Network Slicing Games: Enabling Customization in Multi-Tenant Mobile Networks

Pablo Caballero, Albert Banchs, *Senior Member, IEEE,* Gustavo de Veciana, *Fellow, IEEE,* and Xavier Costa-Pérez, *Senior Member, IEEE*

***Abstract*—Network slicing to enable resource sharing among multiple tenants –network operators and/or services– is considered a key functionality for next generation mobile networks. This paper provides an analysis of a well-known model for resource sharing, the 'share-constrained proportional allocation' mechanism, to realize network slicing. This mechanism enables tenants to reap the performance benefits of sharing, while retaining the ability to customize their own users' allocation. This results in a *network slicing game* in which each tenant reacts to the user allocations of the other tenants so as to maximize its own utility. We show that, for elastic traffic, the game associated with such strategic behavior converges to a Nash equilibrium. At the Nash equilibrium, a tenant always achieves the same, or better, performance than under a static partitioning of resources, hence providing the same level of protection as such static partitioning. We further analyze the efficiency and fairness of the resulting allocations, providing tight bounds for the price of anarchy and envy-freeness. Our analysis and extensive simulation results confirm that the mechanism provides a comprehensive practical solution to realize network slicing. Our theoretical results also fill a gap in the literature regarding the analysis of this resource allocation model under strategic players.**

***Index Terms*—Wireless Networks, 5G, Network Slicing, Game theory, Resource allocation, Multi-tenant networks.**

I. Introduction

THERE is consensus among the relevant industry and standardization communities [1], [2] that a key element in 5G mobile networks will be *network slicing*. The idea is to allow the mobile infrastructure to be "sliced" into logical networks, which are operated by different entities and may be tailored to support specific services. This provides a basis for efficient infrastructure sharing among diverse entities, ranging from classical or virtual mobile network operators to new players that simply view connectivity as a service. Such new players could be, for instance, Over-The-Top (OTT) service providers which use a *network slice* to ensure satisfactory service to their customers (e.g., Amazon Kindle's support for downloading content or a pay TV channel including a premium subscription). In the literature, the term *tenant* is often used to refer to the owner of a network slice.

A network slice is a collection of resources and functions that are orchestrated to support a specific service. This includes software modules running at different locations as well as the nodes' computational resources, and communication resources in the backhaul and radio network. The intention is to only provide what is necessary for the service, avoiding unnecessary overheads and complexity. Thus, network slices enable tenants to compete with each other using the same physical infrastructure, but customizing their slices and network operation according to their market segment's characteristics and requirements. For instance, slices can be geared at supporting various IoT or M2M applications, such as the connectivity required to realize 'intelligent' vehicular systems.

A key problem underlying network slicing is enabling efficient sharing of mobile network resources. One of the frameworks considered in 3GPP suggests that resources could be statically partitioned based on fixed 'network shares' associated to each slice [3]. This framework fits very well some scenarios like, e.g., the case where several operators jointly contribute to a common infrastructure with a fraction of the overall cost and share this infrastructure with the others while being entitled to use an amount of resources that depends on their monetary contribution. For other network slicing scenarios, such as the case where some tenants only need to use a fixed amount of network resources for some limited period of time, other frameworks considered in the standards may be more appropriate.

The focus of this paper is on network slicing for the share-based framework mentioned above. However, given that slices' loads may be spatially inhomogenous and time varying, rather than statically partitioning the resources at each base station, it is deemed desirable to allow resource allocations to be dependent on the slices' loads at different base stations. At the same time, tenants should be protected from one another, and retain the ability to autonomously manage their slice's resources, in order to better customize allocations to their customers. To that end, it is desirable to adopt resource allocation models in which tenants can communicate their preferences to the infrastructure (say by dynamically subdividing their

Manuscript received December 11, 2017; revised November 30, 2018; accepted January 10, 2019; approved by IEEE/ACM TRANSACTIONS ON NETWORKING Editor M. Schapira. Date of publication xxxx xx, xxxx; date of current version xxxx xx, xxxx. The work of P. Caballero and G. de Veciana was supported in part by a gift from Cisco, and the work of A. Banchs was supported by the H2020 5G-MoNArch project (Grant Agreement No. 761445) and the 5GCity project of the Spanish Ministry of Economy and Competitiveness (TEC2016-76795-C6-3-R). *(Corresponding author: Albert Banchs.)*

P. Caballero and G. de Veciana are with the University of Texas at Austin, Austin, TX 78712 USA.

A. Banchs is with the University Carlos III of Madrid, 28911 Leganés, Spain, and also with the IMDEA Networks Institute, 28918 Leganés, Spain (email: banchs@it.uc3m.es).

X. Costa-Pérez is with NEC Laboratories Europe, 69115 Heidelberg, Germany.

Digital Object Identifier xx.xxxx/TNET.xxxx.xxxxxx

network share amongst their customers) and then have base stations' resources allocated according to their preferences (e.g., proportionally to the customers' shares).

Under such a dynamic resource allocation model, a tenant might exhibit strategic behavior, by adjusting its preferences depending on perceived congestion at resources, so as to maximize its own utility. Such behavior could in turn have adverse effects on the network; for instance, the overall efficiency may be harmed, or one may see instability in slice requests. The focus of this paper is on (i) the analysis and performance of this simple resource allocation model, and (ii) the validation of its feasibility as a means to enable tenants to customize resource allocation within their slice while protecting them from one another. The analysis of this paper concentrates on elastic traffic; the case of inelastic traffic has been addressed by the authors in [4].

Related work

The resource allocation mechanism informally described above, aligned with the fixed 'network shares' model considered in 3GPP, corresponds to a Fisher market. This is a standard framework in economics; in such markets, buyers (in our case slices) have fixed budgets (in our case network shares) and (according to their preferences) bid for resources within their budget, which are then allocated to buyers proportionally to their bids. Analysis of the Fisher market shows that, as long as buyers are price-taking (i.e., they do not anticipate the impact of their bids on the price – in our case, the impact of the slices' preferences on the overall congestion), the Nash equilibrium is socially optimal, and distributed algorithms can be easily devised to reach it [5]. This assumption may be reasonable for markets where the impact of a single buyer on a resource's price is negligible, but does not apply to our case where a relatively small number of active tenants might be sharing resources.

There is a substantial literature on Fisher markets with strategic buyers, which, as will be studied in this paper, anticipate the impact of their bids [6]. The analysis, so far, has been limited to the case of buyers with *linear utility* functions of the allocated resources, which can lead to extremely unfair allocations. While such utility functions may be suitable for goods, they are not an appropriate model for tenants wishing to customize allocations amongst their customers. This paper includes a comprehensive analysis for a wide set of slice utility functions, including the convergence of best response dynamics and other results which to our knowledge are new.

A related resource allocation model often considered in the networking field is the so-called 'Kelly's mechanism' [7]; this mechanism allocates resources to players proportionally to their bids and, assuming that they are price-taking, converges to a social optimum. Follow-up work has considered price-anticipating players in this setting; for example, [8] analyze efficiency losses, while [9] devise a scalar-parametrized modification that is once again socially optimal for price-anticipating players. However, in Kelly's mechanism players respond to their payoff (given by the utility minus cost) whereas in our model tenants' behavior is only driven by their utilities (since they have a fixed budget: the network share). Consequently, results on the analysis of Kelly's mechanism are not applicable to our setting

In the context of the existing resource allocation models described above, this paper addresses the following gap in the literature: the analysis of *budget-constrained resource allocation* under *price-anticipating users* with *nonlinear utilities*. This requires novel analysis that differs substantially from previous work in the literature. Table I summarizes some of the main resource allocation models for this problem, highlighting the most relevant contributions for each case and situating the contribution of this work.

	price taking	**price anticipating**	
	scalar bid	**scalar bid**	**vector bid**
non fixed budget	[7] Kelly's mechanism (conv, efficiency)	VCG-Kelly mechanism [9] Hajek/Yang [10] Johari/Tsitsiklis	Johari/Tsitsiklis [8] Efficiency of congestion games
	concave utilities	**linear utilities**	**concave utilities**
fixed budget	[5] Zhang (convergence)	[6] Feldman (conv, efficiency)	*This work* (conv, efficiency)

TABLE I: Resource allocation models.

All the analyses mentioned above, as well as that conducted in this paper, consider concave utility functions, which reflects the behavior of elastic traffic [11]. In contrast to elastic applications, inelastic applications typically require a minimum amount of resources to provide an acceptable experience to the users, and their performance degrades drastically if resources fall below this minimum. The case of inelastic traffic has been addressed by the authors in [4], leading to quite different outcomes; indeed, in contrast to the results obtained in this paper, for inelastic traffic the existence of a Nash Equilibrium is not guaranteed, best response dynamics may not converge and the Price of Anarchy is not bounded.

Beyond the Fisher market model, there have been a number of works in the literature that address game theory and α-fairness, as we do in this paper. The work in [12] analyzes the game resulting from allowing users to select the access network, when resources in each network are allocated based on α-fairness; in contrast, our game is played by selecting the bid submitted by each tenant, rather than the access network. In [13], the authors prove the convergence of games where individual user choices are driven by a convex optimization (as in our case); however, they require some properties not met by our game. The analysis of [14] shows the convergence of a game with some similarities to ours, building on potential game theory which cannot be applied in our case. In [15], a number of methodologies are proposed to analyze equilibria in wireless games, yet none of those games coincides with ours. To the best of our knowledge, even though there is a vast literature addressing equilibria and convergence of wireless games, none of the existing tools can be applied to the specific problem addressed here.

In order to design algorithms that converge to a Nash Equilibrium, some work in the literature has proposed leveraging reinforcement learning techniques [16]–[20]. A key advantages of such approaches is that they do not require all the knowledge involved in computing the best response. However, while this is an essential feature for the systems addressed by those papers, where best response requires substantial information, in our system the best response requires only limited information, and hence a practical approach can be built based on best responses without having to resort to reinforcement learning.

From a more practical perspective, multi-tenant sharing has been studied from different points of view, including planning, economics, coverage, performance, etc. [21], [22]. This paper focuses specifically on the design of algorithms for resource sharing among tenants, which has been previously addressed by [23]–[26]. The work of [26] considers sharing via a bid-based auction, which may incur substantial overhead and complexity; in contrast, our approach relies on fixed (pre-negotiated) network shares. The works of [23]–[25] also fix a network share per slice, but consider approaches where the infrastructure makes centralized decisions on the resources allocated to each tenant's customers; hence, these approaches do not enable tenants to make their own decisions on how to allocate resources to their customers.

Network slicing has emerged as a desirable feature for 5G [1]. 3GPP has started work on defining requirements for network slicing [2], whereas the Next Generation Mobile Network (NGMN) alliance has identified network sharing among slices (the focus of this paper) as a key issue [27]. In spite of these efforts, most of the work so far has addressed architectural aspects with only a limited focus on resource allocation algorithms [28], [29]. To the best of our knowledge, this is the first work investigating how to enable tenants to customize their allocations in a dynamic slicing model; there is wide consensus that such an ability to custumize tenants' allocations is needed to efficiently satisfy their very diverse requirements (see, e.g., [30] for examples of vertical tenants).

Key contributions

The rest of the paper is organized as follows. After introducing our system model (Section II), we show that with the resource sharing model under study, each slice has the ability to achieve the same or better utility than under static resource slicing irrespective of how the other slices behave, which confirms that this model effectively protects slices from one another (Section IV-A). Next we show that if tenants exhibit strategic behavior (i.e, optimize their utilities), then (i) a Nash equilibrium exists under mild conditions; and (ii) the system converges to such an equilibrium when tenants sequentially take their best response (Sections III-B and III-C). The resulting efficiency and fairness among tenants are then studied, providing: (i) a tight bound on the Price of Anarchy of the system, and (ii) a bound on the Envy-freeness (Section IV). Our results are validated via simulation, confirming that the approach provides substantial gains, protects network slices from each other, operates close to optimal performance and is effectively envy-free (Section V).

II. System model

We consider a wireless network consisting of a set of resources $\mathcal{B}$ (the base stations or sectors) shared by a set of network slices $\mathcal{O}$ (the tenants). At a given point in time, each slice supports a set of users (the customers or devices). The wireless network is operated by an infrastructure provider (hereafter the 'infrastructure'). Each tenant is the owner of a slice and requests resources for this slice to the infrastructure provider. The network resources allocated to the slice are then shared among the users of that slice (according to the preferences expressed by the tenant of the slice).

A. Resource allocation model

As indicated in the introduction, we focus on a well established resource sharing model known in economics as a Fisher market; we will refer to this model as the 'Share-Constrained Proportional Allocation' (SCPA) mechanism.

Hereafter, we refer to the set of users supported by the network as $\mathcal{U}$, which can be divided into subsets $\mathcal{U}_b$ (the users at base station b), $\mathcal{U}^o$ (the users of slice o) and $\mathcal{U}_b^o$ (their intersection). For any user $u \in \mathcal{U}$, we let $b(u)$ denote the base station it is currently associated with.

In our setting, each slice o is allocated a network share s_o (corresponding to its budget) such that $\sum_{o\in\mathcal{O}} s_o = 1$. The slice is at liberty in turn to distribute its share amongst its users, assigning them weights (corresponding to the bids): w_u for $u \in \mathcal{U}^o$, such that $\sum_{u\in\mathcal{U}_o} w_u = s_o$. We let $\mathbf{w}^o = (w_u \colon u \in \mathcal{U}_o)$ be the weights of slice o, $\mathbf{w} = (w_u \colon u \in \mathcal{U})$ those of all slices and $\mathbf{w}^{-o} = (w_u \colon u \in \mathcal{U} \setminus \mathcal{U}^o)$ the weights of all users excluding those of slice o.

In this paper, we adopt a generic formulation for resources that can be applied to a variety of technologies. The specific definition of resource will depend on the underlying technology; for instance, in LTE/LTE-A resources refer to physical Resource Blocks, in FDM to bandwidth and in TDM to the fraction of time. We shall assume users are allocated a fraction of resources at their base station proportionally to their weights w_u. Thus, the rate of user u is given by

$$r_u(\mathbf{w}) = \frac{w_u}{\sum_{v\in\mathcal{U}_{b(u)}} w_v} c_u = \frac{w_u}{l_{b(u)}(\mathbf{w})} c_u$$

where $l_b(\mathbf{w}) = \sum_{u\in\mathcal{U}_b} w_u$ denotes the overall load at b and c_u is the user's achievable rate, defined as the rate that the user would see if she had the entire base station to herself. Note that c_u depends on the modulation and coding scheme selected for the current radio conditions, which accounts for noise as well as the interference from the neighboring base stations. Following similar analyses in the literature (see, e.g., [23], [25]), we shall assume that c_u is fixed for each user at a given time.

To implement the above resource allocation, a slice needs to communicate the weights of its users $\mathbf{w}^o$ to the infrastructure. In turn, the infrastructure needs to communicate to the slice the overall load at each base station, so that the slice can select the weights of its users.[1] We argue that this is a relatively light

[1] Note that, even if the load was not communicated by the infrastructure, a slice could infer it by varying its users' weights and observing the resulting resource allocations.

exchange of information; as a matter of fact, there are already some interfaces defined in 3GPP, such as the X2 interface, which share this kind of information. Note that, by sharing information in this way, the weights of a given tenant are not disclosed to the others, but only the overall load at each base station.

In the case where a slice o is the only one with users at a given base station b, we shall assume that the slice's users are allocated the entire capacity at that base station independent of their weights. Thus such a slice would set $w_u = 0$ for these users, allowing them to receive all the resources of this base station without consuming any share. In order to avoid dealing with this special case, and without loss of generality, we will make the following assumption for the rest of the paper.

Assumption 1. *(Competition at all resources) We assume that all resources have active users from at least two slices.*

B. Network Slice Utility and Service Differentiation

Network slices may support services and customers of different types and needs. Alternatively, competing slices with similar customer types may wish to differentiate the service they provide. To that end, we assume each network slice has a *private* utility that reflects the benefit obtained by the slice from a given allocation and is given by

$$U^o(\mathbf{w}) = \sum_{u \in \mathcal{U}^o} \phi_u f_u(r_u(\mathbf{w})), \tag{1}$$

where ϕ_u is the relative priority of user u, with $\phi_u \geq 0$ and $\sum_{u\in\mathcal{U}_o} \phi_u = 1$, and $f_u(\cdot)$ is a utility function associated with the user. In the sequel, we will often focus on the following well-known class of utility functions [31]:

Definition 1. *A network slice o has a homogenous α_o-fair utility if for all $u \in \mathcal{U}^o$ we have that*

$$f_u(r_u) = \begin{cases} \frac{(r_u)^{1-\alpha_o}}{(1-\alpha_o)}, & \alpha_o \neq 1 \\ \log(r_u), & \alpha_o = 1. \end{cases}$$

With the above setting, a slice is free to choose different fairness criteria in allocating resources across its users, by selecting the appropriate α_o parameter. Note that $\alpha_o = 1$ corresponds to the widely accepted proportional fairness criterion, while $\alpha_o = 2$ corresponds to potential delay fairness, $\alpha_o \to \infty$ to max-min fairness and $\alpha_o = 0$ to linear sum utility.

Note that the chosen utility function $f_u(r_u)$ is concave, with a concavity level that can be adjusted with the α_o parameter. Such a utility is appropriate to represent the behavior of elastic traffic, the performance of which mainly depends on rate and exhibits diminishing utility improvements as throughput increases [11].

C. Baseline allocations

Next, we introduce some resource allocation comparative baselines.

a) Socially Optimal Allocations (SO): If slices were to share their utility functions with a centralized authority, one could in principle consider a socially optimal allocation of weights and resources. These would be given by the maximizer to the *overall network utility* $U(\mathbf{w})$ given by (see [25]):

$$\begin{aligned} \max_{\mathbf{w}\geq 0} \quad & U(\mathbf{w}) := \sum_{o\in\mathcal{O}} s_o U^o(\mathbf{w}) \\ \text{s.t.} \quad & r_u(\mathbf{w}) = \frac{w_u}{l_{b(u)}(\mathbf{w})} c_u, \quad \forall u \in \mathcal{U} \\ & \sum_{u\in\mathcal{U}^o} w_u = s_o, \quad \forall o \in \mathcal{O}. \end{aligned}$$

Note that (as in [25]) we have weighted the slices' utilities to reflect their shares (thus prioritizing those with higher shares). We shall denote the resulting optimal weight and resource allocations under the socially optimal allocations by $\mathbf{w}^*$ and $\mathbf{r}^* = (r_u^* : u \in \mathcal{U})$, respectively.

b) Static Slicing (SS): By static slicing (also known as static splitting [32]) we refer to a complete partitioning of resources based on the network shares $s_o, o \in \mathcal{O}$. In this setting, each slice o receives a fixed fraction s_o of each resource and can unilaterally optimize its weight allocation as follows:

$$\begin{aligned} \max_{\mathbf{w}^o\geq 0} \quad & U^o(\mathbf{w}^o) = \sum_{u\in\mathcal{U}^o} \phi_u f_u(r_u(\mathbf{w}^o)) \\ \text{s.t.} \quad & r_u(\mathbf{w}^o) = \frac{w_u}{\sum_{v\in\mathcal{U}^o_{b(u)}} w_v} s_o c_u \quad \forall u \in \mathcal{U}^o \\ & \sum_{u\in\mathcal{U}^o} w_u = s_o, \end{aligned}$$

where we have abused notation to indicate that, in this case, U^o and r_u depend only on $\mathbf{w}^o$. We shall denote the resulting optimal weight and resource allocations under static slicing for all slices by $\mathbf{w}^{ss}$ and $\mathbf{r}^{ss} = (r_u^{ss} : u \in \mathcal{U})$ respectively, where

$$r_u^{ss} = \frac{w_u^{ss}}{\sum_{v\in\mathcal{U}^o_{b(u)}} w_v^{ss}} s_o c_u \quad \forall u \in \mathcal{U}^o, \forall o \in \mathcal{O}. \tag{2}$$

c) Optimal Dynamic Pricing: An alternative resource allocation model to the one considered in Section II-A, where slices have a fixed budget, is to let slices bid for individual base station resources and let the infrastructure provider set a price to the slices. Under such a model, the payoff obtained by the slice is given by $\Pi^o = U^o - p_o$, where U^o is the utility obtained by slice o with the allocated resources, given by (1), and p_o is the total price set to the slice for such resources.

A particularly interesting strategy within dynamic pricing is the optimal dynamic pricing approach proposed in [8], which sets the price for slice o equal to the total utility loss caused to the other slices, i.e., $p_o = \sum_{o'\in\mathcal{O}\setminus o} U^{o',-o} - U^{o'}$, where $U^{o',-o}$ and $U^{o'}$ are the utility of slice o' with and without slice o participating in the game, respectively. This strategy drives the system to the social optimal (SO) allocation; thus, when comparing the performance of our system against the SO in Sections IV and V, the results also apply to the performance attained by an optimal dynamic pricing approach.

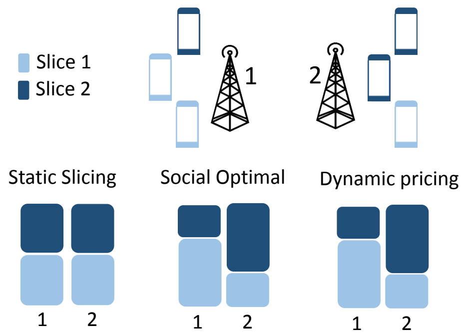

Fig. 1: Allocated resources for the three benchmark approaches.

While dynamic pricing may yield optimal system performance, it also suffers from significant drawbacks as compared to our model involving a fixed budget: (i) to set the optimal prices, the infrastructure provider needs to know the utility function of all slices, which involves a significant complexity and overhead; and (ii) as argued in [33], for practical purposes tenants typically prefer to deal with predictable pricing strategies and costs driven by market considerations, rather than by the instantaneous demands of the other tenants.

Figure 1 shows the allocation corresponding to each of the above baseline allocations for a network with two base stations and two slices with equal shares ($s_1 = s_2 = 0.5$). We observe that static slicing allocates one half of each base stations' resources to each slice, independent of the number of users of each slice, while the other two allocations take into account the number of users of the slices at each base station.

III. Network Slicing Game

Under the SCPA resource allocation model, it is reasonable to assume that a player (network slice) would 'strategically' optimize the weight allocation of its users to maximize its own utility (and thus the service delivered to its customers). In the following, we analyze the game resulting from such a strategic behavior.

A. Game formulation

Since the resources allocated to a user depend on the weight allocations of the other slices, the behavior of a slice will be predicated on the aggregate weight of the other slices at each resource. From the point of view of slice o, the overall load at resource b can be decomposed as

$$l_b(\mathbf{w}) = a_b^o(\mathbf{w}^{-o}) + d_b^o(\mathbf{w}^o)$$

where

$$a_b^o(\mathbf{w}^{-o}) = \sum_{o' \in \mathcal{O}\setminus\{o\}} \sum_{u \in \mathcal{U}_b^{o'}} w_u \quad \text{and} \quad d_b^o(\mathbf{w}^o) = \sum_{u \in \mathcal{U}_b^o} w_u,$$

correspond to the aggregate weight of the other slices and that of slice o, respectively. In our model, each slice is informed by the infrastructure of the overall load at each base station $\mathbf{l} = (l_b : b \in \mathcal{B})$; from this, the slice can obtain $\mathbf{a}^o = (a_b^o : b \in \mathcal{B})$, by subtracting d_b^o from the l_b values. Then, based on $\mathbf{a}^o$, it can choose the weight setting $\mathbf{w}^o$ that maximizes its utility. This leads to the following game.

Definition 2. *In the network slicing game, each slice o is aware of the aggregated weight of the other slices at each base station, $\mathbf{a}^o = (a_b^o : b \in \mathcal{B})$, and chooses the weight allocation $\mathbf{w}^o$ that maximizes its utility.*

B. Existence and Uniqueness of Nash Equilibrium

Next we study whether there exists a Nash equilibrium (NE) under which no slice can benefit by unilaterally changing its weight allocation. To that end, we first characterize the best response of a slice. Given the weights of the other slices, $\mathbf{w}^{-o}$, the best response of slice o is the unique maximizer $\mathbf{w}^o$ of its utility, i.e.,

$$\max_{\mathbf{w}'^o \geq 0} \sum_{u \in \mathcal{U}_o} \phi_u f_u \left(\frac{w'_u c_u}{a^o_{b(u)}(\mathbf{w}^{-o}) + d^o_{b(u)}(\mathbf{w}'^o)} \right)$$
$$\text{s.t} \sum_{u \in \mathcal{U}_o} w'_u = s_o.$$

The following lemma characterizes the best response for a network slice with homogenous α_o-fair utility (see [6] for the best response when $\alpha_o = 0$).

Lemma 1. *Suppose slice o has a homogeneous α_o-fair utility (with $\alpha_o > 0$). Given the weights of the other slices $\mathbf{w}^{-o} > 0$, slice o's best response $\mathbf{w}^o$ is the unique solution to the following nonlinear set of equations:*

$$w_u = \frac{\beta_u \frac{\left(a^o_{b(u)}(\mathbf{w}^{-o})\right)^{\frac{1}{\alpha_o}}}{\left(a^o_{b(u)}(\mathbf{w}^{-o}) + d^o_{b(u)}(\mathbf{w}^o)\right)^{\frac{2}{\alpha_o}-1}}}{\sum_{v \in \mathcal{U}_o} \beta_v \frac{\left(a^o_{b(v)}(\mathbf{w}^{-o})\right)^{\frac{1}{\alpha_o}}}{\left(a^o_{b(v)}(\mathbf{w}^{-o}) + d^o_{b(v)}(\mathbf{w}^o)\right)^{\frac{2}{\alpha_o}-1}}} s_o, \quad \forall u \in \mathcal{U}^o, \quad (3)$$

where $\beta_u := (\phi_u)^{\frac{1}{\alpha_o}} (c_u)^{\frac{1}{\alpha_o}-1}$.

Note that slice o need only know $\boldsymbol{a}^o(\mathbf{w}^{-o})$ to compute its best response. Building on this characterization, we will study the game in which all slices choose to allocate their weights based on their best response. The following theorem proves that this game admits a Nash equilibrium, i.e., there is a weight allocation $\mathbf{w}$ such that no slice can improve its utility by modifying its weights unilaterally.[2]

Theorem 1. *Suppose all slices have homogenous α_o-fair utilities (with possibly different $\alpha_o > 0$). Then, there exists a (not necessarily unique) Nash equilibrium satisfying* (3) *for each slice.*

The above theorem covers any finite α_o value, but leaves out the case $\alpha_o \to \infty$, which yields a utility function $U^o(\mathbf{w}) = \min_{u \in \mathcal{U}^o} (r_u(\mathbf{w}))$ and corresponds to max-min fairness. The following lemma shows that in this case the existence of a NE is not guaranteed.

Lemma 2. *Let $U^o(\mathbf{w}) = \min_{u \in \mathcal{U}^o} (r_u(\mathbf{w}))$ for two or more slices. Then, the existence of a NE cannot be guaranteed.*

[2]The existence of a NE had already been proven by [5] for the case $\alpha_o = 0 \ \forall o$. Here we extend this result to any combination of α_o values.

C. Convergence of Best Response Dynamics

Below, we will consider best response dynamics wherein slices realize their best responses in rounds, either (i) updating their weights $(\mathbf{w}^o)$ sequentially, one at a time and in the same fixed order, in response to the other slices' weights $(\boldsymbol{a}^o)$; or (ii) having all slides update their weights simultaneously in each round in response to the other slices' weights in the previous round.

Theorem 2. *If slices have homogeneous α_o-fair utilities, possibly with different $\alpha_o \in [1,2]$ for $o \in \mathcal{O}$, then the best response game converges to a Nash equilibrium. This result holds both for sequential and for simultaneous updates.*

Note that the value of α_o impacts a slice's best response and consequently the game dynamics. As seen in Lemma 1, the best response weights are proportional to:

$$w_u \propto g(a_b^o, d_b^o) := \frac{(a_b^o)^{\frac{1}{\alpha_o}}}{(a_b^o + d_b^o)^{\frac{2}{\alpha_o}-1}},$$

where we have suppressed the dependency of a_b^o on $\mathbf{w}^{-o}$ and d_b^o on $\mathbf{w}^o$. The function $g(\cdot,\cdot)$ has different properties depending on α_o which are shown in Table II. The regime where $1 \le \alpha_o \le 2$, considered in Theorem 2, is of particular interest since it includes proportional $(\alpha_o = 1)$ and potential delay $(\alpha_o = 2)$ fairness. It is known that convergence is not ensured when $\alpha_o = 0$ for all slices (see [6]); for other regimes, we resort to the simulations results of Section V, which suggest convergence for any $\alpha_o > 0$ since they are different problems in nature and therefore the analysis require a distinct approach.

	$\alpha_o = 0$	$0 < \alpha_o < 1$	$1 \le \alpha_o \le 2$	$2 < \alpha_o < \infty$
g w.r.t. d_b^o	linear	convex	convex	concave
g w.r.t. a_b^o	linear	convex	concave	concave
NE existence	✓ [6]	✓Theorem 1 for heterogeneous α_o		
convergence	× [6]	✓simulations	✓Theorem 2	✓simulations

TABLE II: Impact of α_o on slice's Best Responses.

Perhaps surprisingly, the above result is quite challenging to show. The key challenge lies in the "price-anticipating" aspect of the best response, in which players anticipate the impact of their own allocation (indeed, as mentioned in the introduction, there are very few results in the literature on the convergence of price-anticipating best response dynamics).

IV. Performance analysis

In this section we analyze the performance of the Nash equilibrium in terms of: (i) the *gain over static slicing*, which is the benchmark allocation where resources are statically partitioned, (ii) the *price of anarchy*, which gives the loss in overall utility resulting from slices' strategic behavior, and (iii) *envy-freeness*, which captures the degree to which a slice would prefer another slice's allocations across the network resources. The first result holds for any utility function, while the other two assume that slice utilities are 1-fair homogeneous i.e., $U^o(\mathbf{w}) = \sum_{u\in\mathcal{U}^o} \phi_u \log(r_u(\mathbf{w}))\ \forall o \in \mathcal{O}$ – a widely accepted case leading to the well-known proportionally fair allocations.

A. Protection: Gain over Static Slicing

We first analyze if strategic behavior on the part of network slices may result in allocations that are worse that those under static slicing. Note that static slicing provides complete isolation among slices but potentially poor utilization. A critical question is whether dynamic sharing, which achieves a higher resource utilization, also provides the same level of protection. This is confirmed by the following result.[3]

Lemma 3. *Consider slice o and any feasible weight allocation $\mathbf{w}^{-o}$ for other slices satisfying the network share constraints. Then, there exists a weight allocation $\mathbf{w}^o$ for slice o, possibly dependent on $\mathbf{w}^{-o}$, such that the resulting weight allocation $\mathbf{w}$ satisfies $r_u(\mathbf{w}) \ge r_u^{ss}$ for all $u \in \mathcal{U}_o$.*

The lemma is easily shown by choosing $\mathbf{w}^o$ such that

$$w_u = \frac{w_u^{ss}}{\sum_{u\in\mathcal{U}_{b(u)}^o} w_u^{ss}} \frac{a_{b(u)}^o(\mathbf{w}^{-o})}{\sum_{b'\in\mathcal{B}_o} a_{b'}^o(\mathbf{w}^{-o})} s_o,\ \forall u \in \mathcal{U}^o$$

where $\mathcal{B}^o$ is the set of base stations where slice o has users. The intuitive interpretation for this choice is that by distributing its weights proportionally to the load at each base station, slice o can achieve the same resource allocation as static slicing at each base station. Further, by redistributing these allocations amongst its user in the same manner as static slicing, it achieves at least as much rate per user.

It follows immediately from the above lemma that under the SCPA resource allocation model, if all slices exhibit strategic behavior attempting to maximize their utilities, they necessarily achieve a higher utility than under static slicing. This result does not require slices to have homogenous or concave utilities, just that they be increasing in the users' rate allocations.

Theorem 3. *If the game where each network slice maximizes its utility has a Nash equilibrium, then each slice achieves a higher utility than under static slicing.*

The above results guarantee a *form of resource isolation*, since they show that (i) a slice can always choose a weight assignment that provides its users with the same or higher rates than those provided by static slicing (i.e., full isolation), and (ii) by choosing a smarter weight allocation, a slice is guaranteed to experience better performance than that achieved with full isolation.

B. Efficiency: Price of Anarchy

In the following, we analyze the *price of anarchy* for $\alpha_o = 1$ (i.e., 1-fair homogenous utilities). We define the price of anarchy as the difference between the *overall network utility* resulting from the socially optimal allocation, $U(\mathbf{w}^*)$, and that obtained at a Nash equilibrium of the SCPA resource allocation mechanism, $U(\mathbf{w})$; such a notion captures the *efficiency* of the

[3]The proofs of all lemmas and theorems are provided in the Appendix or as supplementary material.

proposed approach as it shows how far it performs from the optimal.[4]

The following result characterizes the socially optimal allocation of resources considered in the above definition for the price of anarchy (see [34]).

Fact 1. *For slices with* 1*-fair homogenous utilities, the socially optimal allocation of resources* $\mathbf{w}^*$ *is such that* $w_u^* = \phi_u s_o, \quad \forall u \in \mathcal{U}^o$ *and* $\forall o \in \mathcal{O}$.

Building on the above result, the following theorem bounds the price of anarchy – the proof is provided in the Appendix.

Theorem 4. *If all slices have* 1*-fair homogenous utilities, then the Price of Anarchy (PoA) associated with a Nash equilibrium* $\mathbf{w}$ *satisfies*

$$PoA := U(\mathbf{w}^*) - U(\mathbf{w}) \leq \log(e).$$

Furthermore, there exists a game instance for which this bound is tight.

Note that, with 1-fair utilities, if we increase the capacity of all resources by a factor Δc, we have a utility increase of $\log(\Delta c)$. Thus, the performance improvement achieved by the socially optimal allocation over SCPA is (in the upper bound) equivalent to having a capacity e times larger, i.e., almost the triple capacity. While there are some (pathological) cases in which such a bound can be achieved, our simulation results show that for practical scenarios the actual performance difference between the two allocations is much smaller, confirming that (for $\alpha_o = 1$) the flexibility gained with the SCPA mechanism comes at a very small price in performance.

C. Fairness: Envy-freeness

Next we consider a Nash equilibrium $\mathbf{w}$ and analyze whether a slice, say o, with utility $U^o(\mathbf{w})$, might have a better utility if it were to exchange its resources with those of another slice, say o'. To that end, we denote by $\tilde{\mathbf{w}}$ the resulting weight allocation when the users of slices o and o' exchange their allocated resources. It is easy to see that $\tilde{\mathbf{w}}^o$ is such that

$$\tilde{w}_u^o = \frac{\phi_u}{\sum_{v \in \mathcal{U}_b^o} \phi_v} d_b^{o'}(\mathbf{w}) \text{ for all } b \in \mathcal{B} \text{ and all } u \in \mathcal{U}_b^o, \quad (4)$$

i.e., slice o takes the aggregate weight of o' at base station b under the Nash equilibrium, $d_b^{o'}(\mathbf{w})$, and allocates it proportionally to its user priorities. Clearly, $\tilde{\mathbf{w}}^{o'}$ is defined similarly and the remaining slices weights remain unchanged under $\tilde{\mathbf{w}}$.

We define the envy of slice o for o''s resources under the Nash equilibrium $\mathbf{w}$ by

$$E^{o,o'} := U^o(\tilde{\mathbf{w}}) - U^o(\mathbf{w}).$$

Note that envy is a "directed" concept, i.e., it is defined from slice o's point of view. When $E^{o,o'} \leq 0$, we say slice o is not envious. The following theorem provides a bound on $E^{o,o'}$.

[4]Note that the price of anarchy is typically defined as the ratio of the social optimal network utility and the network utility under the NE. However, since with $\alpha_o = 1$ we have logarithmic shaped utilities, the utility loss in this case is better expressed by the subtraction rather than by the ratio; as a matter of fact, the ratio in linear scale corresponds to a subtraction in the logarithmic scale.

Theorem 5. *Consider a slice* o *with* 1*-fair homogeneous utilities and the remaining slices* $\mathcal{O} \setminus \{o\}$ *with arbitrary slice utilities. Consider a slice* o' *such that* $s_o = s_{o'}$*. Let* $\mathbf{w}$ *denote a Nash equilibrium and* $\tilde{\mathbf{w}}$ *denote the resulting weights when* o *and* o' *exchange their resources. Then, the envy of slice* o *for* o' *satisfies*

$$E^{o,o'} = U^o(\tilde{\mathbf{w}}) - U^o(\mathbf{w}) \leq 0.060.$$

Furthermore, there is a game instance where $0.041 \leq E_{o,o'}$.

Given that, if one increases the rates of all users by a factor Δr this yields a utility increase of $\log(\Delta r)$, one can interpret this result as saying that, by exchanging resources with o', slice o may see a gain equivalent to increasing the rate of all its users by a factor between 4.1% and 6.1% (given by the lower and upper bounds of the above theorem). This is quite low and, moreover, simulation results show that in practical settings there is actually (almost) never any envy, confirming that our system is (practically) envy-free.

V. Performance Evaluation

Next, we evaluate the performance of the SCPA resource allocation mechanism via simulation. The mobile network scenario considered is based on the IMT-A evaluation guidelines for dense 'small cell' deployments [35], which consider base stations with an intersite distance of 200 meters in a hexagonal cell layout with 3 sector antennas and a network size $|\mathcal{B}|$ of 57 sectors.[5] Unless otherwise stated, users move according to the Random Waypoint Model (RWP) and user association follows the strongest signal policy. The Signal Interference to Noise Ratio of user u (SINR$_u$) is computed based on physical layer network model specified in [35]: $\text{SINR}_u = P_b g_{ub} / (\sum_{k \in \mathcal{B}, k \neq b} P_k g_{uk} + \sigma^2)$, where P_b is the transmit power and g_{ub} denotes the channel gain between user u and base station b, which includes path loss, shadowing, fast fading and antenna gain. Following [35], we set $P_b = 41$ dBm, $\sigma^2 = -104$dB, a path loss equal to $36.7 \log_{10}(\text{dist}) + 22.7 + 26 \log_{10}(f_c)$ for carrier frequency $f_c = 2.5$GHz, and an antenna gain of 17 dBi. The shadowing factor is given by a log-normal function with a standard deviation of 8dB (as in [36]) updated every second, and fast fading follows a Rayleigh distribution dependent of the user speed and the angle of incidence (as in [37]). Achievable rates are then computed with the Shannon formula, $\text{BW} \log_2(1 + \overline{\text{SINR}}_u)$, for the average $\overline{\text{SINR}}_u$ and a channel bandwidth of $\text{BW} = 10$MHz [38]. Finally, the modulation-coding scheme is selected according to the $\overline{\text{SINR}}_u$ thresholds reported in [39]. For all our simulation results, we obtained 95% confidence intervals with relative errors below 1% (not shown in the figures).

Given the nature of elastic traffic, the performance of which mainly depends on rate, the evaluation conducted in this section primarily focuses on this metric. However, we note that as long as users are allocated a sufficiently high rate for the traffic they generate, a slice can ensure low delays and drop rates for its users. This can be achieved, e.g., by acquiring

[5]Note that, in this setting, users associate with sectors rather than the base stations we used in the mechanism description and analysis.

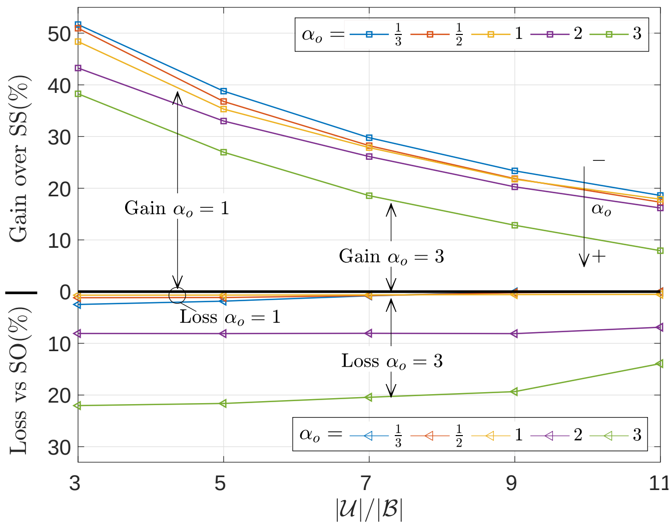

Fig. 2: Average Gain over Static slicing and Loss against Social optimum for different scenarios.

an appropriate share s_0 and/or performing proper admission control on the users entering the slice.

A. Overall performance

Throughout the paper we have used *static slicing* and the *socially optimal* resource allocations as our baselines. In order to confirm our analytical results and gain additional insights, we have evaluated the performance of the SCPA mechanism versus these two baselines via simulation. As an intuitive metric for comparison, we have used the extra capacity required by these baseline schemes to deliver the same performance as SCPA: (i) *Gain over SS*: additional resources required by *static slicing* to provide the same utility as SCPA (in %); and (ii) *Loss versus SO*: additional resources required by SCPA to provide the same utility as the *socially optimal* allocation (in %). Note that the latter metric is closely related to the *Price of Anarchy* analyzed in Section IV-B; indeed, while the Price of Anarchy reflects the loss of efficiency in terms of utility, the Loss versus SO reflects the loss of efficiency in terms of resources.

The results shown in Figure 2 are for different user densities ($|\mathcal{U}|/|\mathcal{B}|$) and different slice utilities (α_o parameter). As expected, the SCPA mechanism always has a gain over static slicing and a loss over the social optimal. However, for $\alpha_o = 1$ the loss is well below the bound given in Section IV-B. We further observe that performance is particularly good as long as α_o does not exceed 1 (Gain over SS up to 50% and Loss over SO below 5%), and it degrades mildly as α_o increases.

B. Fairness

In addition to overall performance, it is of interest to evaluate the fairness of the resulting allocations. While in Section IV-C we derived analytically a bound on the envy, we have further explored this via simulation by evaluating up to 10^7 randomly generated scenarios, with parameters drawn uniformly in the ranges: $|\mathcal{O}| \in [2, 12]$, $|\mathcal{B}| \in [10, 90]$, $|\mathcal{U}|/|\mathcal{B}| \in [3, 15]$, $\alpha_o \in [0.01, 30]$ and $\boldsymbol{\phi}$ vectors in the simplex. Our results show that $E^{o,o'} < 0$ holds for *all* the cases explored, confirming that in practice the system is envy-free.

C. Protection against other slices

One of the main objectives of our proposed framework is to enable slices to customize their resource allocations. This can be done by adjusting (i) the user priorities ϕ_u, and (ii) the parameter α_o, which regulates the desired level of fairness among the slice's users. In order to evaluate the impact that these settings have amongst slices, we simulated a scenario with three slices: Slice 1 has $\alpha_1 = 1$, Slice 2 has $\alpha_2 = 4$, and Slice 3 has α_3 with varying values. For simplicity, we set the priorities ϕ_u equal for all users.

Figure 3 shows the rate distributions of the 3 slices. We observe that the choice of α_3 is effective in adjusting the level of user fairness for Slice 3; indeed, as α_3 grows, the rate distribution becomes more homogeneous. Such customization at Slice 3 has a higher impact on Slice 1 than on Slice 2. This is the case because, as α_2 is quite large, the distribution of Slice 2's rates remains homogeneous, making the slice fairly insensitive to the choices of the other slices. As can be seen in the subplots, the utilities of Slices 1 and 2 are not only larger than the utility of static slicing, but remain fairly insensitive to α_3, showing that in both cases we have a good level of protection between slices.

D. Gains over Static Slicing

In order to gain additional insight into the impact of the various factors, Figure 4 displays the influence of the number of slices ($|\mathcal{O}|$) and the average load per base station sector ($|\mathcal{U}|/|\mathcal{B}|$) on the gain over static slicing (given by the additional resources in % required by static slicing to provide the same performance as SCPA). The results show that the gains are higher with a large number of slices and small load. This is rather intuitive: (i) if the slice has a large number of users, its multiplexing gain is already high without sharing the infrastructure, and hence there is little gain from sharing, and (ii) if we have a small number of slices, each of them is already using a large fraction of the network resources and the impact of sharing is smaller. It is also worth noting that α_o has a relatively small impact on the gains.

E. File download times

To provide additional insights on the achieved gains from an end-user perspective, we compare the file download times achieved by SCPA against the static slicing baseline, when base stations have the same capacity in both cases and users download and endless sequence of files of fixed size. Let us define the file download time gain as $G_D = (D_{SS} - D_{SCPA})/D_{SS}$, where D_{SS} is the average file download time with the static slicing approach and D_{SCPA} with the SCPA approach. The gains achieved are shown in Figure 5 for a file size of 20 Mb (results with other file sizes, not reported here for space reasons, show a similar trend). We observe the gains are substantial, and (as above) they are larger for low user densities and a large number of slices.

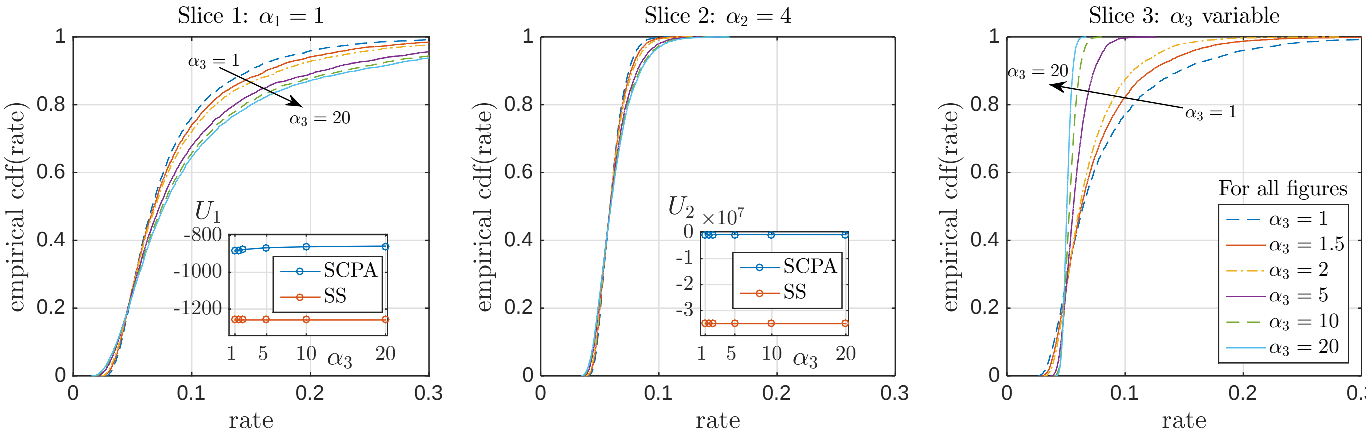

Fig. 3: Impact of α_3 decision on the slice rate distributions.

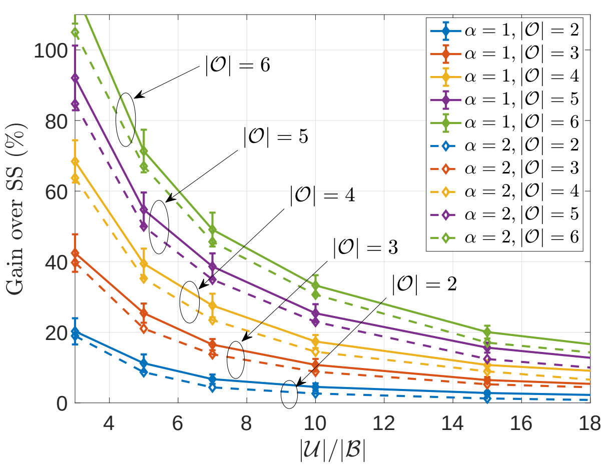

Fig. 4: Gains over static slicing for different settings.

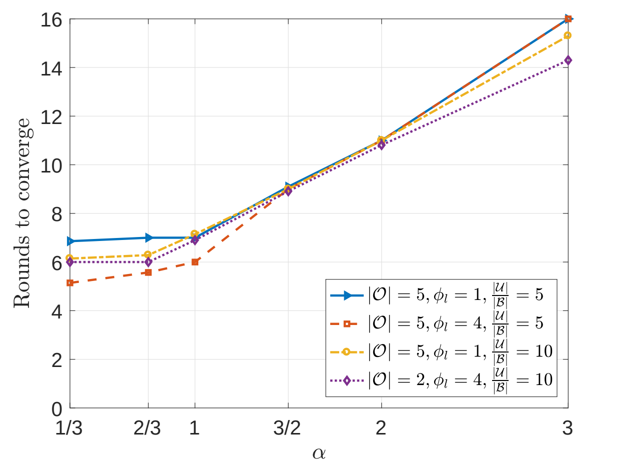

Fig. 6: Average number of rounds until convergence for different scenarios.

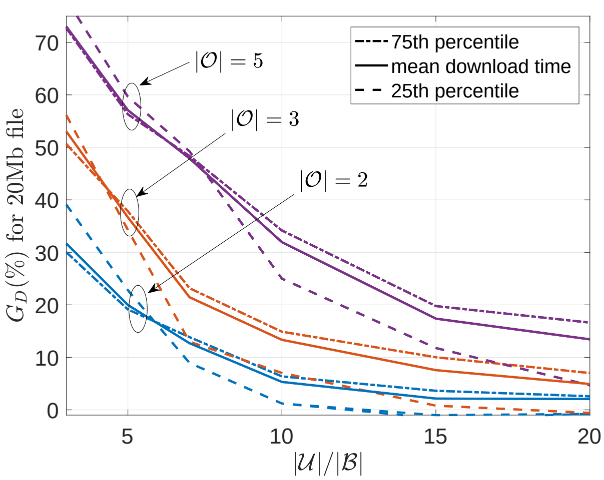

Fig. 5: Gains in terms of file download times.

F. Convergence speed

The existence of a Nash equilibrium and the convergence of Best Response Dynamics are essential for the system stability. While the existence of a Nash equilibrium has been proven for all α_o values, convergence has only been shown for $\alpha_o \in [1, 2]$. In order to confirm the convergence for other α_o values, we have conducted extensive simulations implementing sequential best response updates for up to 10^7 randomly generated scenarios within the same parameter space as in Section V-B. Our results confirmed the convergence of the best response game in all cases. Moreover, they also showed that convergence speed mainly depends on α_o, while it is fairly insensitive to the user priorities and the network size. According to the results, convergence is very quick for $\alpha_o \leq 1$ (about 8 rounds) and increases slightly as α_o grows (about 16 rounds for $\alpha_o = 3$). The average number of rounds needed for the Best Response dynamics to converge are shown in Figure 6.

G. Impact of user mobility

The above results assume a Random Waypoint mobility model where users are (on average) uniformly distributed across space. To understand the impact of other user distributions, we evaluated the *Gain over SS* for the following user mobility patterns: (i) *uniform*: all slices with a uniform spatial load distribution; (ii) *overlapping hotspots*: all slices with the same non-uniform spatial load distribution; (iii) *non-overlapping hotspots*: different slices with different non-uniform spatial load distributions; and (iv) *mixed*: half of

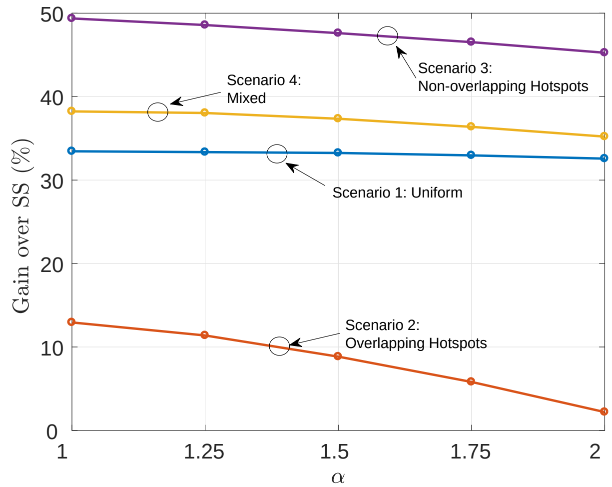

Fig. 7: Gain over Static slicing for different traffic models and α values.

the slices with a uniform spatial load distribution and the other half with a non-uniform one. In all cases, we have 4 slices with equal shares. The results, depicted in Figure 7, show that the gains are larger for scenarios with uneven and complementary traffic loads; indeed, in this case different slices need their resources at different base stations and thus there is a higher gain from dynamically sharing the resources. We further observe that larger α values result in smaller gain; this is due to the fact that slices are less elastic with larger α, which limits the ability to exploit statistical multiplexing.

VI. Conclusions

In this paper we have analyzed a 'share-constrained proportional allocation' framework for network slicing. The framework allows slices to customize the resource allocation to their users, leading to a *network slicing game* in which each slice reacts to the settings of the others. Our main conclusion is that the framework provides an *effective* and *implementable* scheme for dynamically sharing resources across slices. Indeed, this scheme involves simple operations at base stations and incurs a limited signaling between the slices and the infrastructure. Our results confirm system stability (*best response dynamics* converge), substantial gains over *static slicing*, and fairness of the allocations (*envy-freeness*). Moreover, as long as the majority of the slices do not choose α_o values larger than 1 (i.e., they do not all demand very homogeneous rate distributions), the overall performance is close to optimal (*price of anarchy* is very small). Thus, in this case the flexibility provided by this framework comes at no cost. If a substantial number of slices choose higher α_o's, then we pay a (small) price for enabling slice customization.

Appendix

In the following, we give the proofs of Lemma 3 and Theorems 3, 4 and 5. The proofs of Lemmas 1 and 2 and Theorems 1 and 2 have been provided as supplementary material.

Proof of Lemma 3

Given the weight allocation under static slicing, $\mathbf{w}^{ss}$, and the weights of the other slices under dynamic sharing, $\mathbf{w}^{-o}$, we consider the following weight allocation for slice o:

$$w_u = \frac{w_u^{ss}}{\sum_{u\in\mathcal{U}^o_{b(u)}} w_u^{ss}} \frac{a^o_{b(u)}(\mathbf{w}^{-o})}{\sum_{b'\in\mathcal{B}_o} a^o_{b'}(\mathbf{w}^{-o})} s_o, \tag{5}$$

where $\mathcal{B}_o$ is the set of base stations where slice o has customers.

We define $\rho^o_u(\mathbf{w}^{ss})$ as the ratio between the weight of user u under static slicing and the sum of the weights of all the users of the same slice in the base station, i.e.,

$$\rho^o_u(\mathbf{w}^{ss}) \doteq \frac{w_u^{ss}}{\sum_{u\in\mathcal{U}^o_{b(u)}} w_u^{ss}}$$

where we have dropped the terms $\mathbf{w}^{-o}$ and $\mathbf{w}^{ss}$ from $a^o_{b(u)}(\mathbf{w}^{-o})$ and $\rho^o_u(\mathbf{w}^{ss})$ for readability purposes.

With the allocation given by (5), for two users u and u' of slice o it holds

$$\frac{w_u}{w_{u'}} = \frac{\rho^o_u}{\rho^o_{u'}} \frac{a^o_{b(u)}}{a^o_{b(u')}} \tag{6}$$

Furthermore, it also holds that

$$d^o_b = \sum_{u\in\mathcal{U}^o_b} w_u = \sum_{u\in\mathcal{U}^o_b} \rho^o_u \frac{a^o_{b(u)}}{\sum\limits_{b'\in\mathcal{B}_o} a^o_{b'}} s_o = \frac{a^o_{b(u)}}{\sum\limits_{b'\in\mathcal{B}_o} a^o_{b'}} s_o = \frac{w_u}{\rho^o_u}$$

for $u \in \mathcal{U}^o_b$.

From the above expression, we have

$$\frac{\rho^o_u l_{b(u)}(\mathbf{w})}{\rho^o_{u'} l_{b(u')}(\mathbf{w})} = \frac{\rho^o_u \left(a^o_{b(u)} + d^o_{b(u)}\right)}{\rho^o_{u'} \left(a^o_{b(u')} + d^o_{b(u')}\right)} = \frac{\rho^o_u \left(a^o_{b(u)} + \frac{w_u}{\rho^o_u}\right)}{\rho^o_{u'} \left(a^o_{b(u')} + \frac{w_{u'}}{\rho^o_{u'}}\right)},$$

and combining this with (6):

$$\begin{aligned}\frac{\rho^o_u l_{b(u)}(\mathbf{w})}{\rho^o_{u'} l_{b(u')}(\mathbf{w})} &= \frac{\rho^o_u \left(a^o_{b(u)} + \frac{a^o_{b(u)}}{a^o_{b(u')}} \frac{w_{u'}}{\rho^o_{u'}}\right)}{\rho^o_{u'} \left(a^o_{b(u')} + \frac{w_{u'}}{\rho^o_{u'}}\right)} \\ &= \frac{\rho^o_u a^o_{b(u)} \left(1 + \frac{w_{u'}}{a^o_{b(u')}\rho^o_{u'}}\right)}{\rho^o_{u'} a^o_{b(u')} \left(1 + \frac{w_{u'}}{a^o_{b(u')}\rho^o_{u'}}\right)} = \frac{\rho^o_u a^o_{b(u)}}{\rho^o_{u'} a^o_{b(u')}}\end{aligned}$$

From the above,

$$\begin{aligned} w_u &= \frac{w_u}{\sum_{u'\in\mathcal{U}^o} w_{u'}} s_o = \frac{s_o}{\sum_{u'\in\mathcal{U}^o} \frac{w_{u'}}{w_u}} = \\ &= \frac{s_o}{\sum\limits_{u'\in\mathcal{U}^o} \frac{\rho^o_{u'} a^o_{b(u')}}{\rho^o_u a^o_{b(u)}}} = \frac{s_o}{\sum\limits_{u'\in\mathcal{U}^o} \frac{\rho^o_{u'} l_{b(u')}(\mathbf{w})}{\rho^o_u l_{b(u)}(\mathbf{w})}} \\ &= \frac{\rho^o_u l_{b(u)}(\mathbf{w})}{\sum_{u'\in\mathcal{U}^o} \rho^o_{u'} l_{b(u')}(\mathbf{w})} s_o = \frac{\rho^o_u l_{b(u)}(\mathbf{w})}{\sum_{b'\in\mathcal{B}_o} l_{b'}(\mathbf{w})} s_o \end{aligned}$$

Since $\mathcal{B}_o \subseteq \mathcal{B}$:

$$\sum_{b\in\mathcal{B}_o} l_b(\mathbf{w}) \leq \sum_{b\in\mathcal{B}} l_b(\mathbf{w}) = 1$$

and thus

$$w_u \geq \rho^o_u l_{b(u)}(\mathbf{w}) s_o,$$

from which

$$r_u(\mathbf{w}) = \frac{w_u}{l_{b(u)}(\mathbf{w})} c_u \geq \frac{\rho_u^o l_{b(u)}(\mathbf{w}) s_o}{l_{b(u)}(\mathbf{w})} c_u = \rho_u^o s_o c_u = r_u^{ss}.$$

The above holds for all $u \in \mathcal{U}$, which proves the lemma. ∎

Proof of Theorem 3

This result follows from Lemma 3. Given the configuration of the other slices, there exists a configuration for a given slice under which all its users obtain at least the same throughput as with static slicing, and thus the slice's utility with this configuration is at least as high. As a consequence, in a NE the slice will receive a utility no smaller than this value. ∎

Proof of Theorem 4

We first show that an optimal (not necessarily unique) solution to the centralized problem is given by $\mathbf{w}^*$ which assigns weights to all users of a given slice proportionally to their priorities, i.e., $w_u^* = \phi_u s_o, \ \forall u \in \mathcal{U}_o$. To prove this we only need to show that $U(\mathbf{w}^*) \geq U(\mathbf{w})$ for any other feasible weight vector $\mathbf{w}$. To that end, consider

$$U(\mathbf{w}^*) - U(\mathbf{w}) = \sum_{o\in\mathcal{O}} \sum_{u\in\mathcal{U}_o} \phi_u \left(\log\left(\frac{w_u^* c_u}{l_b(\mathbf{w}^*)} \right) - \log\left(\frac{w_u c_u}{l_b(\mathbf{w})} \right) \right)$$

Let us denote the distributions induced by $\mathbf{w}^*$ and $\mathbf{w}$ respectively as: $\mathbf{p}^b(\mathbf{w}) = (p_u^b(\mathbf{w}) = \frac{w_u}{l_b(\mathbf{w})} : u \in \mathcal{U}_b)$ and $\mathbf{p}^b(\mathbf{w}^*) = (p_u^b(\mathbf{w}^*) = \frac{w_u}{l_b(\mathbf{w}^*)} : u \in \mathcal{U}_b)$. Since $\boldsymbol{\phi} = \mathbf{w}^*$, we have

$$\begin{aligned} U(\mathbf{w}^*) - U(\mathbf{w}) &= \sum_{b\in\mathcal{B}} l_b(\mathbf{w}^*) \sum_{o\in\mathcal{O}} \sum_{u\in\mathcal{U}_o^b} p_u^b(w^*) \log\left(\frac{p_u^b(w^*)}{p_u^b(w)} \right) \\ &= \sum_{b\in\mathcal{B}} l_b(\mathbf{w}^*) D(\mathbf{p}^b(\mathbf{w}^*) || \mathbf{p}^b(\mathbf{w})) \end{aligned}$$

where $D(\mathbf{p}^b(\mathbf{w}^*)||\mathbf{p}^b(\mathbf{w}))$ is the Kullback-Leibler divergence, between the distributions induced by $\mathbf{w}^*$ and $\mathbf{w}$ respectively, i.e., $\mathbf{p}^b(\mathbf{w}^*)$ and $\mathbf{p}^b(\mathbf{w})$. It is known [40] that $D(\mathbf{p}^b(\mathbf{w}^*)||\mathbf{p}^b(\mathbf{w})) \geq 0$ and 0 only when $\mathbf{p}^b(\mathbf{w}) = \mathbf{p}^b(\mathbf{w}^*)$ Hence it follows that $\mathbf{w}^*$ is optimal.

We next show that $U(\mathbf{w}^*) - U(\mathbf{w}) \leq \log(e)$ holds when $\mathbf{w}$ is a Nash Equilibrium of the distributed resource allocation game and $\mathbf{w}^*$ an optimal solution. To show this, we proceed as follows. Since in the Nash Equilibrium each slice maximizes its utility given the allocation of the other slices,

$$\sum_{u\in\mathcal{U}^o} \phi_u \log\left(\frac{w_u}{l_{b(u)}(\mathbf{w})} \right) \geq \sum_{u\in\mathcal{U}^o} \phi_u \log\left(\frac{w_u^*}{d_{b(u)}^o(\mathbf{w}^*) + a_{b(u)}^o(\mathbf{w})} \right)$$

Given that $d_{b(u)}^o(\mathbf{w}^*) + a_{b(u)}^o(\mathbf{w}) \leq l_{b(u)}(\mathbf{w}) + l_{b(u)}(\mathbf{w}^*)$,

$$\sum_{u\in\mathcal{U}^o} \phi_u \log\left(\frac{w_u}{l_{b(u)}(\mathbf{w})} \right) \geq \sum_{u\in\mathcal{U}^o} \phi_u \log\left(\frac{w_u^*}{l_{b(u)}(\mathbf{w}) + l_{b(u)}(\mathbf{w}^*)} \right)$$

From the above it follows that

$$\begin{aligned} &\sum_{u\in\mathcal{U}^o} \phi_u \log(r_u(\mathbf{w}^*)) - \sum_{u\in\mathcal{U}^o} \phi_u \log(r_u(\mathbf{w})) \\ &\quad \leq \sum_{u\in\mathcal{U}^o} \phi_u \log\left(\frac{w_u^* c_u}{l_{b(u)}(\mathbf{w}^*)} \right) - \sum_{u\in\mathcal{U}^o} \phi_u \log\left(\frac{w_u^* c_u}{l_{b(u)}(\mathbf{w}) + l_{b(u)}(\mathbf{w}^*)} \right) \\ &\quad = - \sum_{u\in\mathcal{U}^o} \phi_u \log\left(\frac{l_{b(u)}(\mathbf{w}^*)}{l_{b(u)}(\mathbf{w}) + l_{b(u)}(\mathbf{w}^*)} \right) \end{aligned}$$

Summing the above over all slices weighted by the corresponding shares yields

$$U(\mathbf{w}^*) - U(\mathbf{w}) \leq - \sum_{u\in\mathcal{U}} \phi_u s_o \log\left(\frac{l_{b(u)}(\mathbf{w}^*)}{l_{b(u)}(\mathbf{w}) + l_{b(u)}(\mathbf{w}^*)} \right)$$

Given $w_u^* = \phi_u s_o$, we have

$$\begin{aligned} U(\mathbf{w}^*) - U(\mathbf{w}) &\leq - \sum_{b\in\mathcal{B}} \log\left(\frac{l_b(\mathbf{w}^*)}{l_b(\mathbf{w}) + l_b(\mathbf{w}^*)} \right)^{\sum_{u\in\mathcal{U}_b} w_u^*} \\ &= - \sum_{b\in\mathcal{B}} \sum_{u\in\mathcal{U}_b} w_u \log\left(\frac{l_b(\mathbf{w}^*)}{l_b(\mathbf{w}) + l_b(\mathbf{w}^*)} \right)^{\frac{\sum_{v\in\mathcal{U}_b} w_v^*}{\sum_{v\in\mathcal{U}_b} w_v}} \\ &= - \sum_{b\in\mathcal{B}} \sum_{u\in\mathcal{U}_b} w_u \log\left(\frac{l_b(\mathbf{w}^*)/l_b(\mathbf{w})}{1 + l_b(\mathbf{w}^*)/l_b(\mathbf{w})} \right)^{\frac{l_b(\mathbf{w}^*)}{l_b(\mathbf{w})}} \end{aligned}$$

and, given that $(x/(1+x))^x > 1/e$ for $x \geq 0$, this yields

$$U(\mathbf{w}^*) - U(\mathbf{w}) \leq \sum_{b\in\mathcal{B}} \sum_{u\in\mathcal{U}_b} w_u \log(e) = \log(e).$$

Finally, we show that there exists some scenario for which $U(\mathbf{w}^*) - U(\mathbf{w}) = \log(e)$. Let us consider a scenario with two slices with shares s_1 and s_2, respectively. There are two base stations. Slice 1 has $m+1$ users, m associated to base station 1 and one associated to base station 2. Slice 2 has one user associated to base station 2. All users have $c_{ub} = 1$. Under the optimal allocation:

$$\begin{aligned} U(\mathbf{w}^*) = &\frac{s_1}{m+1} m \log\left(\frac{1}{m} \right) + \frac{s_1}{m+1} \log\left(\frac{\frac{s_1}{m+1}}{\frac{s_1}{(m+1)} + s_2} \right) \\ &+ s_2 \log\left(\frac{s_2}{\frac{s_1}{(m+1)} + s_2} \right), \end{aligned}$$

and under the Nash equilibrium

$$\begin{aligned} U(\mathbf{w}) = &\frac{s_1}{m+1} m \log\left(\frac{1}{m} \right) + \frac{s_1}{m+1} \log\left(\frac{s_1}{s_1 + s_2} \right) \\ &+ s_2 \log\left(\frac{s_2}{s_1 + s_2} \right). \end{aligned}$$

For $m \to \infty$ this yields $U(\mathbf{w}^*) = s_1 \log\left(\frac{1}{m}\right) + s_2 \log(1)$ and $U(\mathbf{w}) = s_1 \log\left(\frac{1}{m}\right) + s_2 \log\left(\frac{s_2}{s_1+s_2}\right)$. From this,

$$U(\mathbf{w}) - U(\mathbf{w}^*) = s_2 \log\left(\frac{s_2}{s_1 + s2} \right)$$

which tends to $-\log(e)$ when $s_1 \to 1$ and $s_2 \to 0$. ∎

Proof of Theorem 5

Let us consider two slices, o and o', that have the same share s_o. Let the utility function of slice o be $U^o = \sum_{u\in\mathcal{U}^o} \phi_u \log(r_u)$. We first show that it holds

$$U^o(\tilde{\mathbf{w}}_\mathbf{o}) - U^o(\mathbf{w}_\mathbf{o}) \leq 0.060$$

In order to bound the envy $U^o(\tilde{\mathbf{w}}) - U^o(\mathbf{w})$ at the NE, we will construct a weight allocation $\mathbf{m}$ that satisfies $U^o(\mathbf{m}) \leq U^o(\mathbf{w})$ and $U^o(\tilde{\mathbf{m}}) \geq U^o(\tilde{\mathbf{w}})$ – where $\tilde{\mathbf{w}}$ and $\tilde{\mathbf{m}}$ are the allocations resulting from exchanging the resources of slices o and o' in $\mathbf{w}$ and $\mathbf{m}$, respectively. It then follows that $U^o(\tilde{\mathbf{m}}) - U^o(\mathbf{m})$ is an upper bound on the envy.

Specifically, the weight allocation $\mathbf{m}$ will be chosen such that: (i) for all slices different from o, the weights remain the same as in the NE, i.e, $\mathbf{m}^{-o} = \mathbf{w}^{-o}$; and (ii) the weights of slice o are chosen so as to maximize $U^o(\mathbf{m})$ subject to $d_b^o(\mathbf{m}^o) = \sum_{u\in\mathcal{U}_b^o} m_u \leq a_b^o(\mathbf{m}^{-o})\ \forall b \in \mathcal{B}$ and slice o's share constraint. Note that with this weight allocation we have $a_{b(u)}^o(\mathbf{m}^{-o}) = a_{b(u)}^o(\mathbf{w}^{-o})$ – for readability purposes, we will use just $a_{b(u)}^o$. Note also that the weights that slice o would have with the resources of o' remain the same, i.e. $\tilde{\mathbf{m}}^o = \tilde{\mathbf{w}}^o$.

By following a similar argument to that of Lemma 2, it can be seen that the above leads to the weights m_u for $u \in \mathcal{U}^o$ solving the set of equations below, which have a feasible solution as long as $s_o < \sum_{u\in\mathcal{U}^o} a_{b(u)}^o(\mathbf{m}^{-o})$ (we deal with the case $\sum_{b\in\mathcal{B}_o} a_b^o < s_o$ later).

$$m_u = \begin{cases} a_{b(u)}^o \dfrac{\phi_u}{\sum_{v\in\mathcal{U}_{b(u)}^o}\phi_v}, & a_{b(u)}^o = d_{b(u)}^o(\mathbf{m}^o) \\ \dfrac{\phi_u \frac{a_{b(u)}^o}{a_{b(u)}^o + d_{b(u)}^o(\mathbf{m}^o)}}{\sum\limits_{v\in\hat{\mathcal{U}}^o} \phi_v \frac{a_{b(v)}^o}{a_{b(v)}^o + d_{b(v)}^o(\mathbf{m}^o)}} s_o', & a_{b(u)}^o > d_{b(u)}^o(\mathbf{m}^o) \end{cases}$$

where $\hat{\mathcal{U}}^o$ is the set of users of slice o for which $a_{b(u)}^o > d_{b(u)}^o(\mathbf{m}^o)$ and $s_o' = s_o - \sum_{u\in\mathcal{U}^o\setminus\hat{\mathcal{U}}^o} m_u$.

It is clear that with this weight allocation we have $U^o(\mathbf{m}) \leq U^o(\mathbf{w})$. Indeed, only the weights of slice o have changed and (as mentioned before) $\mathbf{w}^o$ is the best response of the slice o, hence any other weight setting for this slice will provide a lower utility.

To show $U^o(\tilde{\mathbf{m}}) \geq U^o(\tilde{\mathbf{w}})$ we proceed as follows. The base stations that initially had a load for operator o larger than a_b^o ($d_{b(u)}^o(\mathbf{m}^o) > a_b^o$) decrease their load with the new allocation, while the others increase it. Let us denote the first set of base stations as $\mathcal{B}_1$ and the other set as $\mathcal{B}_2$. Since the base stations of set $\mathcal{B}_1$ decrease their load in the new allocation and the base stations of set $\mathcal{B}_2$ increase it, we can move from the initial allocation to the new one by iteratively selecting one base station of set $\mathcal{B}_1$ and one of set $\mathcal{B}_2$ and moving load from the first one to the second until one of them reaches its target load. When decreasing the load of base station b and increasing that of base station b' by δ we have

$$\frac{dU^o(\tilde{\mathbf{w}})}{d\delta} = -\frac{\sum\limits_{u\in\mathcal{U}_{b'}^o}\phi_u}{l_{b'}(\tilde{\mathbf{w}})} + \frac{\sum\limits_{u\in\mathcal{U}_b^o}\phi_u}{l_b(\tilde{\mathbf{w}})}$$

If we can show at the beginning (before increasing/decreasing the load of any base station), for any $b \in \mathcal{B}_1$ and $b' \in \mathcal{B}_2$ it holds

$$\frac{\sum\limits_{u\in\mathcal{U}_b^o}\phi_u}{l_b(\tilde{\mathbf{w}})} \geq \frac{\sum\limits_{u\in\mathcal{U}_{b'}^o}\phi_u}{l_{b'}(\tilde{\mathbf{w}})} \tag{7}$$

we will have the value of $\sum_{u\in\mathcal{U}_b^o} \frac{\phi_u}{l_b}$ for any base station of set $\mathcal{B}_1$ will always be larger than for any base station of set $\mathcal{B}_2$, since it are larger at the beginning and it increases in the intermediate steps, while it decreases for a base station of $\mathcal{B}_2$.
With this, $dU^o(\tilde{\mathbf{m}})/d\delta$ is positive at the beginning and will continue to be positive in the intermediate steps, yielding to an increase in $dU^o(\tilde{\mathbf{m}})$.

To show (7), we proceed as follows. It holds that

$$\begin{aligned}\frac{d_b^o(\mathbf{m}^o)}{d_{b'}^o} &= \frac{\sum_{u\in\mathcal{U}_b^o}\phi_u}{\sum_{u'\in\mathcal{U}_{b'}^o}\phi_{u'}} \frac{\frac{1}{1+d_b^o(\mathbf{m}^o)/a_b^o}}{\frac{1}{1+d_{b'}^o(\mathbf{m}^o)/a_{b'}^o}} \\ &= \frac{\sum_{u\in\mathcal{U}_b^o}\phi_u}{\sum_{u'\in\mathcal{U}_{b'}^o}\phi_{u'}} \left(\frac{1 + \frac{d_{b'}^o(\mathbf{m}^o)}{a_{b'}^o}}{1 + \frac{d_b^o(\mathbf{m}^o)}{a_b^o}} \right)\end{aligned}$$

For $b \in \mathcal{B}_1$ and $b' \in \mathcal{B}_2$ (since $a_b^o < d_b^o(\mathbf{m}^o)$ and $a_{b'}^o > d_{b'}^o(\mathbf{m}^o)$)

$$\frac{d_b^o(\mathbf{m}^o)}{d_{b'}^o(\mathbf{m}^o)} < \frac{\sum_{u\in\mathcal{U}_b^o}\phi_u}{\sum_{u'\in\mathcal{U}_{b'}^o}\phi_{u'}}$$

and thus

$$\begin{aligned}\frac{l_b}{\sum_{u\in\mathcal{U}_b^o}\phi_u} &= \frac{a_b^o + d_b^o(\mathbf{m}^o)}{\sum_{u\in\mathcal{U}_b^o}\phi_u} < \frac{2d_b^o(\mathbf{m}^o)}{\sum_{u\in\mathcal{U}_b^o}\phi_u} < \frac{2d_{b'}^o(\mathbf{m}^o)}{\sum_{u\in\mathcal{U}_{b'}^o}\phi_u} \\ &\leq \frac{a_{b'}^o + d_{b'}^o(\mathbf{m}^o)}{\sum_{u\in\mathcal{U}_{b'}^o}\phi_u} = \frac{l_{b'}}{\sum_{u\in\mathcal{U}_{b'}^o}\phi_u}\end{aligned}$$

which proves (7), and thus $U^o(\tilde{\mathbf{m}}) \geq U^o(\tilde{\mathbf{w}})$.

We now go back to the case $\sum_{b\in\mathcal{B}_o} a_b^o < s_o$. Following the above procedure, in this case we can find an allocation $\mathbf{m}^o$ that satisfies: (i) $U^o(\mathbf{m}) \leq U^o(\mathbf{w})$, (ii) $U^o(\tilde{\mathbf{m}}) \geq U^o(\tilde{\mathbf{w}})$ and (iii) $d_b^o(\mathbf{m}^o) \geq a_b^o\ \forall b$. In this case we then have $U^o(\tilde{\mathbf{w}}) - U^o(\mathbf{w}) \leq U^o(\tilde{\mathbf{m}}) - U^o(\mathbf{m}) \leq 0$.

To find an upper bound on $U^o(\tilde{\mathbf{m}}) - U^o(\mathbf{m})$, recall that

$$U^o(\tilde{\mathbf{m}}) = \sum_{u\in\mathcal{U}^o} \phi_u \log\left(\frac{\tilde{m}_u c_u}{l_b(\tilde{\mathbf{m}})}\right),$$

and

$$U^o(\mathbf{m}) = \sum_{u\in\mathcal{U}^o} \phi_u \log\left(\frac{m_u c_u}{l_b(\mathbf{m})}\right).$$

Given that $l_b(\tilde{\mathbf{m}}) = l_b(\mathbf{m})$ and $\tilde{m}_u = m_u$ for $u \notin \hat{\mathcal{U}}^o$, this yields

$$U^o(\tilde{\mathbf{m}}) - U^o(\mathbf{m}) = \sum_{u\in\hat{\mathcal{U}}^o} \phi_u \log(\tilde{m}_u) - \sum_{u\in\hat{\mathcal{U}}^o} \phi_u \log(m_u).$$

Since $\sum_{u\in\hat{\mathcal{U}}^o} \log(\tilde{m}_u)$ subject to $\sum_{u\in\hat{\mathcal{U}}^o} \tilde{m}_u = s_o'$ takes a maximum at $\tilde{m}_u = \hat{\phi}_u s_o'$ (where $\hat{\phi}_u = \phi_u / \sum_{v\in\hat{\mathcal{U}}^o}\phi_v$),

$$\begin{aligned}U^o(\tilde{\mathbf{m}}) - U^o(\mathbf{m}) &\leq \sum_{u\in\hat{\mathcal{U}}^o} \phi_u \log(\hat{\phi}_u s_o') - \sum_{u\in\hat{\mathcal{U}}^o} \phi_u \log(m_u) \\ &\leq \sum_{u\in\hat{\mathcal{U}}^o} \hat{\phi}_u \log(\hat{\phi}_u s_o') - \sum_{u\in\hat{\mathcal{U}}^o} \hat{\phi}_u \log(m_u)\end{aligned} \tag{8}$$

In order to bound the term $\sum_{u\in\hat{\mathcal{U}}^o} \hat{\phi}_u \log(m_u)$ above, we look for a bound on $\frac{m_u}{m_v}$. Given that $a_b^o \geq d_b^o(\mathbf{m}^o)$ holds for all b, we have for $u, v \in \hat{\mathcal{U}}^o$:

$$\frac{m_u}{m_v} = \frac{\phi_u}{\phi_v} \frac{\frac{a_{b(u)}^o}{a_{b(u)}^o + d_{b(u)}^o(\mathbf{m}^o)}}{\frac{a_{b(v)}^o}{a_{b(v)}^o + d_{b(v)}^o(\mathbf{m}^o)}} > \frac{\phi_u}{\phi_v} \frac{\frac{a_{b(u)}^o}{a_{b(u)}^o + a_{b(u)}^o}}{\frac{a_{b(v)}^o}{a_{b(v)}^o}} = \frac{1}{2}\frac{\hat{\phi}_u}{\hat{\phi}_v}.$$

It can be seen that $\sum_{u\in\hat{\mathcal{U}}^o} \hat{\phi}_u \log(m_u)$ subject to $\frac{m_u}{m_v} \geq \frac{1}{2}\frac{\hat{\phi}_u}{\hat{\phi}_v}$ and $\sum_{u\in\hat{\mathcal{U}}^o} \hat{\phi}_u = 1$ is maximized when the $\frac{m_u}{\hat{\phi}_u}$ of all users but one is equal to the lower bound given by the constraint, which yields

$$\frac{m_u}{\hat{\phi}_u} = \frac{1}{2}\frac{m_v}{\hat{\phi}_v},\ \forall u \neq v. \tag{9}$$

This is shown by contradiction. Let us imagine that in the weight allocation that maximizes (8) there exists some other user u for which $\frac{m_u}{\hat{\phi}_u} > \frac{m_v}{2\hat{\phi}_v}$, where v is the user with the largest $m_v/\hat{\phi}_v$ of that allocation. Then, if we increase m_v by δ and decrease m_u by δ we have

$$\frac{d}{d\delta}\sum_{u\in\mathcal{U}^o}\hat{\phi}_u \log\left(\frac{\hat{\phi}_u s_o'}{m_u}\right) = -\frac{\hat{\phi}_v}{m_v} + \frac{\hat{\phi}_u}{m_u} > 0$$

and thus (8) increases, which contradicts our assumption that (8) was already maximum. From (9) we have

$$m_u = \frac{\hat{\phi}_u s_o}{\sum_{u'\in\mathcal{U}^o\setminus\{v\}}\hat{\phi}_u + 2\hat{\phi}_v}, \text{ and } m_v = \frac{2\hat{\phi}_v s_o}{\sum_{u'\in\mathcal{U}^o\setminus\{v\}}\hat{\phi}_u + 2\hat{\phi}_v}$$

Combining this with (8) we obtain

$$\begin{aligned} U^o(\tilde{\mathbf{w}}_o^*) - U^o(\mathbf{w}_o^*) \leq &\sum_{u\in\mathcal{U}^o\setminus\{v\}}\hat{\phi}_u \log\left(\sum_{u'\in\mathcal{U}^o\setminus\{v\}}\hat{\phi}_u + 2\hat{\phi}_v\right) \\ &+ \hat{\phi}_v \log\left(\frac{1}{2}\sum_{u'\in\mathcal{U}^o\setminus\{v\}}\hat{\phi}_u + 2\hat{\phi}_v\right) \\ &= \log(1+\hat{\phi}_v) + \hat{\phi}_v\log(1/2) \end{aligned}$$

If we now compute the $\hat{\phi}_v$ that maximizes this expression we obtain $\hat{\phi}_v = \frac{1}{\log 2} - 1$, and substituting this value

$$U^o(\tilde{\mathbf{w}}_o^*) - U^o(\mathbf{w}_o^*) \leq -\log(\log 2) - \left(\frac{1}{\log 2} - 1\right)\log 2$$

As mentioned at the beginning, the above bounds also applies to $U^o(\tilde{\mathbf{w}}_\mathbf{o}) - U^o(\mathbf{w}_\mathbf{o})$.

We next show that the worst case envy is lower bounded by 0.041, by finding a game instance for which $U^o(\tilde{\mathbf{w}}_\mathbf{o}) - U^o(\mathbf{w}_\mathbf{o}) = 0.041$. Let us consider a scenario with 2 base stations. Let slice o have a share of s_o and one user at each base station with priorities ϕ_1 and ϕ_2. Let the loads of the other slices in these two base stations be $a_1 = 1 - s_o - x\phi_2 s_o$ and $a_2 = x\phi_2 s_o$. for a fixed $x > 0$. Let s_o be sufficiently small such that $a_1 > \phi_2 s_o$.

In this setting, the weights of slice o at each station are given by

$$d_1^1 = \frac{s_o\phi_1\frac{a_1}{a_1+d_1^1}}{\phi_1\frac{a_1}{a_1+d_1^1} + \phi_2\frac{a_2}{a_2+d_2^1}}, \text{ and } d_2^1 = \frac{s_o\phi_2\frac{a_2}{a_2+d_2^1}}{\phi_1\frac{a_1}{a_1+d_1^1} + \phi_2\frac{a_2}{a_2+d_2^1}}$$

We distinguish the cases (i) $x \geq 1$ and (ii) $x < 1$.

(i) For $x \geq 1$, we consider slice o' with share $s_{o'} = s_o$ with priorities $\tilde{\phi}_1$ and $\tilde{\phi}_2$, where

$$\frac{\tilde{\phi}_1}{\tilde{\phi}_2} = \frac{\phi_1}{\phi_2}\frac{\frac{a_2-\phi_2 s_o+d_2^1}{a_2+d_2^1}}{\frac{a_1-\phi_1 s_o+d_1^1}{a_1+d_1^1}}$$

We further consider a third slice with only one user in the first base station with $s_3 = a_1 - \phi_1 s_o$ and a fourth slice with a one user in the second base station with $s_4 = a_2 - \phi_2 s_o$. This leads to $d_1^2 = \phi_1 s_o$ and $d_2^2 = \phi_2 s_o$.

If we now let $s_o \to 0$,

$$d_2^1 = \frac{\phi_2 x\phi_2 s_o}{\phi_1(x\phi_2 s_o + d_2^1) + \phi_2 x\phi_2 s_o}s_o = \phi_2\frac{x\phi_2 s_o}{x\phi_2 s_o + d_2^1}s_o$$

From the above, $d_2^1 = \hat{x}\phi_2 s_o$, where $\hat{x}$ is the unique solution to the equation $x = (x+\hat{x})\hat{x}$. Then, $d_1^1 = s_o - \hat{x}\phi_2 s_o$. From this, we have that in this case

$$\begin{aligned} U^o(\tilde{\mathbf{w}}) - U^o(\mathbf{w}) &= \phi_1\log\left(\frac{\phi_1 s_o}{s_o - \hat{x}\phi_2 s_o}\right) + \phi_2\log\left(\frac{\phi_2 s_o}{\hat{x}\phi_2 s_o}\right) \\ &= \phi_1\log\left(\frac{\phi_1}{1-\hat{x}+\hat{x}\phi_1}\right) - (1-\phi_1)\log(\hat{x}) \end{aligned}$$

(ii) In case that $x < 1$, we consider slice o' has priorities $\tilde{\phi}_1$ and $\tilde{\phi}_2$, where

$$\frac{\tilde{\phi}_1}{\tilde{\phi}_2} = \frac{\phi_1}{\phi_2}\frac{\frac{a_2-x\phi_2 s_o+w_2}{a_2+w_2}}{\frac{a_1-s_o-x\phi_2 s_o+w_1}{a_1+w_1}}$$

which leads to $\tilde{w}_1 = (1-x\phi_2)s_o$ and $\tilde{w}_2 = x\phi_2 s_o$. We further consider a third slice in the first base station with $s_3 = a_1 - (1-x\phi_2)s_o$. If we now let $s_o \to 0$, we have the same expressions as above for w_1 and w_2, from which

$$\begin{aligned} U^o(\tilde{\mathbf{w}}_o) - U^o(\mathbf{w}_o) &= \phi_1\log\left(\frac{s_o - x\phi_2 s_o}{s_o - \hat{x}\phi_2 s_o}\right) + \phi_2\log\left(\frac{x\phi_2 s_o}{\hat{x}\phi_2 s_o}\right) \\ &= \phi_1\log\left(\frac{1-x+x\phi_1}{1-\hat{x}+\hat{x}\phi_1}\right) - (1-\phi_1)\log\left(\frac{x}{\hat{x}}\right) \end{aligned}$$

By putting together the cases $x \geq 1$ and $x < 1$, we can obtain a lower bound for the worst-case envy by finding the values of x and ϕ_1 over $x > 0$ and $\phi_1 \in [0,1]$ that minimize the following expression

$$\begin{cases} \phi_1\log\left(\frac{\phi_1}{1-\hat{x}+\hat{x}\phi_1}\right) - (1-\phi_1)\log(\hat{x}), & \text{for } x \geq 1 \\ \phi_1\log\left(\frac{1-x+x\phi_1}{1-\hat{x}+\hat{x}\phi_1}\right) - (1-\phi_1)\log\left(\frac{x}{\hat{x}}\right), & \text{for } x < 1 \end{cases}$$

By performing the above search numerically, we find a scenario with the following envy level:

$$U^o(\tilde{\mathbf{w}}_o) - U^o(\mathbf{w}_o) = 0.041$$

which terminates the proof of the theorem. ■

References

[1] NGMN Alliance, "Description of Network Slicing Concept," NGMN 5G P1, Jan. 2016.

[2] 3GPP, "Study on Architecture for Next Generation System," TR 23.799, v0.5.0, May 2016.

[3] ——, "Study on Radio Access Network (RAN) sharing enhancements," TS 22.101, v15.1.0, Jun. 2017.

[4] P. Caballero, A. Banchs, G. de Veciana, X. Costa-Pérez, and A. Azcorra, "Network Slicing for Guaranteed Rate Services: Admission Control and Resource Allocation Games," *IEEE Trans. Wireless Commun.*, vol. 17, no. 10, pp. 6419–6432, Oct. 2018.

[5] L. Zhang, "Proportional response dynamics in the Fisher market," *Theoretical Computer Science*, vol. 412, no. 24, pp. 2691–2698, May 2011.

[6] M. Feldman, K. Lai, and L. Zhang, "The Proportional-Share Allocation Market for Computational Resources," *IEEE Trans. Parallel Distrib. Syst.*, vol. 20, no. 8, pp. 1075–1088, Aug. 2009.

[7] F. P. Kelly, A. K. Maulloo, and D. Tan, "Rate control for communication networks: Shadow prices, proportional fairness, and stability," *Operational Research*, vol. 49, pp. 237–252, Mar. 1998.

[8] R. Johari and J. N. Tsitsiklis, “Efficiency Loss in a Network Resource Allocation Game,” *Mathematics of Operations Research*, vol. 29, no. 3, pp. 407–435, Aug. 2004.
[9] S. Yang and B. Hajek, “VCG-Kelly Mechanisms for Allocation of Divisible Goods: Adapting VCG Mechanisms to One-Dimensional Signals,” *IEEE J. Sel. Areas Commun.*, vol. 25, no. 6, pp. 1237–1243, Aug. 2007.
[10] R. Johari and J. N. Tsitsiklis, “Efficiency of scalar-parameterized mechanisms,” *Operations Research*, vol. 57, no. 4, pp. 823–839, Aug. 2009.
[11] S. Shenker, “Fundamental Design Issues for the Future Internet,” *IEEE J. Sel. Areas Commun.*, vol. 13, no. 7, pp. 1176–1188, Sep. 2006.
[12] E. Aryafar, A. Keshavarz-Haddard, M. Wang, and M. Chiang, “RAT selection games in HetNets,” in *Proc. IEEE INFOCOM*, Turin, Italy, Apr. 2013, pp. 998–1006.
[13] G. Scutari, F. Facchinei, J.-S. Pang, and D. P. Palomar, “Real and Complex Monotone Communication Games,” *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 4197–4231, Jul. 2014.
[14] Y. Yang, F. Rubio, G. Scutari, and D. P. Palomar, “Multi-Portfolio Optimization: A Potential Game Approach,” *IEEE Trans. Signal Process.*, vol. 61, no. 22, pp. 5590–5602, Nov. 2013.
[15] S. Lasaulce, M. Debbah, and E. Altman, “Methodologies for analyzing equilibria in wireless games,” *IEEE Signal Process. Mag.*, vol. 26, no. 5, pp. 41–52, Sep. 2009.
[16] M. Chen *et al.*, “Caching in the Sky: Proactive Deployment of Cache-Enabled Unmanned Aerial Vehicles for Optimized Quality-of-Experience,” *IEEE J. Sel. Areas Commun.*, vol. 35, no. 5, pp. 1046–1061, May 2017.
[17] M. Chen, W. Saad, C. Yin, and M. Debbah, “Echo State Networks for Proactive Caching in Cloud-Based Radio Access Networks with Mobile Users,” *IEEE Trans. Wireless Commun.*, vol. 16, no. 6, pp. 3520–3535, Jun. 2017.
[18] M. Bennis, S. Perlaza, and M. Debbah, “Learning Coarse-Correlated Equilibria in Two-Tier Networks,” in *Proc. IEEE ICC*, Ottawa, ON, Canada, Jun. 2012, pp. 1592–1596.
[19] M. Bennis and S. Perlaza, “Decentralized Cross-Tier Interference Mitigation in Cognitive Femtocell Networks,” in *Proc. IEEE ICC*, Kyoto, Japan, Jun. 2011.
[20] M. Bennis, S. M. Perlaza, P. Blasco, Z. Han, and H. V. Poor, “Self-Organization in Small Cell Networks: A Reinforcement Learning Approach,” *IEEE Trans. Wireless Commun.*, vol. 12, no. 7, pp. 3202–3212, Jul. 2013.
[21] P. Di Francesco, F. Malandrino, T. K. Forde, and L. A. DaSilva, “A Sharing- and Competition-Aware Framework for Cellular Network Evolution Planning,” *IEEE Trans. on Cogn. Commun. Netw.*, vol. 1, no. 2, pp. 230–243, Jun. 2015.
[22] B. Leng, P. Mansourifard, and B. Krishnamachari, “Microeconomic Analysis of Base-station Sharing in Green Cellular Networks,” in *Proc. of IEEE INFOCOM*, Toronto, ON, Canada, Apr. 2014, pp. 1132–1140.
[23] A. Gudipati, L. E. Li, and S. Katti, “RadioVisor: A Slicing Plane for Radio Access Networks,” in *Proc. HotSDN*, Chicago, IL, USA, Aug. 2014, pp. 237–238.
[24] I. Malanchini, S. Valentin, and O. Aydin, “Generalized resource sharing for multiple operators in cellular wireless networks,” in *Proc. IWCMC*, Nicosia, Cyprus, Aug. 2014, pp. 803–808.
[25] R. Mahindra, M. Khojastepour, H. Zhang, and S. Rangarajan, “Radio Access Network sharing in cellular networks,” in *Proc. IEEE ICNP*, Gottingen, Germany, Oct. 2013.
[26] S. Rathinakumar and M. K. Marina, “GAVEL: strategy-proof ascending bid auction for dynamic licensed shared access,” in *Proc. ACM MobiHoc*, Paderborn, Germany, Jul. 2016, pp. 121–130.
[27] NGMN Alliance, “5G White Paper,” NGMN deliverable, Feb. 2015.
[28] X. Zhou, R. Li, T. Chen, and H. Zhang, “Network slicing as a service: enabling enterprises’ own software-defined cellular networks,” *IEEE Commun. Mag.*, vol. 54, no. 7, pp. 146–153, Jul. 2016.
[29] K. Samdanis, X. Costa-Perez, and V. Sciancalepore, “From Network Sharing to Multi-tenancy: The 5G Network Slice Broker,” *IEEE Commun. Mag.*, vol. 54, no. 7, pp. 32–39, Jul. 2016.
[30] The 5G Infrastructure Public Private Partnership (5GPPP), “5G Empowering vertical industries,” 5GPPP White Paper, Feb. 2016.
[31] J. Mo and J. Walrand, “Fair end-to-end window-based congestion control,” *IEEE/ACM Trans. Netw.*, vol. 8, no. 5, pp. 556–567, Oct. 2000.
[32] X. Costa-Perez, J. Swetina, T. Guo, R. Mahindra, and S. Rangarajan, “Radio access network virtualization for future mobile carrier networks,” *IEEE Commun. Mag.*, vol. 51, no. 7, pp. 27–35, Jul. 2013.
[33] L. Buttyan and J.-P. Hubaux, *Security and Cooperation in Wireless Networks*. Cambridge University Press, 2008.
[34] L. Li, M. Pal, and Y. R. Yang, “Proportional fairness in multi-rate wireless LANs,” in *Proc. IEEE INFOCOM*, Phoenix, AZ, USA, Apr. 2008, pp. 1004–1012.
[35] ITU-R, “Report ITU-R M.2135-1, Guidelines for evaluation of radio interface technologies for IMT-Advanced,” Technical Report, Dec. 2009.
[36] Q. Ye *et al.*, “User Association for Load Balancing in Heterogeneous Cellular Networks,” *IEEE Trans. Wireless Commun.*, vol. 12, no. 6, pp. 2706–2716, Jun. 2013.
[37] H. Dhillon, H. S. Dhillon, R. K. Ganti, F. Baccelli, J. G. Andrews , “Modeling and Analysis of K-Tier Downlink Heterogeneous Cellular Networks,” *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 550–560, Apr. 2012.
[38] V. Sciancalepore, V. Mancuso, A. Banchs, S. Zaks, A. Capone , “Interference coordination strategies for content update dissemination in LTE-A,” in *Proc. IEEE INFOCOM*, Toronto, ON, Canada, Apr. 2014, pp. 1797–1805.
[39] 3GPP, “Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures,” TS 36.213, v12.5.0, Mar. 2015.
[40] S. Kullback and R. A. Leibler, “On Information and Sufficiency,” *Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, Mar. 1951.

Pablo Caballero received his B.S. in telecommunications and his M.S. in telematics engineering respectively from the University Carlos III of Madrid in 2013 and 2015, respectively. In 2015, he joined the Wireless Networking and Communications Group at the University of Texas at Austin, where he recently got his Ph.D. degree. Previously, Pablo worked as Research Assistant at IMDEA Networks Institute and as Research Intern at NEC Laboratories Europe. His research interests lie in the design and performance evaluation of communication networks, game theory and algorithm analysis.

Albert Banchs (M’04-SM’12) received his M.Sc. and Ph.D. degrees from the Polytechnic University of Catalonia (UPC-BarcelonaTech) in 1997 and 2002, respectively. He is currently a Full Professor with the University Carlos III of Madrid (UC3M), and has a double affiliation as Deputy Director of the IMDEA Networks institute. Before joining UC3M, he was at ICSI Berkeley in 1997, at Telefonica I+D in 1998, and at NEC Europe Ltd. from 1998 to 2003. He is Editor of IEEE Transactions on Wireless Communications and IEEE/ACM Transactions on Networking. His research interests include the performance evaluation and algorithm design in wireless and wired networks.

Gustavo de Veciana (S’88-M’94-SM’01-F’09) received his B.S., M.S, and Ph.D. in electrical engineering from the University of California at Berkeley in 1987, 1990, and 1993 respectively, and joined the Department of Electrical and Computer Engineering where he is currently a Cullen Trust Professor of Engineering. His research focuses on communication and computing networks; data-driven decision-making in man-machine systems, and applied probability and queueing theory. Dr. de Veciana served as editor and is currently serving as editor-at-large for the IEEE/ACM Transactions on Networking. In 2009 he was designated IEEE Fellow for his contributions to the analysis and design of communication networks. He currently serves on the board of trustees of IMDEA Networks.

Xavier Costa-Pérez (M’06-SM’18) is Head of 5G Networks R&D and Deputy General Manager of the Security & Networking Research Division at NEC Laboratories Europe. His team contributes to products roadmap evolution, EU H2020 collaborative projects and related standardization bodies, and has received several R&D Awards for successful technology transfers. Xavier is a 5GPPP Technology Board member and the Technical Manager of the 5G-Transformer project, has served on the TPC of several conferences and holds several patents. He received his M.Sc. and Ph.D. degrees from the Polytechnic University of Catalonia (UPC) and was the recipient of a national award for his Ph.D. thesis.