Incentivizing Efficient Load Repartition in Heterogeneous Wireless Networks with Selfish Delay-Sensitive Users

Vladimir Fux Patrick Maillé

Institut Mines-Telecom; Telecom Bretagne
2, rue de la Chataigneraie CS 17607
35576 Cesson-Sevigne, FRANCE
e-mail: {first.last}@telecom-bretagne.eu

***Abstract*—Almost all modern mobile devices are equipped with a number of various wireless interfaces simultaneously, so that each user is free to select between several types of wireless networks. This opportunity raises a number of challenges, since in general selfish choices do not lead to a globally efficient repartition of users over networks. The most popular approach in this context is to charge an extra tax for connecting to overloaded networks, thus incentivizing users to choose less congested alternatives.**

In this paper we apply that idea to a system where several networks with a common coverage area coexist. Moreover we assume that users –or the applications they use– are heterogeneous in their sensitivity to the congestion-varying Quality-of-Service (QoS). We show the technical and computational feasibility of computing taxes leading to a globally optimal outcome for any number of networks and application types (QoS-sensitivities), hence generalizing the results from previous works.

I. Introduction

Wireless networks technologies such as 3G, WiFi (IEEE 802.11 a/b/g/n/ac), or LTE, are becoming more and more crucial and widespread. Each technology has its own advantages and drawbacks, in terms of throughput, geographic area covered, energy consumption, etc. Moreover, recent mobile terminals are equipped with a number of different network interfaces, offering the possibility to connect through different technologies to a variety of networks concurrently. Wireless network users can then switch from one network to another, for example using the IEEE 802.21 standard [7].

Switching between networks implementing different technologies is referred to as vertical handover. We expect that one of the major objectives in future generations of mobile networks would be to find a solution for the vertical handover decision, satisfying both mobile users and providers. Indeed, allowing each user to select at any time its most suitable wireless network, i.e., to be *always-best-connected* [10], could cause the overload of some technologies and the under-utilization of others. This is due to user selfishness: users ignore the negative consequences of their actions on others when making their choices, which can lead to inefficient situations. In order to cope with that problem and profit from the diversity of technologies, operators have to improve resource management.

A number of recent papers in the transportation science literature addressed that same problem (see [4], [8], [12]). They discuss the introduction of incentive tools, interpreted as taxes, which could influence user choices towards a more efficient situation. In this paper, we focus on applying that idea to a situation when we need to influence user's choices between several wireless heterogeneous networks. Due to the specificity of the wireless framework, our problem can be modeled as a routing game simpler than the general ones studied in [4], [8], [12], which allows us to reach analytical results.

II. Related Work

Various works in the literature investigate how the selfish behavior of users in networks can be regulated through incentive tools, such as taxation or penalties. The idea being that users select the cheapest path from their position to their destination node in the network, taking into account the cost (latency, or delay, that is sensitive to congestion) of the paths but also possibly some additional (monetary) costs imposed by the network manager. So that a proper definition of the tax levels influences user choices. In the homogeneous case, i.e., when all users have the same sensitivity to the taxation, Beckmann et al. [2] showed that the so-called Pigovian taxes–applied on each link, and computed using the derivative of the cost functions of the links–produce a minimum-latency (delay) traffic routing (see [18]). In [11], Pigovian taxes are also used to influence user preferences, and induce a repartition of flows among the available access networks that optimizes the overall network performance.

Reference [4] considers the case when users may perceive differently the relative costs of delay and taxes. The authors were the first to study this setting, for a situation when all users have the same source and destination, with any network topology in between. For that scenario, it is shown that there exist taxes so that the resulting user flow minimizes the average latency. Those results have been generalized to the

ISBN 978-3-901882-53-1, 9th CNSM 2013: Workshop ICQT 2013

multicommodity setting (i.e., several source-destination pairs) in [12], [13]. A constructive proof is given to show that taxes inducing the minimum average latency multicommodity flow exist for both the cases of elastic (i.e., cost-dependent) and nonelastic demands.

In the articles evoked above, users are sensitive to the latency caused by congestion; however there are several papers where other congestion-dependent costs are considered. In [3], three different cost functions are proposed: two of them depend both on the interference level and the transmission rate, and the third one depends only on the interference level. In [17] users are supposed to have information about the geographical locations and current loads of network access points, and are able to move between the coverage areas of different networks. Thus users face a trade-off between the load level of their current access point and the distance they have to travel. In both works, the authors take into account only the user behavior, i.e., mobile users select the access network selfishly, hence a noncooperative game. The interaction between mobile users and the operator is not considered there, while in the current paper we consider the impact of the operator's actions (the incentives).

A totally different approach is to seek for an optimal *user admission* policy in a network, through SMDPs (Semi Markov Decision Processes). This approach is applied to the problem of global expected throughput maximization with the help of a central controller (taking admission decisions) in [5], [6], [14]. The methods consider that the user arrival process and the time spent in the network are known stochastically. An admission policy maximizing total throughput can then be derived. However, the presence of an authority making decisions instead of users could be perceived negatively. In our model, users make their own choices, the operator's intervention consisting only in adding incentives.

III. Model and Problem Formulation

We consider a system with n heterogeneous wireless networks covering the same area. This model is a generalization of the one in [9] where only two networks and two user (or application) classes were considered. The users situated in the common coverage area of these networks seek for an Internet connection. We assume that they could easily handover from one network to another, thus choosing at every moment the most suitable one. Users select which network to connect to based on the QoS they experience and on the prices charged by the network owner. We investigate how users make their decisions, what is the outcome of these decisions, how far that outcome is from the optimum situation from the point of view of the network owner, and, finally, how the network owner could stimulate users to act efficiently.

A typical application case of our approach is that of network off-loading, with the objective to reach the most efficient load balance between indoor and outdoor coverage technologies.

A. Mathematical formulation

We identify all parameters related to a specific network i through the use of the lower index $_i$, for $1 \le i \le n$. Each network i has a QoS-related cost function $\ell_i(f_i)$ that we will call the latency function, where f_i is the flow (cumulated throughput) on network i. All networks are owned by the same provider, which is aiming to minimize some cost function and could influence users behavior through charging a tax τ_i on each network i.

We assume a total user demand D coming from users' applications. Since QoS requirements can vary depending on the applications used and on user preferences, the trade-offs between QoS and monetary cost shall differ, which we model through the sensitivity to the monetary cost (or equivalently, the ratio of the price sensitivity to the latency sensitivity). We can represent this variability by considering price sensitivities of *users* and price sensitivities of *applications*, so that each pair (*user,application*) would lead to a specific sensitivity value. Assuming a finite number of application types and of user types, we would have a finite number of overall sensitivities. To simplify notations, without loss of generality we will treat a user running q applications with different requirements as q separate users, each one running one application. Therefore from now on we only evoke users, each one having a given price sensitivity. This simplification can be done because the interactions among flows from a single user are negligible due to a non-atomicity assumption explained below: no user can improve his utility by coordinating his own flows, so we can treat those flows as being issued by distinguished (non-cooperating) users.

We consider m classes of users, implying that users from the same class have the same price sensitivity value. We write all the parameters related to class j with the upper index ^{j}, $1 \le j \le m$; users in class j have tax sensitivity $\alpha^j \ge 0$ and the total demand from class-j users is denoted by d^j, so that $\sum_{j=1}^m d^j = D$.

We assume that the cost perceived by a class-j user connected to network i is a combination of QoS (through the latency function) and price

$$C_i^j(f) = \ell_i(f_i) + \alpha^j \tau_i, \tag{1}$$

and that every user seeks for a connection which minimizes this cost. The following assumption specifies the type of latency functions we use in our model:

Assumption A: Each network i has a capacity c_i, and a load-sensitive latency function corresponding to the mean sojourn time in an M/M/1 queue:

$$\ell_i(f_i) = \begin{cases} (c_i - f_i)^{-1} & \text{if } f_i < c_i, \\ \infty & \text{if } f_i \ge c_i, \end{cases} \tag{2}$$

with f_i the total flow on network i.

With this type of latency function we also have to assume that $D < \sum_{i=1}^n c_i$, in other words the aggregated capacity is enough to treat all demand. We assume that the provider owning all considered networks is interested in minimizing the social cost (or total cost) expressed as:

$$C(f) = \sum_{i=1}^{n} f_i \ell_i(f_i), \tag{3}$$

where $f = (f_1, \ldots, f_n)$ is the flow distribution vector, with $\sum_{i=1}^{n} f_i = D$. That cost corresponds to the aggregated latencies undergone by users.

B. Routing game interpretation

Assuming that only radio links incur QoS-related costs (i.e., latency), the setting described above could be seen as a routing problem, with a common source for all users, represented by the common coverage area of the networks, and one common destination (the Internet). Each user forwards his flow through one of n routes, which are the n networks, with a routing cost equal to the cost in (1), as depicted in Figure 1. When users selected their route, their interactions form a *noncooperative routing game*.

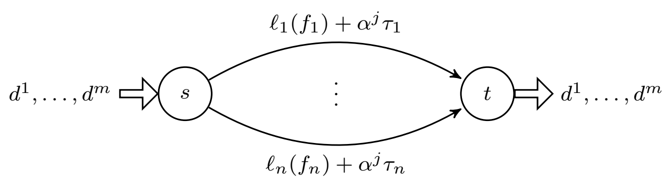

Fig. 1. Logic representation of the network selection problem as a routing problem: the perceived cost on each route i depends on the load f_i and the tax τ_i, but also on the user type j through the sensitivity α^j.

In this paper we assume that users are *non-atomic* [1], i.e., the individual impact of each player on the network loads is negligible. Those games have been extensively studied since the seminal work of Wardrop [19]. In particular, for our routing game there are theoretical results proving the existence of optimal taxes, i.e., taxes driving the system to a situation with minimum social cost [12]. For the specific latency functions considered in this paper, we find an analytical expression for those optimal taxes.

Note that our model does not include network attachment costs: adding such costs (possibly different among networks) would affect the attachment decisions of users but also possibly deter them from performing vertical handovers during the connection (due to varying network conditions over time). Such considerations are left for future work.

C. The case of several providers

In this paper we consider that all networks are owned and controlled by the same entity, that we call the provider. The objective for the provider here is to make the best use of the network resource, in the sense of the aggregated user cost of Equation (3). Hence the provider is not directly driven by revenue, the taxes imposed on network are only used as incentives to reach the best flow repartition.

Considering several providers managing the different networks would totally change the paradigm, since those providers would compete to attract customers and make revenue, and would use taxes for that purpose. We would then have a non-cooperative game played among providers deciding their tax levels, and anticipating user reactions when making those decisions. Such situations of competing providers have been studied in [15] with cost functions similar to ours, but with few positive analytical results: even the existence of a Nash equilibrium of the tax-setting game is not guaranteed. However, if such an equilibrium exists, it can reasonably be expected to benefit to users (a general property of competition) with respect to a case where a single entity controls all networks and sets prices to maximize revenue (not the case treated here).

The case when several providers perfectly cooperate to optimize network usage would be equivalent to the one-provider case. However there are some in-between situations, where providers may partially compete and cooperate: for example they may have roaming agreements, or may have to share the capacity of their access networks. Those aspects are partially treated in [16] but would deserve more attention.

IV. User equilibrium and optimal situations

In this section we define the user equilibrium of the routing game, and compare the equilibrium without taxes to an optimal situation from the point of view of social cost (3). To simplify notations, we assume without loss of generality that:

Assumption B:

1) $c_1 \geq c_2 \geq \ldots \geq c_n$
2) $\alpha^1 < \alpha^2 < \ldots < \alpha^m$

A. User equilibrium

When users act selfishly, each one choosing a network minimizing his individual cost (1), then the game has an equilibrium, i.e., a situation such that no user can reduce his cost by a route change. We call that situation *user equilibrium* or *Wardrop equilibrium*, and it is characterized by Wardrop's principle [19].

Definition 1: A Wardrop equilibrium is a flow repartition $f = (f_i^j)_{1\leq i\leq n, 1\leq j\leq m}$, such that $\begin{cases} f_i^j \geq 0 & \forall i,j \\ d^j = \sum_{i=1}^{n} f_i^j & \forall j \end{cases}$ and such that

$$\forall i, i', j \qquad f_i^j > 0 \Rightarrow \ell_i(f_i) + \alpha^j \tau_i \leq \ell_{i'}(f_{i'}) + \alpha^j \tau_{i'}, \tag{4}$$

with $f_i = \sum_{j=1}^{m} f_i^j$. The quantity f_i^j represents the flow from class-j users that is routed through network i (recall that d^j is the total flow of class-j users).

In other words, at a Wardrop equilibrium, the cost of each used route is lower (for the users taking that route) than the cost of any other.

B. User equilibrium without taxes

Consider the case when the provider does not charge taxes for using his networks (or equivalently all taxes are the same), and thus users make their choices without any intervention from the provider. Then the flows at a Wardrop equilibrium have the form stated in the following proposition.

Proposition 1: Under Assumptions A and B, at a Wardrop equilibrium f^{WE} with no taxes being applied, we have:

$$f_i^{\mathrm{WE}} = \begin{cases} \frac{D-\sum_{q=1}^{t} c_q + tc_i}{t} & \text{if } i \le t, \\ 0 & \text{otherwise,} \end{cases} \tag{5}$$

where $1 \le t \le n$ is the maximum index for which

$$D - \sum_{i=1}^{t} c_i + tc_t > 0, \tag{6}$$

and represents the number of used networks.

The proof comes quite directly from Definition 1, since without taxes all users should perceive the same cost on all used routes. The proof details are omitted due to lack of space.

Proposition 1 provides a way to compute the equilibrium flows (in a time linear in the number n of flows).

C. Optimal situation

In this section we investigate the optimum situation, which we later intend to reach by introducing appropriate taxes. An optimal flow assignment $f^{\mathrm{opt}} = (f_1^{\mathrm{opt}}, \ldots, f_n^{\mathrm{opt}})$ which minimizes social cost (3) is the solution of the following mathematical program:

$$\min_{f_1,\ldots,f_n} \sum_{i=1}^{n} f_i \ell_i(f_i) \tag{7}$$

$$\text{s.t.} \begin{cases} \sum_{i=1}^{n} f_i = D \\ f_i \ge 0, \text{ for } i = 1, \ldots, n \end{cases} \tag{8}$$

Note that this problem does not distinguish among user classes, it only involves aggregate flows on each network. With the specific latency functions (2) we can express the optimal flows analytically.

Proposition 2: Optimal flows $(f_i^{\mathrm{opt}})_{1\le i\le n}$ minimizing (3) are unique and given by:

$$f_i^{\mathrm{opt}} = \begin{cases} c_i - \frac{\sqrt{c_i}(\sum_{j=1}^{k} c_j - D)}{\sum_{j=1}^{k}\sqrt{c_j}} & \text{if } i \le k, \\ 0 & \text{otherwise,} \end{cases} \tag{9}$$

where $1 \le k \le n$ is the maximum index for which

$$c_i - \frac{\sqrt{c_i}(\sum_{j=1}^{k} c_j - D)}{\sum_{j=1}^{k}\sqrt{c_j}} \ge 0. \tag{10}$$

Proof: We apply the following result from [2]:

Lemma 1 (Beckmann et al., 1956): For any non-atomic routing game with latency functions (ℓ_i), the optimal flows minimizing social cost (3) correspond to the Wardrop equilibrium flows of a modified game where latency functions are

$$\bar{\ell}_i(f_i) = \ell_i(f_i) + f_i \ell_i'(f_i). \tag{11}$$

Therefore, applying the equilibrium conditions (4) there exists $H > 0$ such that for all $i, 1 \le i \le n$:

$$\begin{cases} f_i^{\mathrm{opt}} > 0 \Rightarrow \ell_i(f_i^{\mathrm{opt}}) + f_i^{\mathrm{opt}} \ell_i'(f_i^{\mathrm{opt}}) = H, \\ f_i^{\mathrm{opt}} = 0 \Rightarrow \ell_i(f_i^{\mathrm{opt}}) + f_i^{\mathrm{opt}} \ell_i'(f_i^{\mathrm{opt}}) = \ell_i(0) \ge H. \end{cases} \tag{12}$$

With our latency functions (2), we immediately remark that

$$f_i^{\mathrm{opt}} > 0 \Leftrightarrow \frac{1}{c_i} < H, \tag{13}$$

thus from Assumption B there exists k (the number of used networks at the optimal situation) such that ($f_i^{\mathrm{opt}} > 0 \Leftrightarrow i \le k$). From (12) we get

$$f_i^{\mathrm{opt}} = c_i - \frac{\sqrt{c_i}}{\sqrt{H}}, i = 1, \ldots, k, \tag{14}$$

and the condition $\sum_{i=1}^{k} f_i^{\mathrm{opt}} = D$ yields $H = \frac{(\sum_{i=1}^{k}\sqrt{c_i})^2}{(\sum_{i=1}^{k} c_i - D)^2}$. Plugging that last expression into (14) gives (5), while plugging it into (13) leads to the characterization (10) for k. ■

Similarly to Proposition 1 for equilibrium flows, Proposition 2 implicitly defines a linear-time algorithm to compute optimal (i.e., globally cost-minimizing) flows. Note that to compute optimal (as well as equilibrium) flows we only need to know the network capacities $(c_i)_{1\le i\le n}$ and the total demand D, that do not depend on any characteristics of user classes.

V. ELICITING OPTIMAL USER-NETWORK ASSOCIATIONS WITH TAXES

To reduce the total cost the provider has to give an incentive to some users to switch networks, so as to provide higher QoS to the majority of users and lower QoS to some others, instead of providing the same QoS to everyone (what we get at the Wardrop equilibrium without taxes). Here the provider introduces special taxes, such that the flow assignment in the Wardrop equilibrium induced by these taxes is the optimum flow assignment. Previous works (see [4]) ensure that those taxes exist, and the following lemma will help to compute them.

Lemma 2: Under Assumptions A and B, optimal taxes are such that $\tau_1 \ge \tau_2 \ge \ldots \ge \tau_k$, where k is the number of networks used (i.e., networks with positive flows) at the optimal situation. For networks $i > k$, it is sufficient to have $\tau_i \ge \tau_k$.

Proof: Let us first consider used networks, i.e. networks $1, \ldots, k$. From Lemma 1, for $i, i' \le k$ we have

$$\frac{c_i}{(c_i - f_i^{\mathrm{opt}})^2} = \frac{c_{i'}}{(c_{i'} - f_{i'}^{\mathrm{opt}})^2} := K^2 \tag{15}$$

for some constant K.

Suppose that $\tau_i < \tau_{i+1}$ for some $i < k$, and that those taxes lead to an equilibrium coinciding with the optimal situation. Then for a class of users j choosing network $i+1$, we have from the equilibrium conditions

$$\ell_{i+1}(f_{i+1}^{\mathrm{opt}}) + \alpha^j \tau_{i+1} \le \ell_i(f_i^{\mathrm{opt}}) + \alpha^j \tau_i,$$

hence $\ell(f_{i+1}^{\mathrm{opt}}) < \ell(f_i^{\mathrm{opt}})$.

But $\ell_i(f_i^{\text{opt}}) = 1/(c_i - f_i^{\text{opt}}) = K/\sqrt{c_i}$ from (15), therefore since $c_i \geq c_{i+1}$ we have $\ell(f_{i+1}^{\text{opt}}) \geq \ell(f_i^{\text{opt}})$, a contradiction.

Now, we consider networks $k+1, \ldots, n$, which do not carry any flow in the optimal situation: no user should prefer one of those networks to their current one. In particular, denoting by j a class sending flow to network k under optimal taxes, we must have

$$\ell_i(0) + \alpha^j \tau_i \geq \ell_k(f_k^{\text{opt}}) + \alpha^j \tau_k, \ \forall i = k+1, \ldots, n,$$

thus

$$\tau_i \geq \frac{\ell_k(f_k^{\text{opt}}) - \ell_i(0)}{\alpha^j} + \tau_k, \ \forall i = k+1, \ldots, n. \qquad (16)$$

But from (12) we have $\ell_k(f_k^{\text{opt}}) - \ell_i(0) \leq 0$, therefore taking $\tau_i \geq \tau_k$ is sufficient to ensure that (16) holds, i.e., that networks $i = k+1, \ldots, n$ are not chosen by users. ■

Now we provide a method to calculate the optimal taxes:

Proposition 3: Under Assumptions A and B, the following taxes are optimal:

$$\tau_{i+1} = \tau_i + \frac{\ell_i(f_i^{\text{opt}}) - \ell_{i+1}(f_{i+1}^{\text{opt}})}{\alpha^{s_i}}, \qquad (17)$$

for $i = 1, \ldots, n-1$, with τ_1 taken arbitrarily, and with

$$s_i := \min \left\{ j : \sum_{r=1}^{i} f_r^{\text{opt}} \leq \sum_{q=1}^{j} d^q \right\}. \qquad (18)$$

For networks used at the optimal situation (networks with $f_i^{\text{opt}} > 0$), the index s_i represents the class with maximum sensitivity among those sending flow to network i.

Proof: For a network i with positive optimal flow, we define $\alpha_i^{\max}$ and $\alpha_i^{\min}$ as respectively the maximum and minimum sensitivities among classes sending some flow to network i (i.e., classes j such that $f_i^j > 0$). Then the Wardrop equilibrium conditions for classes choosing networks i and $i+1$ (both with positive optimal flows) yield

$$\alpha_i^{\max}(\tau_i - \tau_{i+1}) \leq \ell_{i+1}(f_{i+1}^{\text{opt}}) - \ell_i(f_i^{\text{opt}}) \leq \alpha_{i+1}^{\min}(\tau_i - \tau_{i+1})$$

Since $\tau_i \geq \tau_{i+1}$ from Lemma 2, we obtain $\alpha_i^{\max} \leq \alpha_{i+1}^{\min}$.

• If $\alpha_i^{\max} = \alpha_{i+1}^{\min}$ then a class of users, denoted by j', is indifferent between both networks. From the Wardrop equilibrium conditions we have:

$$\ell_i(f_i) + \alpha^{j'} \tau_i = \ell_{i+1}(f_{i+1}) + \alpha^{j'} \tau_{i+1}. \qquad (19)$$

From this we derive (17), with j' satisfying (18).

• If $\alpha_i^{\max} < \alpha_{i+1}^{\min}$, then this corresponds to a rare case, when two neighbor classes are perfectly divided, and there is no class whose users are indifferent between both networks. One more time using the Wardrop equilibrium conditions we write:

$$\begin{cases} \ell_i(f_i) + \alpha_i^{\max} \tau_i \leq \ell_{i+1}(f_{i+1}) + \alpha_i^{\max} \tau_{i+1} \\ \ell_i(f_i) + \alpha_{i+1}^{\min} \tau_i \geq \ell_{i+1}(f_{i+1}) + \alpha_{i+1}^{\min} \tau_{i+1}. \end{cases} \qquad (20)$$

These two inequalities imply that

$$\tau_i + \frac{\ell_i(f_i) - \ell_{i+1}(f_{i+1})}{\alpha_i^{\max}} \leq \tau_{i+1} \leq \tau_i + \frac{\ell_i(f_i) - \ell_{i+1}(f_{i+1})}{\alpha_{i+1}^{\min}}.$$

So, in this particular case a whole range of taxes for network $i+1$ induce an optimal division of users. Note that our proposition in Equation (17) falls in that range.

For networks with empty flows in the optimal situation, our proposition is still valid. Indeed, since taxes decrease with the network index, the class m with the highest sensitivity to price is the first class which would be interested in connecting to these empty networks. It is easy to see that the taxes defined by (17) will prevent them from doing this. If k is the maximum index of a network with non-empty flow in optimal situation, then from the Wardrop equilibrium conditions we should have:

$$\ell_k(f_k^{\text{opt}}) + \alpha^m \tau_k \leq \ell_i(0) + \alpha^m \tau_i \qquad \forall i > k, \qquad (21)$$

which is verified with the tax defined by (17). ■

Like the two previous propositions in the paper, Proposition 3 implicitly defines an algorithm to compute optimal taxes: Proposition 2 should first be applied to obtain optimal flows, then (18) provides the value of s_i for each network i to be inserted into (17) so as to get the tax value.

The freedom to arbitrary choose τ_1 gives us an interesting feature: the provider could regulate his total revenue by adjusting appropriately τ_1 without any harm to the social cost. For example, τ_1 could be set (to a negative value) such that the total revenue is null.

The intuition behind Proposition 3 is illustrated in Figure 2. We already know from Lemma 2 that the bigger tax should be

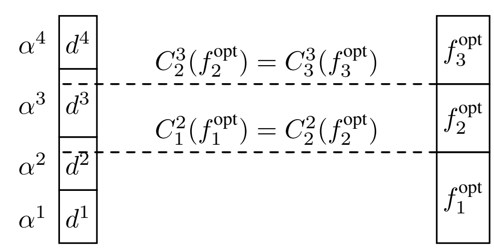

Fig. 2. Example of user distribution among networks with optimal taxes for the case $m = 4, n = 3$: class-1 (resp. class-4) users all attach to network 1 (resp. 3), while class-2 (resp. class-3) users are split among networks 1 and 2 (resp. 2 and 3).

charged on networks with lower indexes (bigger capacities). This in turn means that the "richest" users are connected to them (the smaller their sensitivity values). Thus, the least price-sensitive users will choose network 1. On the example on Figure 2, the total flow of class-1 users is not enough to ensure an optimal flow f_1^{opt} in network 1. So, the following (by sensitivity value) class should fulfill the optimal flow in network 1. The total flow of classes 1 and 2 is bigger than the optimal flow f_1^{opt}, so we have to split users from class 2. Here we should use the Wardrop equilibrium conditions to find an expression for τ_2 depending on τ_1, this condition meaning that users of class 2 are indifferent between networks 1 and 2.

In general, the only computational difficulty is to find a class with users indifferent between two networks with consecutive indices. In the proposed example, it is class 2 for networks 1 and 2, and class 3 for networks 2 and 3.

VI. Efficiency analysis

In this section we present some analytical investigations about the efficiency of our taxation method. As an efficiency measure we use the Price of Anarchy (PoA), which is the ratio between the total cost value achieved from the selfish users behavior and the minimum total cost value that could be reached by coordinating users [12]. This value is larger or equal to one. The larger the PoA, the less efficient the selfish users behavior, while if the PoA equals one, then selfish user behavior leads to an optimal situation and no intervention is needed. Recall that the taxes computed in Proposition 3 drive the system to an optimal situation, i.e., to a situation with PoA equal to one.

A. Influence of heterogeneity on the PoA

At first, we provide the PoA values while varying the heterogeneity among networks, which comes from the different capacities. For simplicity, we consider capacities of the form $c_i = c_0 w^{i-1}$ for $i = 1, \ldots, n$, where we call $w \in (0, 1]$ the homogeneity value. On Figure 3 we plot the PoA for different values of the total user demand D, with c_0 such that the total capacity of the system equals 10 [Mbit/s]. We observe more heterogeneous systems lead to a larger worst-case PoA (higher inefficiency due to user selfishness). It is especially clear when total demand is close to the total capacity value (i.e, the system is congested), but for very heterogeneous systems the PoA is quite high even for small demand values, thus the introduction of taxes would lead to significant performance gains.

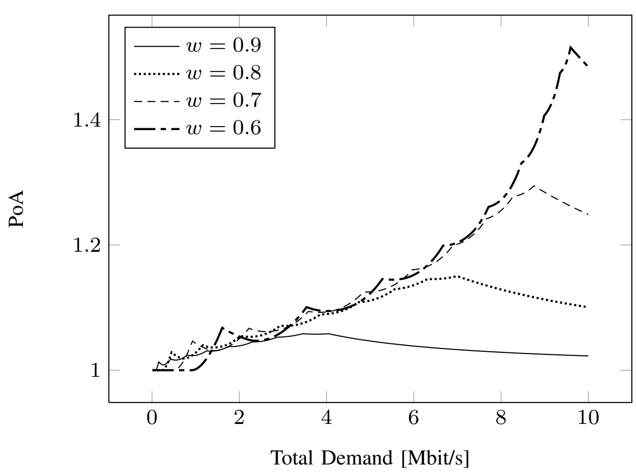

Fig. 3. PoA versus total demand D with $n = 10$ and total capacity equal to 10 [Mbit/s].

B. The PoA interpretation

Finally, we present two counterparts for the Price of Anarchy in our model. For simplicity, we consider only a case with two networks in which $c_1 = 4$ [Mbit/s] and $c_2 = 11$ [Mbit/s]. First, Figure 4 shows how many more users the operator could serve if using network resources in an optimal way for the same total cost, compared to the case when he does not influence users behavior. In a somehow similar way, Figure 5 indicates the capacity (or investment) reduction that would lead to an unchanged total cost, just because of effective resource management. These two values are comparable to the Price of Anarchy, but have the advantage of being convertible into monetary gains, probably more appealing to network providers. These figures have to be understood as follows. Consider a system with relative load equal to 0.7 (dotted curve) and a PoA of 1.02: Figure 4 show that if we optimize resource usage (e.g., through optimal taxes), we could have 2% more users in our system without increasing the total cost. The analogical explanation works for Figure 5: in the same situation, if we introduce optimal taxes, we can decrease our system's capacity by 2% without changing the overall cost perceived by users.

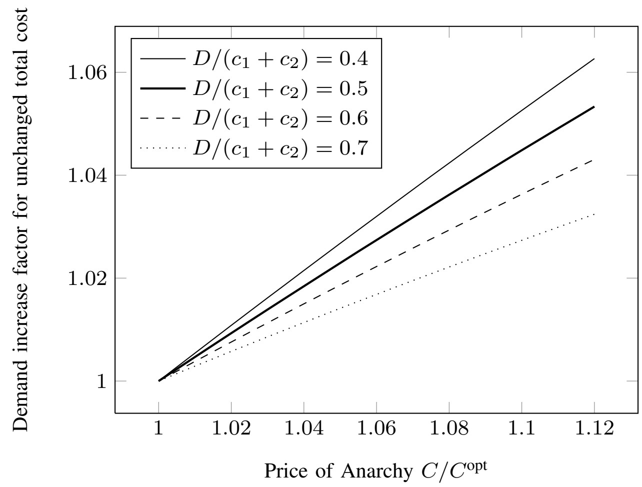

Fig. 4. Demand gain versus PoA, for different demand levels in the case of two networks.

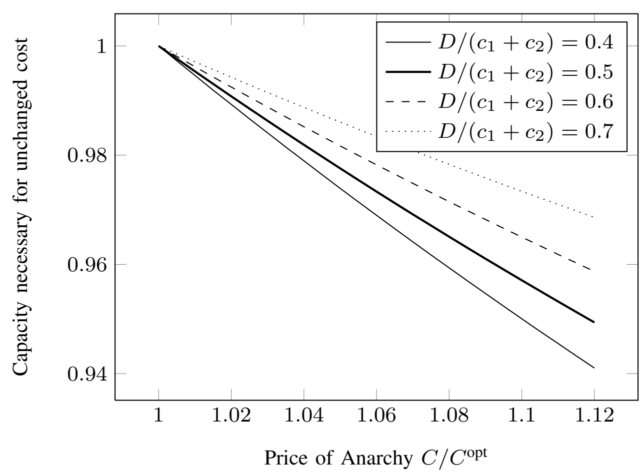

Fig. 5. Capacity gain versus PoA, for different demand levels in the case of two networks.

VII. Conclusions and perspectives

In this paper we have considered the inefficiency of selfish user behavior in heterogeneous wireless systems. We have generalized the results of [9] to a model with an arbitrary number of user classes (corresponding to user-specific and/or application-specific perceptions of price), and also an arbitrary number of networks. We have derived analytical expressions for the optimal taxes, which drive the system to an optimal flow repartition minimizing the total cost. We have showed that the "cost" of inefficiency can have monetary equivalents.

Our model relies on some strong assumptions, one of which is the simple network topology–all networks being supposed to have the same coverage area. Obviously, this topology is quite far from reality, and in the future we aim to consider more complicated systems. Further, we would like to study other–possibly application-specific–cost functions.

Additionally, the non-atomicity assumption significantly simplifies the analysis, however its validity becomes questionable if we consider small-cell networks with only a few users and bandwidth-consuming applications. Extending our work to the atomic case would thus be of high interest; in such a case the decisions made by users could involve attaching simultaneously to several networks and splitting the flows among them (benefiting from protocols such as MultiPath TCP).

Finally, our work did not consider the practical implementation aspects of our mechanism. Those of course need to be examined for our mechanism to be applicable. In particular, measuring precisely the congestion level at the access point, and transmitting this information to users so that they make their decisions, warrants specific investigations. Among the possible tools that can be used for the latter task, one can evoke the 802.21 standard [7] and the Generic Access Network techniques for the management of cross-technology handovers and the information diffusion to users. Also, another path toward proving the applicability and efficiency of our approach would be to observe its behavior on scenarios based on real traffic data (instead of Markovian simulated traffic as we did in [9]).

Acknowledgements

We would like to thank three anonymous reviewers whose constructive comments have helped improve the paper.

This work has been partially funded by the Bretagne Region, through the ARED program, and by the Fondation Telecom through the "Futur&Ruptures" program.

References

[1] R. J. Aumann and L. S. Shapley, *Values of Non-Atomic Games*. Princeton University Press, 1974.

[2] M. Beckmann, C. B. McGuire, and C. B. Winsten, *Studies in the Economics of Transportation*. Yale University Press, 1956.

[3] M. Cesana, N. Gatti, and I. Malanchini, "Game theoretic analysis of wireless access network selection: Models, inefficiency bounds, and algorithms," in *Proceedings of 2nd Worshop on Game Theory in Communication Networks (GameComm)*, October 2008.

[4] R. Cole, Y. Dodis, and T. Roughgarden, "Pricing network edges for heterogeneous selfish users," in *Proceedings of 35th Annual ACM Symposium on Theory of Computing*, 2003.

[5] P. Coucheney, E. Hyon, C. Touati, and B. Gaujal, "Myopic versus clairvoyant admission policies in wireless networks," in *Proceedings of the Fourth International ICST Conference on Performance Evaluation Methodologies and Tools (VALUETOOLS)*, 2009.

[6] M. Coupechoux, J.-M. Kelif, and P. Godlewski, "Network controlled joint radio resource management for heterogeneous networks," in *Proceedings of IEEE Vehicular Technology Conference (VTC)*, 2008.

[7] A. De La Oliva, A. Banchs, I. Soto, T. Melia, and A. Vidal, "An overview of IEEE 802.21: media-independent handover services," *IEEE Wireless Communications*, vol. 15, no. 4, pp. 96–103, August 2008.

[8] L. Fleischer, K. Jain, and M. Mahdian, "Tolls for heterogeneous selfish users in multicommodity networks and generalized congestion games," in *Proceedings of IEEE Symposium on Foundations of Computer Science*, 2004.

[9] V. Fux, P. Maillé, J.-M. Bonnin, and N. Kaci, "Efficiency or fairness: managing applications with different delay sensitivities in heterogeneous wireless networks," in *Proceedings of WoWMoM*, 2013.

[10] E. Gustafsson and A. Jonsson, "Always best connected," *IEEE Wireless Communications*, February 2003.

[11] N. Kaci, P. Maillé, and J.-M. Bonnin, "Performance of wireless heterogeneous networks with always-best connected users,," in *Proceedings of 5th Conference on Next Generation Internet Networks (NGI)*, 2009.

[12] G. Karakostas and S. G. Kolliopoulos, "Edge pricing of multicommodity networks for heterogeneous selfish users," in *Proceedings of 45th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2004.

[13] G. Karakostas, S. G. Kolliopoulos, and G. Stavros, "Edge pricing of multicommodity networks for heterogeneous selfish users with elastic demands," in *Proceedings of 12th annual international conference on Computing and Combinatorics*, 2006.

[14] D. Kumar, E. Altman, and J.-M. Kelif, "User-network association in a WLAN-UMTS hybrid cell: Global individual optimality," in *Proceedings of International Teletraffic Congress (ITC)*, 2006.

[15] P. Maillé, B. Tuffin, and J.-M. Vigne, "Technological investment games among wireless telecommunications service providers," *International Journal on Network Management - Special Issue: Economic Traffic Management*, vol. 21, no. 1, pp. 65–82, 2011.

[16] ——, "Competition between wireless service providers sharing a radio resource," in *Proceedings of 11th IFIP Networking Conference*, Prague, Czech Republic, May 2012.

[17] K. Mittal, E. M. Belding, and S. Suri, "A game-theoretic analysis of wireless access point selection by mobile users," *Computer Communications*, vol. 31, no. 10, 2008.

[18] A. C. Pigou, *The economics of welfare*. Macmillan, 1920.

[19] J. Wardrop, "Some theoretical aspects of road traffic research," *Proceedings of the Institute of Civil Engineers*, 1952.