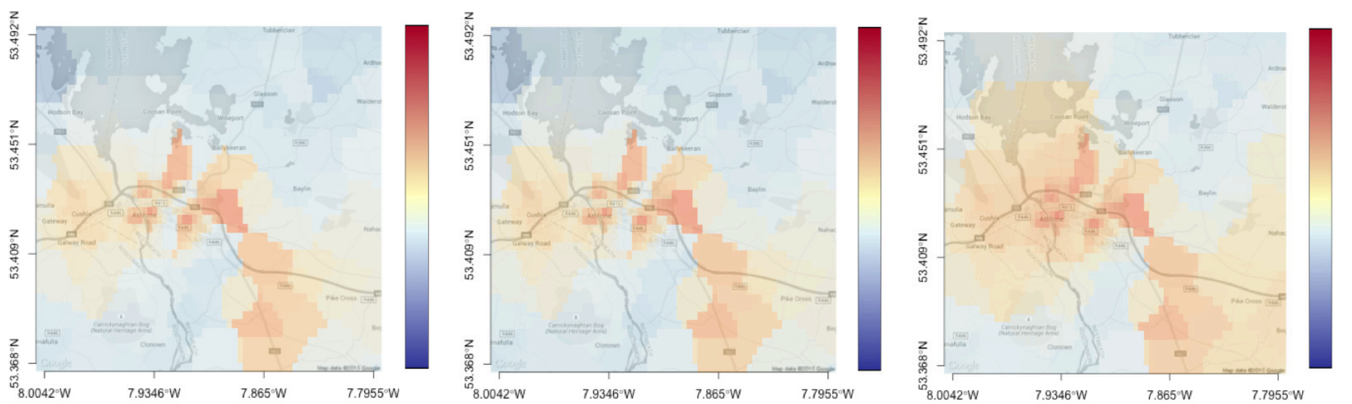


Fig. 9. Unused capacity and cost (as defined in Eq. (5)) as a function of HHI index and for different values of N.



(a) (b) (c)

Fig. 10. Dense urban case, Dublin - initial LTE capacity (a) final LTE capacity when $H_{\max} = 1$ (b), and $H_{\max} = 0.5$ (c). Values are presented normalized.



(a) (b) (c)

Fig. 11. Rural case, Co. Westmeath - initial LTE capacity (a) final LTE capacity when $H_{\max} = 1$ (b), and $H_{\max} = 0.5$ (c). Values are presented normalized.

picture. In rural areas operators have to update much fewer base stations and can decommission many more; furthermore, many base stations can be decommissioned also in urban areas, e.g., in Dublin.

Put together, these results confirm the intuitive notion that network sharing directly translates into better network efficiency. Backed by our real-world demand and deployment traces, we can add that said efficiency is mostly attained by decommissioning underutilized base stations and, to a lesser extent, by pooling updated ones. Location-wise, we can say that network updates have the overall effect of *migrating* capacity from rural areas to urban ones, and sharing makes such an effect more pronounced, taking advantage of the redundancies of deployments, in particular in the rural areas.

Competition regulation brings its own requirements on network planning, in particular mandating a certain level of extra capacity which could be used by a potential new virtual provider. In Fig. 9, we investigate the effects of competition regulation on the evolution of the network infrastructure. Recall that $H_{\max} = 1$ means that there is no regulation in place, while $H_{\max} = 0.5$ corresponds to the most stringent regulation, imposing as much as 50% idle capacity.

Fig. 9 confirms that the tighter the competition regulation, the lower the savings operators are able to achieve. As expected, moving from the most loose to the tightest level of regulation also increases the unused capacity – i.e., from the regulator's viewpoint, the capacity available to new operators. In fact, imposing a minimum competition level has a double effect: first, it forces the operators to perform some updates that otherwise would have not been necessary to meet the capacity goal in Eq. (3); second, it restrains the operators from decommissioning some underutilized base stations. Intuitively, since most of the decommissioning would take place in the rural areas and most of the updates in cities (Fig. 8(c)), the regulation has the effect that users from both sparsely and densely populated areas are able to choose between more (actual or potential) operators.

While Fig. 9 gives a macroscopic view of the overall capacity available, in Fig. 10 and Fig. 11 we compare how the competition affects the LTE capacity supplied in two different localized areas, i.e., a densely populated urban area in Dublin city center, and a sparsely populated rural area respectively. Fig. 10 and Fig. 11 confirm our previous findings. Without regulatory constraints, the operators still make investments to upgrade their infrastructures in densely populated areas, likely the ones with higher returns of investment (ROI), see Fig. 10(a) and Fig. 10(b), while in rural area they do not have incentives to make additional investments, see Fig. 11(a) and Fig. 11(b). On the other hand, regulatory constraints have the effect of stimulating more upgrades in both areas, see Fig. 10(c) and Fig. 11(c).

Fig. 12 gives us further insights on the effect of competition regulation on LTE and 3G coverage. Specifically, it shows the changes in the coverage by presenting the complementary CDF of the subscriber clusters' RSSI in the original LTE and 3G deployment (i.e., $k = 1$), and at the end of the time horizon (i.e., $k = K = 60$) when there is low and high competition in place, e.g. $H_{\max} = 1.0$ and $H_{\max} = 0.5$ respectively. Fig. 12 shows that higher competition stimulates the deployment of newest technology in areas that otherwise the operators would not find attractive, increasing both the number of subscriber clusters with a minimum signal strength and augmenting the overall signal quality. On the other hand, since the regulator enforces a minimum level of coverage for each technology, we do not observe a decrease in the number of subscriber clusters served by the older technology, but rather a small decrease in the signal quality, mainly due to the high redundancies existing in the infrastructure at $k = 1$.

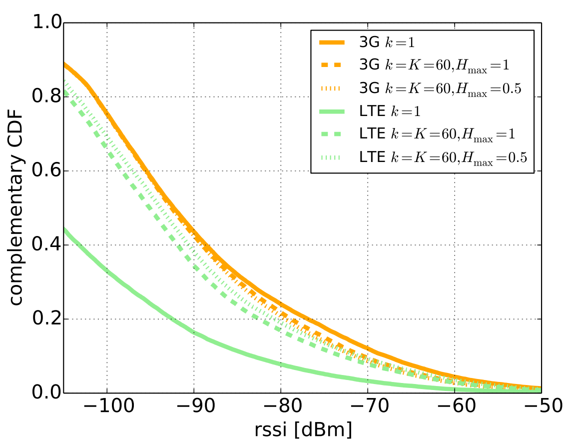


Fig. 12. Complementary CDF of the RSSI for LTE and 3G at time $k = 0$, and at $k = K = 60$ with low competition and high competition.

8 Conclusion

We studied the modernization phase cellular networks will go through in the near future: mobile operators will decommission some underutilized base stations in order to save on costs, and deploy new-generation base stations in order to cope with the increasing demand. Operators can join forces and perform said changes in a shared way, so as to improve the efficiency of their networks. Operators' ability to share infrastructure may be constrained by competition regulation, and the speed with which operators can make changes to their own networks may also be limited by practical considerations. We incorporate both factors in out study of cellular network planning.

Clearly, an operator's decisions about upgrading and decommissioning infrastructure will take into account a complex mix of technical and economic factors, from the OPEX associated with different radio access technologies to possible market advantages over one's competitors. Our model aims to capture the technical constraints on coverage and capacity that can be viewed as a first step in this decision making. As such, we use the number of base stations currently deployed as a rough proxy for the cost faced by the operator. A more sophisticated economic model that also takes into account new candidate locations for deploying new base stations is one possible avenue for the continuation of this work.

Our first contribution is a general framework that describes network modernization scenarios, accounting for real topologies and demand information, and including multiple base station technologies. This model was presented in Sec. 3. Given our model and a limited budget of changes to perform at each time period, we presented in Sec. 4 a family of algorithms able to schedule the changes in a cost effective manner, while satisfying the demand and complying with regulatory constraints. These algorithms work unmodified whether operators perform their updates individually or in a shared fashion and, as shown in Sec. 5, return quasi-optimal solutions with a very small computational complexity, dominated by their sorting stage.

We apply our algorithms in a large-scale scenario, built from real-world demand and deployment traces as described in Sec. 6. As summarized in Sec. 7, we found that network modernization essentially means moving capacity from rural, sparsely populated areas (where many base stations can be decommissioned) to urban ones (where most of the new-generation base stations are located). Allowing sharing, i.e., permitting operators to jointly update and manage their networks, greatly enhances their effectiveness. Such benefits are reduced if tight competition rules are in place, but never entirely jeopardized. Indeed, tight competition regulations have the secondary effect of stimulating operators to extend the capacity and coverage of their new-generation networks, which can therefore serve a larger fraction of their demand. As a result, our model is able to capture the tradeoff between *savings* and *the promotion of innovations* which is one of the main goals of regulators in the telecommunication industry.

Appendix: Proofs

Property 1. *Both goal 1 and goal 2 can be reached through network capacity-preserving changes alone, i.e., changes that comply with Eq. (6).*

Proof: Goal 1 means to satisfy Eq. (3) for all subscriber clusters $c \in \mathcal{C}$ and time periods $k \in \mathcal{K}$. If this is not the case, then the solution is scheduling updates that increase capacity. Decreasing the capacity for some subscriber clusters, i.e., breaking Eq. (6), is never necessary. A similar but slightly different reasoning holds for goal 2. Increasing the σ-values, as we can see from Eq. (2), decreases the HHI. □

Property 2. *If the initial configuration satisfies goal 1, then pursuing goal 2 by scheduling further capacity-preserving changes does not compromise goal 1.*

Proof: Once again, let us look at Eq. (3): if it holds, then all σ-values are no lower than the corresponding τ-ones, therefore, there is no way that further increasing the σ-values can change this. □

Property 3. *The worst-case,* combined *time complexity of all our algorithms is $O(|\mathcal{B}|log|\mathcal{B}|)$.*

Proof: At each iteration of each of our algorithms, we make exactly one decision, i.e., set one x-value to 1. Even if each base station is updated once to each possible type, the total number of decisions is still bounded by $|\mathcal{T}||\mathcal{B}|$, under the reasonable assumption that $|\mathcal{B}| \gg |\mathcal{T}|$, dominated by the sorting in Line 4 in Alg. 3, which has complexity $O(|\mathcal{B}| \log |\mathcal{B}|)$. □

Property 4. *If a set of changes satisfies Eq. (7), then Alg. 1 is able to schedule it.*

Proof: Scheduling a set of changes means enacting each of them at time $\hat{k} \leq k^\star$ no later than its deadline. In other words, every time we reach Line 10 in Alg. 1, the set $\{h \in \mathcal{K} : \sum_{b\in\mathcal{B},t\in\mathcal{T}} x(b,h,t) < N \wedge h \leq k^\star\}$ must be non-empty. In Line 6, we always select to schedule the base station with the lowest value of $k^\star$. This means that if at the current iteration we are scheduling a change due at time period $k^\star$, then all the changes we scheduled so far were due at $k^\star$ or earlier.

Since Eq. (7) holds, the number of such changes it at most $Nk^\star - 1$; therefore, there must be a $\hat{k}$ between 1 and $k^\star$ for which fewer than N changes have been scheduled. Hence, the set is non-empty. □

Property 5. *The schedule returned by Alg. 1 minimizes the quantity in Eq. (8).*

Proof: We prove the property by induction.
Initialization. At the first iteration of Alg. 1, all x-values are set to 0, thus in Line 10 we have $\hat{k} = k^\star$. The value of the quantity in Eq. (8) is zero, hence the schedule is optimal.
Induction step. Suppose all other changes have been scheduled optimally, i.e., they cannot be moved forward in time. Alg. 1 will try (Line 10) to schedule the current change for period $k^\star$, then $k^\star - 1$, and so on, stopping at the latest feasible time. It follows that the resulting schedule still has the lowest possible value of Eq. (8). □

REFERENCES

[1] Cisco, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2013-2018." [Online]. Available: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white_paper_c11-520862.pdf

[2] N. Vallina-Rodriguez, V. Erramilli, Y. Grunenberger, L. Gyarmati, N. Laoutaris, R. Stanojevic, and K. Papagiannaki, "When David Helps Goliath: The Case for 3G Onloading," in *Proceedings of the 11th ACM Workshop on Hot Topics in Networks*, ser. HotNets-XI. New York, NY, USA: ACM, 2012, pp. 85–90. [Online]. Available: http://doi.acm.org/10.1145/2390231.2390246

[3] P. Di Francesco, F. Malandrino, and L. A. DaSilva, "Mobile Network Sharing Between Operators: A Demand Trace-Driven Study," in *ACM SIGCOMM Capacity Sharing Workshop (CSWS)*, Chicago, Aug. 2014.

[4] B. Leng, P. Mansourifard, and B. Krishnamachari, "Microeconomic Analysis of Base-station Sharing in Green Cellular Networks," in *IEEE International Conference on Computer Communications (INFOCOM)*, Toronto, Apr. 2014.

[5] L. Doyle, J. Kibiłda, T. Forde, and L. A. DaSilva, "Spectrum without Bounds, Networks without Borders," *Proceedings of the IEEE*, vol. 102, no. 3, pp. 351–365, 2014.

[6] RTE News, "Vodafone criticises Three-O2 merger terms," 2014. [Online]. Available: http://www.rte.ie/news/business/2014/0529/620446-vodafone-three-criticism/

[7] E. Amaldi, A. Capone, and F. Malucelli, "Planning UMTS Base Station Location: Optimization Models with Power Control and Algorithms," *IEEE Transactions on Wireless Communications*, vol. 2, no. 5, pp. 939–952, 2003.

[8] C. Lee and H. Kang, "Cell Planning with Capacity Expansion in Mobile Communications: a TABU Search Approach," *IEEE Transactions on Vehicular Technology*, vol. 49, no. 5, 2000.

[9] A. Abdel-Khalek, L. Al-Janj, and Z. Dawy, "Optimization Models and Algorithms for Joint Uplink/Downlink UMTS Radio Network Planning with SIR-Based Power Control," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 4, pp. 1612–1625, 2011.

[10] L. Shangyun and M. St-Hilaire, "A Genetic Algorithm for the Global Planning Problem of UMTS Netowrks," in *IEEE Global Communications Conference (GLOBECOM)*, Miami, Dec. 2010.

[11] F. Gordejuela-Sanchez and J. Zhang, "LTE Access Network Planning and Optimization: A Service Oriented and Technology-Specific Prospective," in *IEEE Global Communications Conference (GLOBECOM)*, Honolulu, Dec. 2009.

[12] A. Guo and M. Haenggi, "Spatial Stochastic Models and Metrics for the Structure of Base Stations in Cellular Networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 11, pp. 5800–5812, 2013.

[13] H. Ghazzai, E. Yaacoub, M. S. Alouini, Z. Dawy, and A. Abu-Dayya, "Optimized LTE Cell Planning with Varying Spatial and Temporal User Densities," *IEEE Transactions on Vehicular Technology*, vol. PP, no. 99, 2015.

[14] J. S. Panchal, R. D. Yates, and M. M. Buddhikot, "Mobile Network Resource Sharing Options: Performance Comparisons," *IEEE Transactions on Wireless Communications*, vol. 12, no. 9, pp. 4470–4482, 2013.

[15] S. Hua, P. Liu, and S. S. Panwar, "The Urge to Merge: When Cellular Service Providers Pool Capacity," in *IEEE International Conference on Communications (ICC)*, Ottawa, Jun. 2012.

[16] J. Kibiłda and L. A. DaSilva, "Efficient Coverage through Inter-operator Infrastructure Sharing in Mobile Networks," in *Wireless Days*, Valencia, Nov. 2013.

[17] M. Z. Shafiq, L. Ji, A. Liu, J. J. Pang, and J. Wang, "Characterizing Geospatial Dynamics of Application Usage in a 3G Cellular Data Network," in *IEEE International Conference on Computer Communications (INFOCOM)*, Orlando, Apr. 2012.

[18] R. Keralapura, A. Nucci, Z. L. Zhang, and L. Gao, "Profiling users in a 3G network using hourglass co-clustering," in *ACM International Conference on Mobile Computing and Networking (MobiCom)*, Las Vegas, Sep. 2011.

[19] U. Paul, A. Subramanian, M. Buddhikot, and S. Das, "Understanding Traffic Dynamics in Cellular Data Networks," in *IEEE International Conference on Computer Communications (INFOCOM)*, Shanghai, Apr. 2011.

[20] C. Peng, S.-B. Lee, S. Lu, H. Luo, and H. Li, "Traffic-driven power saving in operational 3G cellular networks," in *ACM International*

Conference on Mobile Computing and Networking (MobiCom), Las Vegas, Sep. 2011.
[21] D. Willkomm, S. Machiraju, J. Bolot, and A. Wolisz, "Primary Users in Cellular Networks: A Large-Scale Measurement Study," in *IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN)*, Chicago, Oct. 2008.
[22] U. Paul, A. Subramanian, M. Buddhikot, and S. Das, "Understanding Spatial Relationships in Resource Usage in Cellular Data Networks," in *IEEE International Workshop on Network Science for Communication Networks (NetSciCom)*, Orlando, Mar. 2012.
[23] C. Courcoubetis and R. Weber, "Economic Issue in Shared Infrastructures," *IEEE/ACM Transactions on Networking*, vol. 20, no. 2, pp. 594–608, 2012.
[24] J. Markendahl and B. G. Molleryd, "On co-opetition between mobile network operators: Why and how competitors cooperate," International Telecommunications Society (ITS), 19th ITS Biennial Conference, Bangkok 2012: Moving Forward with Future Technologies - Opening a Platform for All 72491, 2012.
[25] M. Choudhary, H. Babar, H. Shakeel, and A. Abbas, "Economics of network sharing - a case study of mobile telecom sector in pakistan," in *Collaborative Computing: Networking, Applications and Worksharing, 2009. CollaborateCom 2009. 5th International Conference on*, Washigton DC, Nov. 2009.
[26] M. A. Marsan and M. Meo, "Energy efficient management of two cellular access networks," *SIGMETRICS Perform. Eval. Rev.*, vol. 37, no. 4, pp. 69–73, Mar. 2010. [Online]. Available: http://doi.acm.org/10.1145/1773394.1773406
[27] E. Oh, B. Krishnamachari, X. Liu, and Z. Niu, "Toward dynamic energy-efficient operation of cellular network infrastructure," *Communications Magazine, IEEE*, vol. 49, no. 6, pp. 56–61, June 2011.
[28] U.S. Department of Justice and the Federal Trade Commission, "Horizontal merger guidelines," Tech. Rep., Aug. 2010.
[29] European Commission, "EU Competition Law Rules Applicable to Merger Control Situation as at 1 April 2010," Tech. Rep., 2010.
[30] N. Alon, D. Moshkovitz, and S. Safra, "Algorithmic construction of sets for k-restrictions," *ACM Trans. Algorithms*, vol. 2, no. 2, pp. 153–177, Apr. 2006. [Online]. Available: http://doi.acm.org/10.1145/1150334.1150336
[31] [Online]. Available: http://www.cso.ie/en/census/census2011boundaryfiles/
[32] F. Pozzi and C. Small, "Analysis of urban land cover and population density in the United States," *Photogrammetric Engineering & Remote Sensing*, vol. 71, no. 6, 2005.
[33] M. Hata, "Empirical formula for propagation loss in land mobile radio services," *IEEE Transactions on Vehicular Technology*, 1980.
[34] "LTE Technical Modelling Revised Methodology," *OFCOM White Paper*, 2012. [Online]. Available: http://stakeholders.ofcom.org.uk/binaries/consultations/award-800mhz/annexes/annex14.pdf

Paolo Di Francesco Paolo Di Francesco received the B.S. and M.S. degree in telecommunications engineering from University of Bologna in 2008 and 2011 respectively. He is currently pursuing a PhD. degree in CTVR/CONNECT at Trinity College Dublin. His research interests include resource sharing, network optimization, data analysis, Software Defined Radios.

Francesco Malandrino Francesco Malandrino is currently a Fibonacci Fellow at the Hebrew University of Jerusalem. In 2014, he worked at CTVR, Dublin, as a research fellow. Before that, he was a post-doc at Politecnico di Torino, Italy, where he earned his Ph.D. in 2012. His interests focus on wireless and cellular networks, network planning, and data analysis.

Timothy K. Forde Timothy K. Forde received the B.A.,B.A.I. degree in Electronic and Commuter Engineering and a Ph.D. in Electronic Engineering from the University of Dublin, Trinity College, in 1999 and 2005 respectively. He was a Research Fellow at CTVR, the Irish national telecommunications research centre until 2014. Since January 2015 he is Direct of Innovation at CONNECT, the SFI Research Centre for Future Networks and Communications.

Luiz A. DaSilva Luiz A. DaSilva (SM) holda the chair of Telecommunications at Trinity College Dublin. He also has a research professor appointment in the Bradley Department of Electrical and Computer Engineering at Virginia Tech, USA. His research focuses on distributed and adaptive resource management in wireless networks, and in particular wireless resource sharing, dynamic spectrum access, and the application of game theory to wireless networks. He is currently a Principal Investigator on research projects funded by the National Science Foundation in the United States, the Science Foundation Ireland, and the European Commission under Horizon 2020 and Framework Programme 7. He is a Co-principal Investigator of CONNECT, the Telecommunications Research Centre in Ireland. Prof DaSilva is an IEEE Communications Society Distinguished Lecturer.