# Incentive Mechanisms for Internet Congestion Management: Fixed-Budget Rebate *versus* Time-of-Day Pricing

Patrick Loiseau, *Member, IEEE*, Galina Schwartz, *Member, IEEE*, John Musacchio, Saurabh Amin and S. Shankar Sastry, *Fellow, IEEE*,



***Abstract*—Mobile data traffic has been steadily rising in the past years. This has generated a significant interest in the deployment of incentive mechanisms to reduce peak-time congestion. Typically, the design of these mechanisms requires information about user demand and sensitivity to prices. Such information is naturally imperfect. In this paper, we propose a *fixed-budget rebate mechanism* that gives each user a reward proportional to his percentage contribution to the aggregate reduction in peak time demand. For comparison, we also study a time-of-day pricing mechanism that gives each user a fixed reward per unit reduction of his peak-time demand. To evaluate the two mechanisms, we introduce a game-theoretic model that captures the *public good* nature of decongestion. For each mechanism, we demonstrate that the socially optimal level of decongestion is achievable for a specific choice of the mechanism's parameter. We then investigate how imperfect information about user demand affects the mechanisms' effectiveness. From our results, the fixed-budget rebate pricing is more robust when the users' sensitivity to congestion is "sufficiently" convex. This feature of the fixed-budget rebate mechanism is attractive for many situations of interest and is driven by its closed-loop property, i.e., the unit reward decreases as the peak-time demand decreases.**

***Index Terms*—congestion pricing; lottery-based incentive mechanisms; public good provisioning; probabilistic pricing**

## I. Introduction

The consumer demand for network bandwidth is steadily growing. For instance, mobile data traffic nearly tripled during each of the past three years due to increasing penetration of mobile devices such as smartphones [1]. Numerous studies indicate that this growth will continue as bandwidth intensive applications like video streaming continue to gain popularity [2]. The growing demand for bandwidth forces the Internet Service Providers (ISPs) to adopt congestion management schemes, including capacity expansion and pricing mechanisms. Although the ISPs have historically used flat-rate pricing, many ISPs are now interested in moving to tiered pricing schemes [2], [3]. However, experiments have shown that users prefer flat-rate pricing, and will pay a premium to avoid being metered [4], [5]. This makes the adoption of real-time pricing particularly challenging. Thus, novel pricing mechanisms that balance the conflict between the need for network decongestion and the users' preference for flat prices are of great practical interest.

Network bandwidth (and hence the level of congestion) is not uniform during the course of a day; it drops at night after the prime time evening hours. This variability in demand can be exploited to design variable pricing mechanisms. For instance, time-of-day pricing mechanisms have been designed to incentivize users to shift a part of their demand to the off-peak times [6], [7]. However, such mechanisms typically require information about user demand; in particular, the knowledge of user preferences about shifting their demand from peak to off-peak times. In practice, this information may be inaccurate or just too difficult to obtain due to privacy concerns [7]. Thus, robustness to imperfect information about user preferences must be taken into account in the design of any practically viable mechanism.

Recently, a fixed-budget rebate mechanism (termed "raffle-based scheme") was proposed for decongestion of a shared resource [8]. Decongestion is viewed as a public good: when a user reduces/shifts his demand away from peak times, his contribution benefits all the users sharing the resource. The fixed-budget rebate mechanism in [8] is inspired by economic ideas on incentivizing contributions to provision of public goods [9]. In this mechanism, each user is entitled a reward proportional to his percentage contribution to the total demand reduction. An attractive feature of this mechanism is that, in practice, it can be implemented via a lottery scheme, where each participating user wins a *prize* with a probability equal to the fraction he contributed to the total demand reduction.

In this article, we investigate the fixed-budget rebate mechanism, and compare it with the more traditional time-of-day pricing mechanism for reducing Internet congestion. In Sec. III, we introduce a game-theoretic model with a continuum of non-atomic users. Each user chooses his peak time and off-peak time demand to maximize his utility. The user utility models both his benefit from peak time decongestion, and his willingness to reduce/shift away from the peak time period. The model allows us to compute the user equilibrium welfare for both mechanisms: fixed-budget rebate and time-of-day pricing. We compare their sensitivity to information imperfections for the case when an ISP with imperfect information about user demand chooses the mechanism parameters.

Patrick Loiseau is with EURECOM, Sophia-Antipolis, France. E-mail: patrick.loiseau@eurecom.fr. Part of this work was done while he was at UC Santa Cruz, CA, USA.

Galina A. Schwartz and S. Shankar Sastry are with the department of Electrical Engineering and Computer Science, UC Berkeley, CA, USA. E-mails: {schwartz, sastry}@eecs.berkeley.edu

John Musacchio is with the Technology and Information Management program, UC Santa Cruz, CA, USA. E-mail: johnm@soe.ucsc.edu

Saurabh Amin is with the department of Civil and Environmental Engineering, MIT, MA, USA. E-mail: amins@mit.edu

The authors thank the editor and the anonymous referees for their thorough reviews which lead to significant improvements of the paper.

This research was supported by NSF grant CNS-0910711 and by TRUST (Team for Research in Ubiquitous Secure Technology), which receives support from the NSF (#CCF-0424422) and the following organizations: AFOSR (#FA9550-06-1-0244), BT, Cisco, DoCoMo USA Labs, EADS, ESCHER, HP, IBM, iCAST, Intel, Microsoft, ORNL, Pirelli, Qualcomm, Sun, Symantec, TCS, Telecom Italia, and United Technologies.

Our results in Sec. IV can be summarized as follows:

(i) For any given parameters, for each mechanism, a Nash equilibrium exists, and it is unique.
(ii) For the case when ISP has perfect information about user demand, for each mechanism, the ISP can choose the mechanism parameter to achieve the socially optimal level of decongestion.
(iii) For the case when ISP has imperfect information about user preferences, the fixed-budget rebate mechanism is more robust to the time-of-day pricing mechanism a under mild condition on the users' sensitivity to congestion.

Our analysis reveals several desirable features of fixed-budget rebate mechanism. First, the condition under which it is more robust than the time-of-day-pricing can be interpreted as "convex" user sensitivity to congestion (or delay). This condition is expected to be predominant, especially for today's Internet which supports highly delay-sensitive applications. This robustness of the fixed-budget rebate mechanism is driven by its closed-loop property: as the aggregate demand shifts away from peak time period, the user reward for his per unit contribution decreases. Finally, if an ISP decides to implement the fixed budget rebate mechanism, he knows the total reward (or rebate) that he owes to the users even when the information about user demand characteristics is imperfect. In contrast, under the time-of-day pricing mechanism, the ISP will have to design the mechanism based on an estimate of the total expected reward that he will owe to the users.

The rest of the paper is organized as follows. Sec. II discusses the related literature. We introduce the model in Sec. III. In Sec. IV, we analyze the two incentive mechanisms (Nash equilibrium and social optimum) and compare them in terms of robustness to imperfect information. We conclude in Sec. V. Proofs are relegated to Appendices.

## II. Related work

Many pricing mechanisms have been proposed to manage quality of services (QoS) in networks, see e.g., surveys [10], [11], [12]. For instance, in [13], Honig and Steiglitz propose a usage-based pricing mechanism, and analyze it using a model with delay-sensitive users. Their results show how to find the price that maximizes the provider's revenue by solving a fixed-point equation. A similar model is used in [14] where Başar and Srikant analyze the many-users limit. They show that, as the number of users increases, the provider's revenue per unit of bandwidth increases and conclude that this gives providers an incentive to increase their network capacity. In a number of papers, e.g., [15], [16], [17], pricing mechanisms based on multiple classes of customers with different priorities are proposed and analyzed in terms of equilibrium achieved and optimal price per class. In [18], [19], Shen and Başar investigate the performance of non-linear pricing in a model similar to [14] and find an improvement of up to 38% over linear pricing in some cases. However, in all the aforementioned papers, the demand is assumed stationary and the price is fixed independently of the instantaneous network congestion or of the time of the day. In contrast, in this paper, we investigate linear pricing mechanisms that leverage the time variability of user demand using a single priority class.

A few papers have proposed mechanisms with prices dependent on congestion levels. In [20], Paschalidis and Tsitsiklis propose a congestion-based pricing mechanism in the context of loss networks (i.e., phone). They provide a dynamic programming formulation of the revenue maximization problem and of the welfare maximization problem. Then, they show that this dynamic congestion pricing mechanism can be well approximated by a simpler static time-of-day pricing. An alternative mechanism called "Trade & Cap", was recently proposed by Londoño, Bestavros and Laoutaris [21]. It works in two phases. First, users engage in a trading game where they choose an amount of *reserved* bandwidth slots to buy for hard-constraints traffic. In the second phase, the remaining bandwidth is allocated to users as *fluid* bandwidth, in proportion of their remaining "buying power". They show that this mechanism smoothes the aggregate demand to a certain level. In their model, users have a cost function that increases linearly with the total demand in a given slot. In this paper, we consider simpler one-phase pricing mechanisms with fixed parameters. Our model also differ from these papers in that users have elastic demand and their utility is an arbitrary function of the congestion level.

Two recent papers analyze time-of-day pricing mechanisms over $n$ time slots [6], [7]. In [6], Jiang, Parekh and Walrand consider a model where users have unit demand. Each user chooses one time-slot in which he transmits its entire demand, to maximize his utility. The authors of [6] obtain a bound on the price of anarchy due to users selfishness. In [7], Wong, Ha and Chiang consider a model with users transmitting *sessions* of random length. Sessions arrive as a Poisson process and each session is characterized by a *waiting function* which reflects the willingness of the user to delay his entire session for a given time, in exchange for a reward given by the provider. The authors show how to compute the optimal reward levels in order to maximize the provider profit by balancing the congestion cost due to demand exceeding capacity and the reward amount. Further analysis of this mechanism called "TUBE", as well as implementation are provided in [22]. However, in their model, users are only sensitive to prices (the effect of congestion on the user utility is not considered) and the analysis is not game-theoretic. In this paper, we consider a model with two time slots (peak and off-peak). We provide a game-theoretic analysis. In our model, user utility functions are the closest to [6] where user cost due to latency is an arbitrary (convex) function of the load. However, our setup differs from [6], as each user in our model can shift an arbitrary continuous fraction of his demand from peak time to off-peak time.

In this paper, we show that the problem of decongesting the peak time can be seen as a public good provision problem. Our model is closely related to the "raffle-based" incentive mechanism, which has been recently proposed by Loiseau, Schwartz, Musacchio, Amin and Sastry [8]. That work was inspired by Morgan, who in [9] pioneered the investigation of using the lotteries for public good provision. The public good perspective has been applied in recent works by Sharma and Teneketzis [23], [24] in the context of optimal power allocation for wireless networks. The connection of lottery-

based mechanism with public good provision was originally noted in [9] and received an extensive attention in economic literature (see [25]). The idea of lotteries has also been used in different contexts. For instance, lottery scheduling is a widely applied technique in resource scheduling in computer operating systems [26]. Recent interest in the application of lotteries to congestion management was facilitated by Merugu, Prabhakar and Rama who demonstrated with a field study that lottery-based mechanisms can be effectively used to reduce congestion in transportation systems [27]. In contrast, our contributions are methodological. We use a game-theoretic model to analytically study the performance of a lottery-like mechanism and compare it to a more standard time-of-day pricing mechanism.

## III. Model

Let us consider a shared Internet access point with capacity $C$. Based on the usage patterns, let the day be divided into two time periods: a peak time of duration $T_p$ and an off-peak time of duration $T_o$. We assume that each time period corresponds to a stationary regime with respective loads $\rho_p$ and $\rho_o$.

An access point is typically shared by a finite number of users, each having his own preference for time periods which we model by user *type* (the type of a user will typically depend on the applications that he uses). To account for a large number of users, we model the set of users as a continuum of non-atomic users; i.e., each user contributes a negligible fraction of the total demand. We use the measure-theoretic framework similar to [28], [29]. Let $(\Theta, \mathcal{F})$ be a measurable space where $\Theta$ is the set of user types. We assume that the user types are distributed according to a finite measure $\mu$ on $(\Theta, \mathcal{F})$[1]. While simpler modeling assumptions can be used (e.g., considering only two types), using an arbitrary measure of types $\mu$ gives a higher flexibility that can be interesting to fit real populations.

Note that for simplicity, we describe the population at the granularity of types instead of users as in [28], [29]. This is justified by the strict concavity of the user utilities (see assumptions below) which implies that at Nash equilibrium, all users of the same type choose the same action. As a consequence, although we do require that the measure of users is non-atomic (as for any non-atomic game), we do not require that the measure of types $\mu$ itself is non-atomic. For instance, if all users have the same type, measure $\mu$ is only constituted by one atom. Yet, each user of each type remains infinitesimally small, which means that the action of one user does not affect the aggregate outcome.

### A. User utility

Each user of type $\theta \in \Theta$ chooses his peak-time demand $y_\theta$ and his off-peak time demand $z_\theta$ to maximize his utility

$$u_\theta(y_\theta, z_\theta, y_{-\theta}, z_{-\theta}) = P_\theta(y_\theta) + O_\theta(z_\theta) - y_\theta L_p\left(\int_\Theta y_\theta \mathrm{d}\mu(\theta)\right) - z_\theta L_o\left(\int_\Theta z_\theta \mathrm{d}\mu(\theta)\right) - (y_\theta + z_\theta)\, q - p, \quad (1)$$

where the notation $y_{-\theta}$ and $z_{-\theta}$ is standard: it denotes peak-time and off-peak-time demand choices for all user types but $\theta$. In (1), $P_\theta(\cdot)$ and $O_\theta(\cdot)$ are the utilities that a user of type $\theta \in \Theta$ gets for his demand in the peak time and off-peak time respectively. $L_p(\cdot)$ and $L_o(\cdot)$ are the disutilities due to congestion in the peak time and off-peak time respectively. These disutilities are per unit of demand, hence they are multiplied by the demand in each time. Finally, quantity $q \geq 0$ is a fixed usage-based price (which could be zero) and quantity $p > 0$ is a fixed monthly subscription price.

We assume that utilities $P_\theta(\cdot)$ and $O_\theta(\cdot)$ are twice differentiable increasing strictly concave functions of the demand. We assume that there is a fixed maximum peak-time demand $d_p$ (per day) that could correspond for instance to a subscription daily peak cap, that could be a maximum usable demand (determined by technology limitation), or that could be a maximum daily demand determined from empirical data. For simplicity, we assume that this maximum peak-time demand is the same for each user, but more general cases could be handled easily[2] (in that case, user-dependent prices could also easily be handled). Each user can choose to shift to off-peak time, or to simply not use, part of his maximum peak-time demand. Additionally to the shifted peak-time demand, each user could have an initial off-peak-time demand. However, this additional demand does not modify our analysis as long as the peak time remains more congested than the off-peak time. For simplicity, we assume that the initial off-peak-time demand is zero, i.e., the off-peak-time demand only corresponds to shifted peak-time demand. Then, we have the following constraint on the demands:

$$y_\theta + z_\theta \leq d_p, \quad (\theta \in \Theta). \quad (2)$$

We assume that disutilities $L_p(\cdot)$ and $L_o(\cdot)$ are increasing strictly convex functions of the aggregate demand in each time (a similar assumption is made, e.g., by Jiang, Parekh and Walrand [6]). This assumption is realistic and quite general. As an example, let us focus on the average delay $\delta$ as a measure of the network quality, as in Honig and Steiglitz [13]. Our assumption holds if $(i)$ the disutility is an increasing convex function of the average delay and $(ii)$ the average delay is an increasing strictly convex function of the aggregate demand or equivalently of the load in the corresponding time:

$$\rho_p = (CT_p)^{-1} \int_\Theta y_\theta \mathrm{d}\mu(\theta), \quad \rho_o = (CT_o)^{-1} \int_\Theta z_\theta \mathrm{d}\mu(\theta).$$

Assumption $(i)$ is natural: an increase of the delay from zero to half a second creates no more disutility than from half a second to one second. This assumption is also made in [13]. Assumption $(ii)$ holds for the vast majority of queueing models considered in the literature. For example, it holds for the processor sharing queue (the most classical model for 3G and 4G networks [30]), for which the average delay is

$$\delta(\rho_p) = \frac{\delta_0}{1 - \rho_p}, \quad (3)$$

[1] Throughout the paper, we assume that all functions of $\theta$ are measurable. In [28], Aumann notes that "the measurability assumption is of technical significance only and constitutes no real economic restriction."

[2] If users differ by their maximum peak-time demand, each user could be viewed as an appropriate number of users with identical maximum peak-time demand. The proposed model still applies with measure $\mu$ defined for all subset $\Theta_1 \in \mathcal{F}$ by $\mu(\Theta_1) = \int_\Delta d \cdot \nu(\Theta_1, \mathrm{d}d)$ where $\Delta$ is the set of maximum peak-time demands and measure $\nu$ on $\Theta \times \Delta$ represents the joint distribution of types and demand.

where $\delta_0$ is a constant. It also holds for common models of wired networks such as the M/D/1 model considered in [13] and the M/M/1 model considered by Shen and Başar [18], [19]. Finally, we assume that despite the effect of users shifting part of their demand, the off-peak time remains relatively uncongested so that

$$L_o\left(\int_\Theta z_\theta \mathrm{d}\mu(\theta)\right) \simeq 0. \tag{4}$$

This assumption is not strictly necessary but greatly simplifies the presentation without affecting the important effects that we consider in this model[3].

For numerical illustrations of our model, we use the following example of an Internet access point.

**Example 1** (Internet access point)**.** The capacity is $C = 1$ Gbps. Peak-time lasts $T_p = 2$ h (e.g., 6pm to 8pm), hence $T_o = 22$ h. $\Theta = [0, 1]$ with a uniform distribution of types $\mu(\mathrm{d}\theta) = D_p/d_p \cdot \mathrm{d}\theta$, where $D_p = 7.2 \cdot 10^3$ Gbits and $d_p = 7.2$ Gbits (which corresponds to $1,000$ users with peak-time capacity 1 Mbps). The latency disutility is $L_p(\rho_p) = L_0\delta(\rho_p)$ where $\delta(\rho_p)$ is given by the PS model (3) and $L_0 = 0.065$. Peak-time utility is $P_\theta(x_\theta) = (1+\theta)P_0 \log(1 + x/d_p)$ with $P_0 = 130$ and off-peak time utility is $O_\theta = 1/10 \cdot P_\theta(\cdot)$. The subscription price is $p = \$50$ and the usage-base price is zero.

### B. User type distribution

On the timescale of a day, the population is heterogeneous with user types distributed according to measure $\mu$. However, we assume that each user has a type that varies randomly across the days of a month, with the same distribution $\mu$. Therefore, the population is homogeneous in average on the timescale of a month. In particular, with this assumption, the expected utility of each user on the timescale of a month equals the daily aggregate welfare

$$W = \int_\Theta u_\theta \mathrm{d}\mu(\theta) \tag{5}$$

normalized by $\mu(\Theta)$.

Each user will buy a monthly contract (with subscription price $p$) to use the service if his expected utility over the month is positive, i.e., here if

$$W > 0. \tag{6}$$

We assume that without any incentive mechanism, this condition is satisfied. Then, our assumption guarantees that with any welfare-improving incentive, each user will continue to participate, i.e., to buy the monthly contract.

Note that if the population cannot be assumed homogeneous at the timescale of a month, it is possible to divide it into subpopulations that can be assumed homogeneous and to apply our incentive mechanisms to each of these subpopulations.

### C. Model reduction to one-dimensional strategy space

Before introducing the incentive mechanisms, we show that our model can be reduced to a one-dimensional strategy space focusing on the peak-time demand reduction. With assumption (4), the utility (1) of a user of type $\theta \in \Theta$ can be re-written as

$$\begin{aligned} u_\theta(y_\theta, z_\theta, y_{-\theta}) =& P_\theta\left(y_\theta\right) - y_\theta \cdot L_p\left(\int_\Theta y_\theta \mathrm{d}\mu(\theta)\right) - y_\theta \cdot q \\ &+ O_\theta\left(z_\theta\right) - z_\theta \cdot q - p. \end{aligned} \tag{7}$$

Since we are interested in the reduction of peak time demand, we define the difference between the maximum peak-time demand and the chosen peak-time demand:

$$x_\theta = d_p - y_\theta, \quad (\theta \in \Theta). \tag{8}$$

This peak-time demand *reduction* includes both the unused peak-time demand and the peak-time demand shifted to off-peak time. For a given $x_\theta \in [0, d_p]$, we define the optimal shifted demand:

$$z^*_\theta(x_\theta) = \underset{z_\theta \in [0, x_\theta]}{\arg\max} \left[O_\theta(z_\theta) - z_\theta q\right], \quad (\theta \in \Theta).$$

A user of type $\theta \in \Theta$ maximizing his utility (7) will choose a couple $(x_\theta, z_\theta)$ such that $z_\theta = z^*_\theta(x_\theta)$.[4] As we are interested in the reduction of congestion at peak-time, we restrict our attention to the choice of $x_\theta$. Note that if $q = 0$, then $z^*(x_\theta) = x_\theta$. Indeed, if there is no usage-based cost, off-peak-time demand always gives higher utility than 0.

In the absence of latency, the maximal utility of a user is

$$\bar{u}_\theta = P_\theta\left(d_p - \underline{x}_\theta\right) - \left(d_p - \underline{x}_\theta\right) q + O_\theta\left(z^*_\theta(\underline{x}_\theta)\right) - z^*_\theta(\underline{x}_\theta) q, \tag{9}$$

for all $\theta \in \Theta$, where

$$\begin{aligned} \underline{x}_\theta = \underset{x_\theta \in [0, d_p]}{\arg\max} \Big\{ & P_\theta\left(d_p - x_\theta\right) - \left(d_p - x_\theta\right) q \\ & + O_\theta\left(z^*_\theta(x_\theta)\right) - z^*_\theta(x_\theta) q \Big\}, \quad (\theta \in \Theta), \end{aligned}$$

is the baseline peak-time demand reduction which maximizes the latency-free utility. Latency and incentive mechanisms will only result in users using less of their peak-time demand, i.e., increasing their choice of $x_\theta$ beyond $\underline{x}_\theta$. Then, we define the *cost of shifting* as the loss of utility incurred by a user when reducing his peak-time demand:

$$\begin{aligned} c_\theta(x_\theta) = \bar{u}_\theta - \Big[ & P_\theta\left(d_p - x_\theta\right) - \left(d_p - x_\theta\right) q \\ & + O_\theta\left(z^*_\theta(x_\theta)\right) - z^*_\theta(x_\theta) q \Big], \quad (\theta \in \Theta). \end{aligned} \tag{10}$$

(Note that with a slight abuse of terminology, we call $c_\theta(x_\theta)$ the cost of *shifting* whereas the peak-time demand reduction $x_\theta$ can actually correspond to shifted demand and/or to unused demand.) The definition of the baseline (9) guarantees that $c_\theta(x_\theta)$ is always positive. Moreover, with our assumptions on functions $P_\theta(\cdot)$ and $O_\theta(\cdot)$, the cost of shifting $c_\theta(\cdot)$ is differentiable and strictly convex on $[0, d_p]$; and increasing on $[\underline{x}_\theta, d_p]$ (see details in Appendix A). Finally, to simplify the proofs, we assume that the marginal cost of shifting is bounded by a constant independent of $\theta$.

We view the aggregate peak-time demand reduction

$$G = \int_\Theta x_\theta \mathrm{d}\mu(\theta) \tag{11}$$

as a *public good* to which each user contributes by his choice of $x_\theta$. Indeed, when a user reduces his peak-time demand, the benefits of reduced peak-time congestion is shared by all the users. We define the function

[3] If one wants to consider a non-zero off-peak-time disutility, this assumption could be replaced by the relaxed assumption that when the aggregate shifted demand increases, the marginal peak-time disutility reduction is higher than the marginal off-peak-time disutility increase.

[4] A function $(x_\cdot, z_\cdot)$ corresponding to social welfare maximization also satisfies $z_\theta = z^*_\theta(x_\theta)$ for all $\theta \in \Theta$.

$$h(G) = -L_p\left(D_p - G\right), \tag{12}$$

where $D_p = d_p\mu(\Theta)$ is the aggregate maximum peak-time demand. Function $h(\cdot)$ reflects the notion of how much users benefit from the network decongestion at peak-time. With our assumptions on $L_p(\cdot)$, $h(\cdot)$ is an increasing concave function of the public good level $G$. The term $-y_\theta L_p\left(\int_\Theta y_\theta \mathrm{d}\mu(\theta)\right) = (d_p - x_\theta)h(G)$ in (7) has the interpretation that the benefit a user gets from the peak-time decongestion is the product of his peak-time demand $(d_p - x_\theta)$ times the benefit per unit demand $h(G)$. Notice that $h(G)$ is negative, but its most important characteristic is that it is increasing in $G$, i.e., the disutility due to congestion reduces when $G$ increases.

In summary, in view of (7)-(12), our peak-time decongestion model reduces to a public good provision problem similar to [8]: the utility of a user of type $\theta \in \Theta$ is

$$u_\theta(x_\theta, G) = \bar{u}_\theta + (d_p - x_\theta)h(G) - c_\theta(x_\theta) - p, \tag{13}$$

where $h(\cdot)$ corresponds to the (unit) benefit from the public good and $c_\theta(\cdot)$ corresponds to the cost of contribution. From our assumptions, these functions satisfy:

**(A1)** $h(\cdot)$ is twice differentiable, strictly concave and increasing on $[0, D_p]$;

**(A2)** $c_\theta(\cdot)$ is positive, differentiable and strictly convex on $[0, d_p]$; and increasing on $[\underline{x}_\theta, d_p]$, $\quad(\theta \in \Theta)$;

**(A3)** $\sup_{\theta\in\Theta} c'_\theta(d_p) < \infty$, $\quad(\theta \in \Theta)$.

Notice that in our model, $h(\cdot)$ does not depend on the type. All the type-dependency is carried by the cost of shifting. This modeling choice ensures tractability of the equilibrium.

### *D. Incentive mechanisms*

Individual users maximize their own utility (13), which differs from maximizing (5). Thus, in general, the level of public good and the aggregate user welfare achieved in the individual maximization and in the social optimum differ.

To align Nash equilibrium and social optimum objectives, the service provider can design mechanisms to incentivize users to reduce their peak-time demand. In this paper, we compare two different incentive mechanisms: a fixed-budget rebate mechanism (denoted $R$ or FBR) and a time-of-day pricing mechanism (denoted $T$ or TDP). Each mechanism introduces a reward based on the peak-time demand reduction $x_\theta$ below the maximum $d_p$. For the service provider to finance the respective reward, each mechanism also introduces an increase in the subscription price. However, as we will see (Corollary 1), each user's net utility can be improved even with this price increase. With mechanism $j \in \{R, T\}$, the user utility becomes

$$u^j_\theta(x_\theta, G) = u_\theta(x_\theta, G) + M^j(x_\theta, G), \quad (\theta \in \Theta). \tag{14}$$

The fixed-budget rebate mechanism consists in giving each user a reward proportional to his fraction of the total contribution, i.e., of the functional form:

$$M^R(x_\theta, G) = R \cdot \frac{x_\theta}{G} - \Delta p_R, \quad \text{[fixed-budget rebate]} \tag{15}$$

where $R$ is a parameter of the mechanism chosen *ex-anti* by the provider. In practice, this mechanism could be implemented via randomization. For example, with a finite number of users, it could be implemented by the simplest type of lottery where each user wins the prize $R$ with a probability equal to his percentage contribution to the total amount of peak-time demand reduction. In this case, (14) and (15) would correspond to expected utilities. Other implementations (e.g., deterministic) are also possible. To complete the definition of the fixed-budget rebate mechanism, we assume that if no user reduces his peak-time demand then the reward is not given. However, if the set of users who reduce their peak-time demand is nonempty but of measure zero, then each contributing user receives an infinite reward given in such a way that the integral w.r.t. the measure of users is $R$. This is a technical assumption for the measure-theoretic setting of the non-atomic game. In practice, it reflects the fact that if only a finite number of users contribute, their expected reward relative to their fraction of the total demand grows to infinity as the total number of users goes to infinity.

We notice here that the fixed-budget rebate mechanism introduces uncertainty in the users bill as the reward depends on the amount shifted by the other users. However, this uncertainty is only one-sided: the maximum bill is known and only the reward amount is uncertain. This asymmetry is crucial to ensure good adoption of the mechanism.

The time-of-day pricing mechanism corresponds to a fixed reward per unit of shifted demand:

$$M^T(x_\theta, G) = r \cdot x_\theta - \Delta p_T, \quad \text{[time-of-day pricing]} \tag{16}$$

where $r$ is a parameter of the mechanism chosen *ex-anti* by the provider. This mechanism is a variation of a conventional time-of-day pricing mechanism, with an off-peak price subsidy. Its implementation is straightforward.

In (15) and (16), $\Delta p_j$ denotes the increase in the subscription price that the service provider imposes to finance the reward mechanism. Let $G^{(\text{eq})}$ be the equilibrium level of public good (in the next section, we show that the Nash equilibrium is unique for both mechanism). We assume that the price $\Delta p_j$ is fixed in advance by the service provider to compensate the reward, i.e., such that $\int_\Theta M^j(x_\theta, G^{(\text{eq})})\mathrm{d}\mu(\theta) = 0$ (note that the expression of the aggregate welfare (5) is thus not directly modified by the mechanisms, but only through the chosen contributions $x_\theta$)[5]. Then,

$$\Delta p_R = R \cdot \frac{d_p}{D_p} \quad \text{and} \quad \Delta p_T = rG^{(eq)} \cdot \frac{d_p}{D_p}. \tag{17}$$

From (17), we immediately see that the service provider has to know the equilibrium to determine the price $\Delta p_T$ for the time-of-day pricing mechanism. An error in the estimation of $G^{(\text{eq})}$ could have important consequences. In contrast, such knowledge is not necessary for the fixed-budget rebate mechanism where $\Delta p_R$ only depends on the parameter $R$ chosen by the service provider.

The marginal utility with mechanism $j \in \{R, T\}$ is

$$\frac{\partial u^j_\theta}{\partial x_\theta} = -h(G) - c'_\theta(x_\theta) + M^{j\prime}(G), \quad (\theta \in \Theta), \tag{18}$$

where

$$M^{R\prime}(G) = \frac{R}{G} \quad \text{and} \quad M^{T\prime}(G) = r. \tag{19}$$

[5]If users have different maximum peak-time demand for which they are charged different subscription prices, it is also possible to impose a type-dependent price increase $\Delta p_\theta$ which compensate the reward, i.e., such that $\int_\Theta M^j(x_\theta, G^{(\text{eq})})\mathrm{d}\mu(\theta) = 0$ is still satisfied.

Notice that in (18), there is no $h'(G)$ term corresponding to the variation of the aggregate due to the variation of a user's decision. This is because, in the non-atomic game, users are negligible and do not account for the variation of $G$ induced by their action when taking their decision. In (18) and in all future occasions, we abuse notation by denoting with a partial derivative w.r.t. $x_\theta$ the marginal quantities corresponding to variations following the variation of a user's action.

For both mechanisms, the marginal reward $M^{j\prime}$ is independent of the individual contribution $x_\theta$. Due to the term $-h(G)$ in (18), the marginal utility decreases when $G$ increases. Intuitively, if the congestion is lower at peak time, a user would want to use it more. Hence he would want to reduce less his peak-time demand. This decrease of the marginal utility is accentuated by the term $M^{R'}(G) = R/G$ in the case of the fixed-budget reward mechanism.

## IV. Analysis

In this section,[6] we show that, for each mechanism, there exists a unique Nash equilibrium. Then, we show that for appropriate values of the mechanisms parameters, they achieve social optimum and that for a wide range of parameters, both mechanisms are welfare improving. Finally, we compare the two mechanisms based on their sensitivity to imperfect information about the user utilities.

For clarity, we will use the following notation: $\Gamma_R\left(\Theta, \mu, h, \{c_\theta\}_{\theta\in\Theta}, R\right)$ and $\Gamma_T\left(\Theta, \mu, h, \{c_\theta\}_{\theta\in\Theta}, r\right)$ are the non-atomic games where users selfishly optimize their own utility (14) in the fixed-budget rebate mechanism and in the time-of-day pricing mechanism respectively. We denote with the superscript $^{(\text{eq})}$ the quantities at equilibrium in both games and we explicitly write their dependence on $r$ and $R$ or on other parameters whenever necessary to avoid ambiguity. Similarly, we denote with the superscript $^*$ the social optimum quantities corresponding to the maximization of (5), and denote explicitly their dependence on parameters whenever necessary.

### A. Nash equilibrium existence and uniqueness

We define a Nash equilibrium of the non-atomic game $\Gamma_j$ ($j \in \{R, T\}$) as a function $x^{(\text{eq})} : \Theta \to [0, d_p]$ such that for all $\theta \in \Theta$, $u_\theta^j(x_\theta, x_{-\theta}^{(\text{eq})}) \leq u_\theta^j(x_\theta^{(\text{eq})}, x_{-\theta}^{(\text{eq})}), \forall x_\theta \in [0, d_p]$. Due to the strict concavity of the utility $u_\theta^j$, it is equivalent to $x^{(\text{eq})}$ satisfying the first-order conditions (FOCs)

$$\frac{\partial u_\theta^j}{\partial x_\theta} \begin{cases} \leq 0, & \forall\theta : x_\theta = 0, \\ = 0, & \forall\theta : x_\theta \in (0, d_p), \\ \geq 0, & \forall\theta : x_\theta = d_p, \end{cases} \tag{20}$$

where $\frac{\partial u_\theta^j}{\partial x_\theta}$ is given by (18), and satisfying (11).

The first theorem establishes existence and uniqueness of the Nash equilibrium for both incentive mechanisms.

**Theorem 1.** *For the fixed-budget rebate mechanism, for any $R \geq 0$, there exists a unique Nash equilibrium $x^{(eq)}(R)$.*

*The same result holds for the time-of-day pricing mechanism, for any $r \geq 0$.*

[6]Some of the first results of this section appeared in [8] for the fixed-budget rebate mechanism. They are extended here to handle both mechanisms.

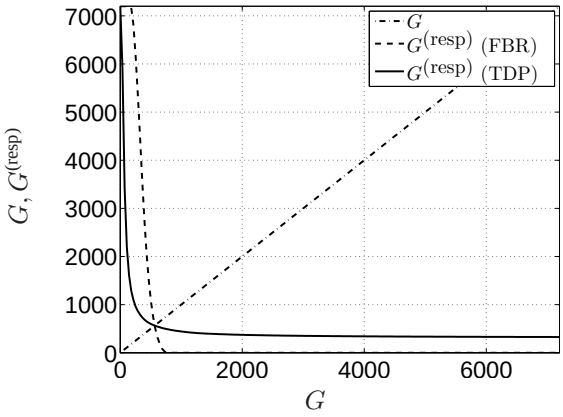


Fig. 1. Illustration of the fixed-point equation (21) for Example 1. The dashdotted line corresponds to the amount $G$ that users shifts (r.h.s. of (21)). The dashed and solid lines correspond to the aggregate amount $G^{(\text{resp})}$ that users would want to shift (l.h.s. of (21)) given $G$ has been shifted for the fixed-budget rebate and time-of-day pricing mechanisms respectively. To obtain $G^{(\text{eq})} = G^* = 78,000$, parameters were set to $R = \$5,500$ and $r = \$9$/Gbit. The corresponding subscription price increase is $\$5.5$.

Intuitively, for a given level of public good $G$, each user of type $\theta \in \Theta$ chooses his best response contribution $x_\theta^{(\text{resp})}(G) \in [0, d_p]$ to maximize his utility. Then, integrating the contribution of each type gives the amount of public good $G^{(\text{resp})}(G)$ that users want to provide in response to a given $G$. An equilibrium occurs when both quantities are equal, which corresponds to solving the fixed-point equation

$$G^{(\text{resp})}(G) = G. \tag{21}$$

Fig. 1 illustrates the two terms of the fixed-point equation for both mechanisms. As we mentioned, a key feature of our model is that the higher the level of public good $G$ is (i.e., the lower peak-time congestion is), the fewer users are willing to reduce their peak-time demand (the marginal utility (18) is decreasing in $G$). Therefore the aggregate best response $G^{(\text{resp})}(G)$ decrease when $G$ increases and this decrease is faster for the fixed-budget rebate mechanism for which the marginal utility decreases faster. Moreover, $G^{(\text{resp})}(G)$ is continuous, which leads to a unique fixed point. The continuity of $G^{(\text{resp})}(G)$ is due to assumption **(A2)** (a linear cost of shifting could induce discontinuities where a slight modification of $G$ would make some users switch from not reducing their peak-time demand to reducing it by $d_p$).

### B. Social optimum

We now show that the social optimum is unique and coincides with the Nash equilibrium of both mechanisms for parameters $R^*$ and $r^*$ given in the next theorem.

**Theorem 2.** *The following characterizes the social optimum:*

*(i) There exists a function $x^*$, uniquely determined almost-everywhere, which maximizes the aggregate welfare* (5).

*(ii) For the fixed-budget rebate mechanism, we have $x^{(eq)}(R) = x^*$ almost-everywhere (and hence $G^{(eq)}(R) = G^*$) for $R = R^*$, where*

$$R^* = G^* h'(G^*)(D - G^*). \tag{22a}$$

*The same result holds for the time-of-day pricing mechanism for $r = r^*$, where*

$$r^* = h'(G^*)(D - G^*). \tag{22b}$$

TABLE I
EFFECT OF THE INCENTIVE MECHANISMS ON CONGESTION FOR EXAMPLE 1 (CF. FIG. 1). THE RIGHT COLUMN CORRESPOND TO ANY OF THE TWO MECHANISM WITH ITS OPTIMAL PARAMETERS $R^* = \$5,500$ AND $r^* = \$9$/GBIT (I.E., TO SOCIAL OPTIMUM).

| | no incentive mechanism | incentive mechanism with optimal parameter (= social optimum) |
|---|---|---|
| G<br>W | 55 Gbits<br>28,000 | 565 Gbits<br>78,000 |
| $\rho_p$<br>$\delta(\rho_p)$ | 0.99<br>130 s | 0.92<br>12 s |
| $\rho_o$<br>$\delta(\rho_o)$ | 0.092<br>1.10 s | 0.098<br>1.11 s |

Intuitively, this result holds because the externality faced by a user ($-h(G) + M^{j'}$) in the game corresponding to any mechanism is independent of his type. Therefore, by fixing a reward that is also independent of the type, it is possible to achieve social optimum (similarly to a Pigovian tax [31]).

For Example 1, Tab. I illustrates the effect of the incentive mechanisms with the optimal parameters of Theorem 2: they permit a $180\%$ increase of the aggregate welfare which, in our model, also correspond to a $180\%$ increase of the average utility of each user over the timescale of a month. Peak-time congestion is significantly decreased: the load is decreased by $7\%$ but the average delay drops by $90\%$. On the other hand, off-peak time decongestion is hardly increased.

### C. Nash equilibrium variation with the mechanism parameters

In this section, we investigate the variation of the equilibrium quantities when the mechanism parameters $r$ and $R$ vary. For ease of exposition, we first assume that the participation constraint (6) is not imposed (we will come back to the effect of the participation constraint later in this section, see Proposition 4). Then, we have the following results on the variations of the equilibrium contributions.

**Proposition 1.** *If the participation constraint* (6) *is not imposed, for the fixed-budget rebate mechanism, we have:*

*(i) For any $R' > R$, $x_\theta^{(eq)}(R') \geq x_\theta^{(eq)}(R)$ ($\forall \theta \in \Theta$); and the inequality is strict if $0 < x_\theta^{(eq)}(R) < d_p$.*

*(ii) For any $R' > R$, $G^{(eq)}(R') \geq G^{(eq)}(R)$; and the inequality is strict if $0 < G^{(eq)}(R) < D_p$.*

*(iii) There exists a threshold $\bar{R} > R^*$ such that, for any $R \geq \bar{R}$, $x_\theta^{(eq)}(R) = d_p$ for all $\theta \in \Theta$ and $G^{(eq)}(R) = D_p$.*[7]

*The same results hold for the time-of-day pricing mechanism by changing $R$ to $r$ everywhere.*

Intuitively, since the marginal utility (18) increases with the reward parameters, the equilibrium contributions of each user increases (result *(i)*); and similarly for the equilibrium level of public good (result *(ii)*). The existence of the thresholds $\bar{R}$ and $\bar{r}$ (result *(iii)*) is a consequence of assumption **(A3)** which means that reducing even the last bit of his peak-time demand implies a finite marginal cost for the user, which can be compensated by a large-enough reward. Clearly, a case with such a large reward will not happen in practice, nevertheless we include it here for completeness of the model analysis.

Proposition 1 implies that for large enough parameters, the equilibrium level of public good will be positive. Let us define, for the fixed-budget rebate mechanism, $\underline{R}$ as the smallest parameter value such that $G^{(eq)}(R) > 0$ for $R > \underline{R}$; and similarly $\underline{r}$ for the time-of-day pricing mechanism. Then we have the following result characterizing these thresholds.

**Proposition 2.** *For the fixed-budget rebate mechanism, $\underline{R} = 0$, i.e., $G^{(eq)}(R) > 0$ for any $R > 0$ (if the participation constraint* (6) *is not imposed).*

*For the time-of-day pricing mechanism, $\underline{r} \geq 0$.*

The intuition behind Proposition 2 is as follows. For the fixed-budget rebate mechanism, for any $R > 0$, the marginal reward is infinite at $G = 0$. All users want to contribute hence this is not an equilibrium. In contrast, for the time-of-day pricing mechanism, the marginal reward is constant. If it is small enough so that the marginal utility of almost-all user types is non-positive at $G = 0$, then it is the equilibrium.

Note that Proposition 2 holds independently of the value of $G^*$ and is consistent with Theorem 2. In particular, if $G^* = 0$, then social optimum is achieved at Nash equilibrium for the fixed-budget rebate mechanism only for $R = R^* = 0$; whereas social optimum is achieved at Nash equilibrium for the time-of-day pricing mechanism for any $r$ smaller than $\underline{r}$.

The next proposition describes the evolution of the aggregate welfare with the mechanism parameters.

**Proposition 3.** *If the participation constraint* (6) *is not imposed, for the fixed-budget rebate mechanism, the equilibrium aggregate welfare $W^{(eq)}(R)$ is increasing in $[0, R^*]$, decreasing in $[R^*, \bar{R}]$ and constant for $R \geq \bar{R}$.*

*For the time-of-day pricing mechanism, the equilibrium aggregate welfare $W^{(eq)}(r)$ is constant on $[0, \underline{r}]$. For $r \geq \underline{r}$, the same results as for the fixed-budget rebate mechanism hold by changing $R$ to $r$ everywhere.*

Proposition 3, illustrated on Fig. 2 shows that the welfare is unimodal. If $G^* > 0$, it increases to its only maximum at $R^*$ or $r^*$ and then decreases. If $G^* = 0$ (hence $R^* = 0$ and $r^* = \underline{r}$), the welfare is maximal at $R = 0$ or $r = 0$ (i.e., with no incentive mechanism) and it only decreases (after a constant phase for the time-of-day pricing mechanism).

In extreme cases where the reward parameter is too large, the equilibrium aggregate welfare may become negative. For instance, consider a case where the usage-based price $q$ is so high compared to the off-peak time utility $O_\theta(\cdot)$ that all users have zero off-peak time demand. If the reward is larger than $\bar{R}$ or $\bar{r}$, then users would not use the service at all and the aggregate welfare would be $-p\mu(\Theta) < 0$. In that case, the participation constraint is not satisfied, hence users will simply not buy the service. The next proposition, which is easily derived using the monotonicity of Proposition 3, describes how the previous results are changed when introducing the participation constraint.

**Proposition 4.** *If the participation constraint* (6) *is imposed, for the fixed-budget rebate mechanism, there exists a threshold*

[7] To avoid ambiguity on the definition of the thresholds $\bar{r}, \bar{R}$, we assume that they are the smallest possible such thresholds.

$R_{max} \in (R^*, \infty]$ *such that*

*(i) For all $R < R_{max}$, all the users buys the monthly subscription and the results of Proposition 1, Proposition 2 and Proposition 3 hold.*

*(ii) For all $R > R_{max}$, no user buys the monthly subscription, hence the welfare is zero.*

*The same results hold for the time-of-day pricing mechanism by changing $R$ to $r$ everywhere.*

The effect of the participation constraint is simple: below a threshold $R_{\max}$ or $r_{\max}$, all the users participate and above this threshold, no users participate. This is due to our assumption that the population is homogeneous at the timescale of a month. Since users are offered a monthly subscription, they will buy it if they expected utility over the month is positive which is equivalent to the aggregate welfare being positive. Due to our assumption that the welfare is positive without any incentive mechanism, we have $R_{\max} > R^*$, i.e., the welfare is positive for any $R \leq R^*$. The threshold $R_{\max}$ can even be infinite if the off-peak time utility is high enough and the usage-based price is small enough so that users have positive utility over the month even without using the peak time.

The last result, which is a direct consequence of the previous results of this section, shows that both mechanisms are welfare improving for a wide range of parameters.

**Corollary 1.** *If $G^* > 0$, the fixed-budget rebate mechanism is strictly welfare improving for any parameter $R$ in a range $(0, R_0)$ where $R_0 \in (R^*, R_{max}]$:*

$$W^{(eq)}(R) > W^{(eq)}(0), \quad \forall R \in (0, R_0).$$

*The same results hold for the time-of-day pricing mechanism by changing $R$ to $r$ everywhere, and $0$ to $\bar{r}$.*

This result is important as it shows that, by implementing an incentive mechanism with a parameter lying in a wide range around an optimal parameter, the provider can increase welfare. Fig. 2 shows that for Example 1, the time-of-day pricing mechanism with any parameter in $(0, 2r^*)$ is welfare improving, and the fixed-budget rebate mechanism with any parameter in $(0, 14R^*)$ is welfare improving However, a consequence of Proposition 3 is that both mechanisms can "overshoot": if $R$ or $r$ is too large (larger that $R^*$ or $r^*$), $G^{(\text{eq})}$ can be larger than $G^*$ and the aggregate user welfare is suboptimal. In a competitive environment, a provider would not intentionally choose an overshooting parameter because it would be a competitive disadvantage as compared to a provider choosing an optimal parameter. However, if the provider has imperfect information about user utilities, it may overshoot unintentionally. Fig. 2 suggests that in this case, the aggregate welfare remains higher for the fixed-budget rebate mechanism than for the time-of-day pricing mechanism. In the next section, we investigate in details the robustness of each mechanism to imperfect information about user utilities.

### *D. Comparison of the two incentive mechanisms*

In this section, we compare the sensitivity of the two incentive mechanisms to imperfect information about user utilities. Let the games $\Gamma_R\left(\Theta, \mu, h, \{c_\theta\}_{\theta\in\Theta}, R^*\right)$ and

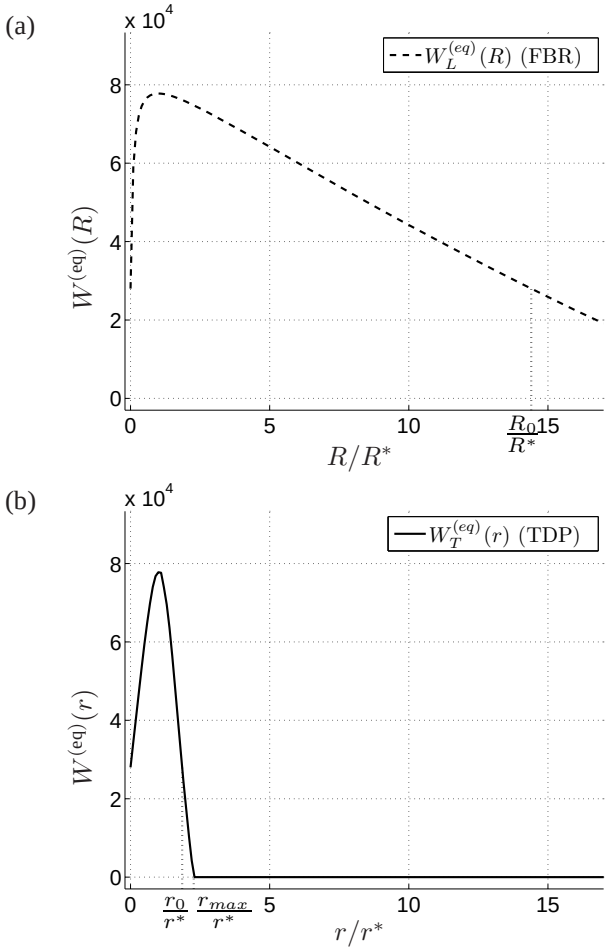


Fig. 2. Variation of the equilibrium aggregate welfare with the reward parameter for Example 1: (a) for the fixed-budget rebate mechanism – (b) for the time-of-day pricing mechanism.

$\Gamma_T\left(\Theta, \mu, h, \{c_\theta\}_{\theta\in\Theta}, r^*\right)$ correspond to the baseline case of perfect information considered in the previous sections and suppose that $R^*$ and $r^*$ have been chosen according to (22) to induce a socially optimal level of public good at equilibrium (i.e., $G^{(\text{eq})} = G^*$). We assume that $G^* \in (0, D_p)$. We analyze the variations in equilibrium and in social optimum when $R^*$ and $r^*$ are maintained for the respective mechanisms and utilities are perturbed (i.e., actual utilities are different from the estimation used by the provider to set the parameters).

We restrict our analysis to the case where only the cost of shifting is perturbed and the rest of the utilities is unchanged. Indeed, we argue that it is more difficult to obtain data on the time preferences (the willingness to move demand from peak time to off-peak time) than on the total demand or on the sensitivity to delay. Therefore, the cost of shifting is more likely to be imperfectly estimated by the provider. We consider the following general form of the perturbed cost of shifting:

$$\tilde{c}_\theta(\cdot) = c_\theta(\cdot) + \epsilon \cdot p_\theta(\cdot), \tag{23}$$

where $\epsilon$ is a real number and $p_\theta : [0, d_p] \to \mathbb{R}$ is a continuously differentiable function satisfying

$$\sup_{\theta\in\Theta} \sup_{x\in[0,d_p]} |p'_\theta(x)| < \infty.$$

Parameter $\epsilon$ is the perturbation magnitude and function $p_\theta(\cdot)$ is the direction of the perturbation. For the analysis, we restrict to small perturbations, i.e., $|\epsilon|$ small. For $|\epsilon|$ small enough, the

perturbed functions $\tilde{c}_\theta(\cdot)$ satisfy assumption **(A2-3)**. We assume that the perturbation direction is such that the aggregate best response has a non-zero perturbation at the order one in $\epsilon$ at the point $G^*$ (the non-perturbed equilibrium). Otherwise, the equilibrium point would not be changed by the perturbation. For numerical illustrations, we will use the following simple perturbation which satisfies the above conditions: $p_\theta(\cdot) = c_\theta(\cdot)$ for all types $\theta \in \Theta$, i.e., $c_\theta(\cdot)$ is scaled by a factor $1+\epsilon$ independent of the type.

Let $G_R^{(\text{eq})}(\epsilon)$ and $G_T^{(\text{eq})}(\epsilon)$ be the equilibrium levels of public good in the games with perturbed utilities $\Gamma_R(\Theta, \mu, h, \{\tilde{c}_\theta\}_{\theta\in\Theta}, R^*)$ and $\Gamma_T(\Theta, \mu, h, \{\tilde{c}_\theta\}_{\theta\in\Theta}, r^*)$, respectively. Let $W_R^{(\text{eq})}(\epsilon)$ and $W_T^{(\text{eq})}(\epsilon)$ be the corresponding equilibrium welfares. Let $G^*(\epsilon)$ and $W^*(\epsilon)$ be the socially optimal level of public good with perturbed utilities, and the corresponding welfare resulting from the maximization of (5) where $c_\theta(\cdot)$ is replaced by $\tilde{c}_\theta(\cdot)$. To evaluate the variation of $G^{(\text{eq})}$ with the perturbation, we need to evaluate the variation of the aggregate best response $G^{(\text{resp})}$ (recall that $G^{(\text{eq})}$ is the fixed-point of $G^{(\text{resp})}(\cdot)$, see (21)). (The variation of $G^*$ is handled similarly since from Theorem 2, the social optimum can also be seen as a Nash equilibrium in a mechanism $SO$ with unit reward given by (24c).) For this purpose, we introduce, for each mechanism $j \in \{R, T, SO\}$, the quantity $\alpha_j$ equal to the opposite of the slope of $G^{(\text{resp})}(\cdot)$ at the common non-perturbed equilibrium point $G^* = G^*(0)$ (see (33)). We define the following conditions:

**(C1)** $\left|\frac{1}{1+\alpha_R} - \frac{1}{1+\alpha_{SO}}\right| < \left|\frac{1}{1+\alpha_T} - \frac{1}{1+\alpha_{SO}}\right|$,

**(C2)** $\left|\frac{1}{1+\alpha_R} - \frac{1}{1+\alpha_{SO}}\right| > \left|\frac{1}{1+\alpha_T} - \frac{1}{1+\alpha_{SO}}\right|$.

If the slopes $\alpha_j$'s for the different mechanisms are close enough, these conditions reduce to the following more intuitive conditions (see details in Appendix G-B):

**(C1′)** $|r'_R(G) - r'_{SO}(G)| < |r'_T(G) - r'_{SO}(G)|$, at $G = G^*(0)$,

**(C2′)** $|r'_R(G) - r'_{SO}(G)| > |r'_T(G) - r'_{SO}(G)|$, at $G = G^*(0)$,

where $r'_R$, $r'_T$, $r'_{SO}$ are the respective derivatives of the unit rewards

$$r_R(G) = \frac{R}{G}, \tag{24a}$$

$$r_T(G) = r, \tag{24b}$$

$$r_{SO}(G) = h'(G)(D - G). \tag{24c}$$

Then we have the following results.

**Proposition 5.** *There exists $\epsilon_m > 0$ such that, for any perturbation* (23) *with $\epsilon \neq 0$ and $|\epsilon| < \epsilon_m$,*

*(i) if condition **(C1)** is satisfied, then*

$$\left|G_R^{(eq)}(\epsilon) - G^*(\epsilon)\right| < \left|G_T^{(eq)}(\epsilon) - G^*(\epsilon)\right|;$$

*(ii) if condition **(C2)** is satisfied, then*

$$\left|G_R^{(eq)}(\epsilon) - G^*(\epsilon)\right| > \left|G_T^{(eq)}(\epsilon) - G^*(\epsilon)\right|.$$

The intuition behind Proposition 5 is the following: the mechanism with the unit reward closer to the optimal unit reward $r_{SO}(G)$ have an equilibrium closer to the social optimum equilibrium $G^*(\epsilon)$. Since $r_R(G)$ and $r_{SO}(G)$ are both decreasing functions, one expects $r_R(G)$ to be closer to $r_{SO}(G)$ than $r_T(G)$. It is often the case. The fact that $r_R(G)$ decreases when $G$ increases is the *closed-loop* effect: the more users reduce their peak-time demand, the lower the incentive to reduce it is. However, if $r_R(G)$ decreases much faster that $r_{SO}(G)$, $r_T(G)$ can be closer to $r_{SO}(G)$. This possibility is covered by case *(ii)* of Proposition 5.

Fig. 3 illustrates the result of Proposition 5 with the perturbation $p_\theta(\cdot) = c_\theta(\cdot)$ for all $\theta \in \Theta$. As it turns out, Example 1 (Fig. 3-(a)) falls in case *(i)* of Proposition 5 (see the unit rewards on Fig. 4); hence the fixed-budget rebate mechanism remains closer to social optimum than the time-of-day pricing mechanism. This is due to fact that the sensitivity to congestion is "strongly convex", i.e., function $h(\cdot)$ (12) is far from linear. Hence, the optimal unit reward (24c) decreases "fast", as for the fixed-budget rebate mechanism. For the sole purpose of illustrating numerically case *(ii)* of Proposition 5, we construct the following example:

**Example 2.** Everything is defined as in Example 1, but the disutility function is artificially contrived to have $h(G) = 1.2 \cdot 10^{-3} \cdot (G^{0.95} - D_p^{0.95})$. (The factor $1.2 \cdot 10^{-3}$ is chosen to yield the same social optimum level of public good $G^*$ than in Example 1 when $\epsilon = 0$.)

Ex. 2 is a contrived example where $h(\cdot)$ is almost linear so that the optimal unit reward $r_{SO}$ is almost constant, as in the time-of-day pricing mechanism. As a result, the time-of-day pricing mechanism is closer to social optimum (Fig. 3-(b)).

Proposition 5 compares the distance between $G^{(\text{eq})}$ and $G^*$ for the two incentive mechanisms, when utilities are perturbed. It holds for all perturbation directions $\{p_\theta(\cdot)\}_{\theta\in\Theta}$ (satisfying the conditions mentioned above). However, the direction of the variation of $G^{(\text{eq})}$ and $G^*$ depends on the perturbation directions $\{p_\theta(\cdot)\}_{\theta\in\Theta}$. With the simple perturbation where $p_\theta(\cdot) = c_\theta(\cdot)$ for all $\theta \in \Theta$, Fig. 3 shows that $G^{(\text{eq})}$ and $G^*$ decrease when $\epsilon$ increases (this could also be easily derived analytically from the proof of Proposition 5). Intuitively, if users are less willing to contribute due to a high cost of shifting, the equilibrium and social optimal amount of public good will be lower. With a general perturbation, the variation of $G^{(\text{eq})}$ and $G^*$ is determined by the variation of the aggregate best response $G^{(\text{resp})}$ at the point $G^*(0)$ when the utilities perturbation is introduced.

From Proposition 5, we deduce the following result.

**Theorem 3.** *There exists $\epsilon_m > 0$ such that, for any perturbation* (23) *with $\epsilon \neq 0$ and $|\epsilon| < \epsilon_m$, we have:*

*(i) if condition **(C1)** is satisfied, then the fixed-budget rebate mechanism is more robust than the time-of-day pricing mechanism:*

$$W_T^{(eq)}(\epsilon) < W_R^{(eq)}(\epsilon) < W^*(\epsilon);$$

*(ii) if condition **(C2)** is satisfied, then the time-of-day pricing mechanism is more robust than the fixed-budget rebate mechanism:*

$$W_R^{(eq)}(\epsilon) < W_T^{(eq)}(\epsilon) < W^*(\epsilon).$$

Theorem 3 is our main robustness result. It establishes which of the two mechanisms remains closer to optimal after the perturbation, in terms of welfare, i.e., in terms of user

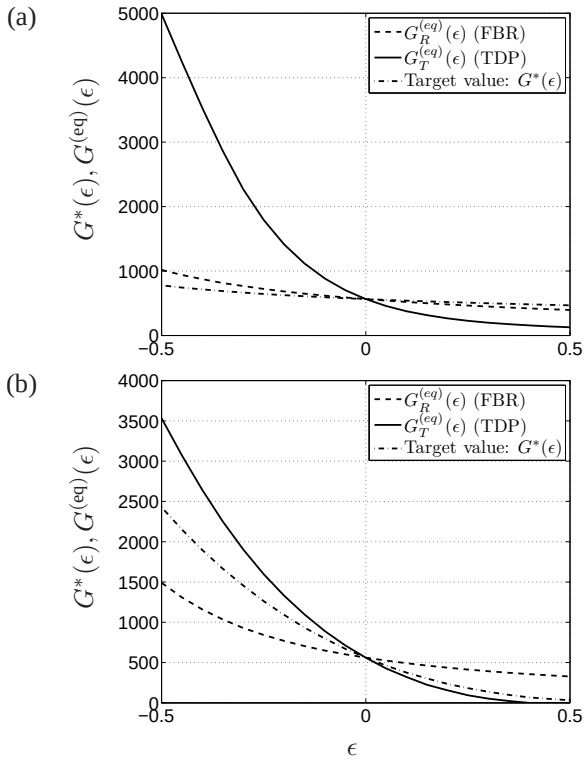


Fig. 3. Variation of the equilibrium amount of public good $G^{(\text{eq})}$ when functions $c_\theta(\cdot)$ are scaled by a factor $(1+\epsilon)$ starting from the baseline case: (a) for Example 1 – (b) for Example 2.

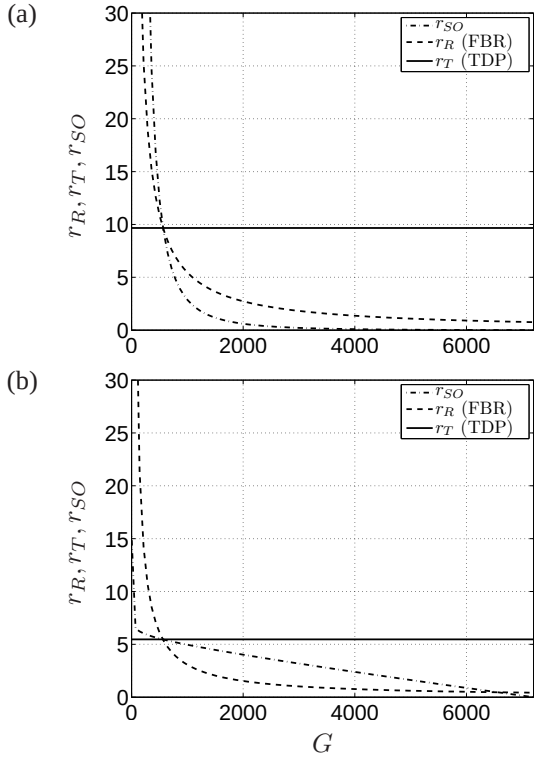


Fig. 4. Unit reward as a function of $G$: (a) for Example 1 – (b) for Example 2.

TABLE II
EQUILIBRIUM FOR EXAMPLE 1, WITH UTILITIES PERTURBED BY A SCALING OF FACTOR $(1+\epsilon)$ WITH $\epsilon = 0.5$. THE PARAMETERS ARE CHOSEN BASED ON UNPERTURBED UTILITIES: $R = \$5,500$ AND $r = \$9$/GBIT (SEE FIG. 1).

| | no mecha. | TDP mecha. | FBR pricing | SO |
|---|---|---|---|---|
| G | 37 Gbits | 127 Gbits | 395 Gbits | 467 Gbits |
| W | $\sim$0 | 61,000 | 75,200 | 75,400 |

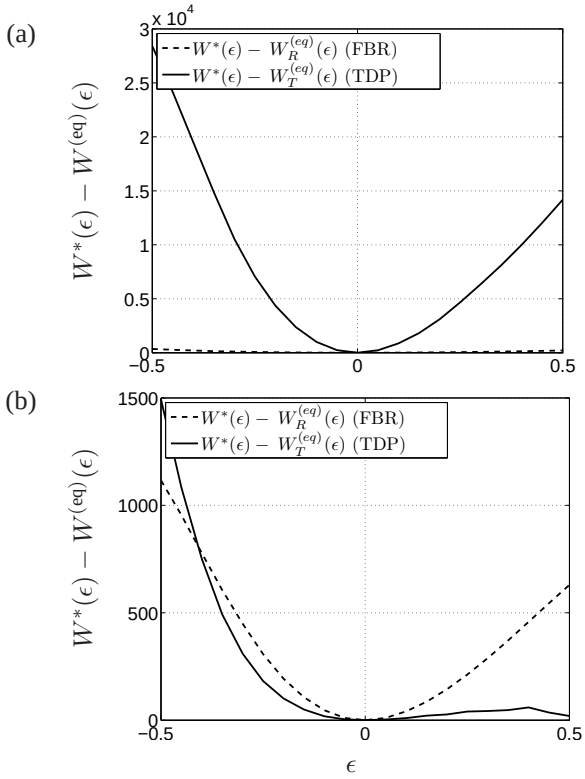


Fig. 5. Variation of the aggregate welfare $W^{(\text{eq})}$ when functions $c_\theta(\cdot)$ are scaled by a factor $(1+\epsilon)$ starting from the baseline case: (a) for Example 1 – (b) for Example 2. For readability and robustness comparison, the difference $W^*(\epsilon) - W^{(\text{eq})}(\epsilon)$ is plotted.

expected utility over the timescale of a month (see Sec. III-B). The conditions of Theorem 3 are the same as in Proposition 5: mechanism $j \in \{R, T\}$ is more robust if its unit reward is closer to the optimal unit reward. Since Example 1 satisfies condition **(C1)** (due to "strong enough" convexity of the sensitivity to congestion), the fixed-budget rebate mechanism is more robust. It means that if the provider chooses the parameters based on an imperfect estimation of the cost of shifting $c_\theta(\cdot)$, the welfare of a population of users whose actual cost of shifting is $\tilde{c}_\theta(\cdot)$ will be higher with the fixed-budget rebate mechanism than with the time-of-day pricing mechanism. Similarly, if the cost of shifting was varying according to a given probability law and parameters $R$ and $r$ were chosen based on expectations, the fixed-budget rebate mechanism would give a higher expected welfare.

Fig. 5 illustrates our robustness results for a simple perturbation. It shows that our analysis with $\epsilon$ close to zero extends to larger perturbations. The numerical values for Example 1 are reported in Tab. II: for a $50\%$ error in the cost-of-shifting estimation ($\epsilon = 0.5$), the welfare is $20\%$ below optimal with the time-of-day pricing mechanism, whereas it is only $0.3\%$ below optimal with the fixed-budget rebate mechanism.

## V. Concluding remarks

This paper provides a comparative analysis of two incentive mechanisms of reducing peak-time congestion in Internet broadband access: a fixed-budget rebate mechanism inspired by the economic literature on public good provision by means of lotteries, and a more standard for the network literature time-of-day pricing mechanism. The fixed-budget rebate mechanism can be interpreted as probabilistic pricing: in this mechanism, each user's reward depends not only on his contribution but also on the contribution of the other users. We suggest that this mechanism has two advantages relative to the time-of-day pricing mechanism with given prices for specific time slots. Firstly, the fixed-budget rebate mechanism is easy to implement via lottery-like scheme(s), for which a total user reward is announced by the ISP in advance. Secondly, it has built-in self-tuning, which appears to be attractive in environments with imperfectly known demand.

Our paper uses a simplified model to provide a theoretical structure that permits to understand the benefits of the fixed-budget rebate mechanism over more standard approaches. The deployment of the mechanism will raise a number of practical questions. In particular, the ISP has to decide at which scale to deploy the mechanism: deploying it at the scale of a base station would involve too precise monitoring whereas deploying it region-wide would face the issue that users do not all share the same access bottleneck. We believe that such decision should be made based on historical statistics on each bottleneck that are accessible to ISPs. Our model also considers only two time periods whereas it could be useful for an ISP to use a finer subdivision of the day. Again, we believe that the number of time periods should be determined using historical data available to the provider. The fixed-budget rebate mechanism could be easily extended to multiple time periods and we believe that it would remain more robust than the time-of-day pricing mechanism.

In our model, we considered a monopolist and the reward was financed by an increase in the subscription price. We showed that both mechanisms still improve each user's average utility. In different scenarios, if the subscription price cannot be increased, it would be possible to finance the reward by a different means, e.g., by the reduction of the congestion cost or by the higher number of customers that the provider could accommodate with the same infrastructure due to lower congestion.

Our model focuses on user welfare maximization rather than on the cost savings for the ISP. However, we believe that both objectives are consistent, as in an competitive environment, increasing user welfare allows either to accommodate more users with the same capacity or to reduce the capacity provisioning costs for the same user base. A quantitative analysis of these questions would require modeling of the cost structure and of the competition (we could typically assume perfect competition) and is left as future work.

To implement in practice the fixed-budget rebate mechanism, an ISP will need to track separately the consumption at peak and off-peak time. However, such separate accounting is also needed for the time-of-day pricing mechanism and is already technically possible in most settings (mobile access, electricity with smart meters, etc.). The use of probabilistic pricing also raises the question of contention billing and verifiability. However, these transparency issues are the same as in many other contexts where they have been successfully solved (state lotteries, casinos, etc.). For instance, inspection techniques similar to those used for gambling [32] could be used here.

While our motivating application is telecommunications, we believe that the fixed-budget rebate mechanism could be modified for use in other applications such as electricity and transportation networks. In the case of electricity demand management, privacy and security considerations make our mechanism advantageous relative to real-time pricing. Indeed, our mechanism requires no real-time user-dispatcher communication. In addition, unlike currently suggested real-time pricing mechanisms (e.g., [33]), our mechanism requires only aggregate data. We plan to explore these other applications in future work.

## References


[1] Cisco Systems, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2010-2015," White Paper. Available at http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_

[2] C. Kang, "Netflix biggest driver of U.S. Internet traffic, puts spotlight on broadband pricing," *Washington Post*, 2011, May 17, available at http://www.washingtonpost.com/blogs/post-tech/post/netflix-biggest-driver-of-us-intern

[3] R. Cheng, "AT&T Pushing To Get Consumers Out Of 'Unlimited' Mindset," *Wall Street Journal*, 2011, March 14, available at http://online.wsj.com/article/BT-CO-20110314-711379.html.

[4] A. Odlyzko, "Internet pricing and the history of communications," *Computer Networks*, vol. 36, pp. 493–517, 2001.

[5] H. Varian, "The Demand for Bandwidth: Evidence from the INDEX Experiment," 2002, Available at http://www.sims.berkeley.edu/~hal/Papers/brookings/brookings.html.

[6] L. Jiang, S. Parekh, and J. Walrand, "Time-dependent network pricing and bandwidth trading," in *Proc. of IEEE NOMS International Workshop on Bandwidth on Demand*, 2008.

[7] C. Wong, S. Ha, and M. Chiang, "Time-dependent broadband pricing: Feasibility and benefits," in *Proc. of ICDCS 2011*, 2011.

[8] P. Loiseau, G. Schwartz, J. Musacchio, S. Amin, and S. S. Sastry, "Congestion pricing using a raffle-based scheme," in *Proc. of NetGCoop*, 2011.

[9] J. Morgan, "Financing public goods by means of lotteries," *Review of Economic Studies*, vol. 67, no. 4, pp. 761–84, October 2000.

[10] T. Henderson, J. Crowcroft, and S. Bhatti, "Congestion pricing: paying your way in communication networks," *IEEE Internet Computing*, vol. 5, no. 5, pp. 85–89, 2001.

[11] B. Tuffin, "Charging the internet without bandwidth reservation: an overview and bibliography of mathematical approaches," *Journal of Information Science and Engineering*, vol. 19, no. 5, pp. 765–786, 2003.

[12] S. Sen, C. Joe-Wong, S. Ha, and M. Chiang, "Incentivizing time-shifting of data: a survey of time-dependent pricing for internet access," *IEEE Communications Magazine*, vol. 50, pp. 91–99, 2012.

[13] M. L. Honig and K. Steiglitz, "Usage-based pricing of packet data generated by a heterogeneous user population," in *Proc. of INFOCOM '95*, 1995, pp. 867–874.

[14] T. Basar and R. Srikant, "Revenue-maximizing pricing and capacity expansion in a many-users regime," in *In proceedings of IEEE INFOCOM*, 2002, pp. 294–301.

[15] H. Mendelson and S. Whang, "Optimal incentive-compatible priority pricing for the M/M/1 queue," *Operations Research*, vol. 38, pp. 870–883, September 1990.

[16] A. Odlyzko, "Paris metro pricing for the internet," in *Proc. of EC '99*, 1999, pp. 140–147.

[17] P. Marbach, "Analysis of a static pricing scheme for priority services," *IEEE/ACM Transactions on Networking*, vol. 12, no. 2, pp. 312–325, Apr. 2004.

[18] H. Shen and T. Başar, "Optimal nonlinear pricing for a monopolistic network service provider with complete and incomplete information," *IEEE Journal on Selected Areas in Communications (JSAC), Special Issue: Non-Cooperative Behavior in Networking*, vol. 25, no. 6, pp. 1216–1223, 2007.

[19] ——, "Pricing under information asymmetry for a large population of users," *Telecommunication Systems*, vol. 47, no. 1-2, pp. 123–136, 2011.

[20] I. C. Paschalidis and J. N. Tsitsiklis, "Congestion-dependent pricing of network services," *IEEE/ACM Transactions on Networking*, vol. 8, no. 2, pp. 171–184, 2000.

[21] J. Londoño, A. Bestavros, and N. Laoutaris, "Trade & cap: a customer-managed, market-based system for trading bandwidth allowances at a shared link," in *Proc. of ACM NetEcon Workshop*, 2010, pp. 7:1–7:7.

[22] S. Ha, S. Sen, C. Joe-Wong, Y. Im, and M. Chiang, "TUBE: Time Dependent Pricing for Mobile Data," in *In proceedings of ACM SIGCOMM*, 2012.

[23] S. Sharma and D. Teneketzis, "An externalities-based decentralized optimal power allocation algorithm for wireless networks," *IEEE/ACM Transactions on Networking*, vol. 17, pp. 1819–1831, December 2009.

[24] ——, "A game-theoretic approach to decentralized optimal power allocation for cellular networks," *Telecommunication Systems*, vol. 47, no. 1-2, pp. 65–80(16), Jun. 2011.

[25] A. Lange, J. A. List, and M. K. Price, "Using lotteries to finance public goods: Theory and experimental evidence*," *International Economic Review*, vol. 48, no. 3, pp. 901–927, 2007.

[26] C. Waldspurger and W. Weihl, "Lottery scheduling: Flexible proportional-share resource management," in *Proc. of USENIX OSDI*, 1994.

[27] D. Merugu, B. Prabhakar, and N. Rama, "An incentive mechanism for decongesting the roads: A pilot program in bangalore," in *Proc. of ACM NetEcon Workshop*, 2009.

[28] R. J. Aumann, "Markets with a continuum of traders," *Econometrica*, vol. 32, no. 1/2, pp. 39–50, 1964.

[29] M. Ali Khan and Y. Sun, "Non-cooperative games with many players," in *Handbook of Game Theory with Economic Applications, Volume 3*, R. Aumann and S. Hart, Eds. Elseiver Science, 2002, pp. 1761–1808.

[30] S. Borst, K. Clarkson, J. Graybeal, H. Viswanathan, and P. Whiting, "User-level QoS and traffic engineering for 3G wireless 1xEV-DO systems," *Bell Labs Technical Journal*, vol. 8, pp. 33–47, 2003.

[31] A. C. Pigou, *The Economics of Welfare*. London: Macmillan and Co., 1932, 4th edition.

[32] New Jersey Department of Law and Public Safety, "Division of gaming enforcement," http://www.nj.gov/oag/ge/index.html.

[33] J.-Y. Le Boudec and D.-C. Tomozei, "Satisfiability of elastic demand in the smart grid," in *ENERGY 2011*, 2011.

[34] E. Zeidler, *Nonlinear Functional Analysis and Its Applications III: Variational Methods and Optimization*. Springer-Verlag, 1985.


## APPENDIX A
## PROPERTIES OF THE COST OF SHIFTING

Let $\theta \in \Theta$. First, we show that the function $x_\theta \mapsto c_\theta(x_\theta)$, defined by (10), is differentiable on $[0, d_p]$. By assumption, the function $x_\theta \mapsto P_\theta(d_p - x_\theta) - (d_p - x_\theta)q$ is differentiable so we only need to show that the function $x_\theta \mapsto O_\theta(z^*_\theta(x_\theta)) - z^*_\theta(x_\theta)q$ at $z_{\theta,max}$ is differentiable on $[0, d_p]$. Since the function $z \mapsto O_\theta(z)$ is twice differentiable increasing strictly concave by assumption, the function $z \mapsto O_\theta(z) - zq$ is twice differentiable strictly concave. Let $z_{\theta,max}$ be its maximum on $[0, d_p]$. Then,

$$z^*_\theta(x_\theta) = \begin{cases} x_\theta \text{ if } x_\theta \leq z_{\theta,max}, \\ z_{\theta,max} \text{ if } x_\theta > z_{\theta,max}. \end{cases}$$

If $z_{\theta,max} = 0$ (hence $z^*_\theta(x_\theta) = z_{\theta,max}, \forall x \in [0, d_p]$), or if $z_{\theta,max} = d_p$ (hence $z^*_\theta(x_\theta) = x_\theta, \forall x \in [0, d_p]$), then $x_\theta \mapsto O_\theta(z^*_\theta(x_\theta)) - z^*_\theta(x_\theta)q$ is clearly differentiable on $[0, d_p]$. If $z_{\theta,max} \in (0, d_p)$, we need to show that $x_\theta \mapsto O_\theta(z^*_\theta(x_\theta)) - z^*_\theta(x_\theta)q$ is differentiable at the point $z_{\theta,max}$. But since $z_{\theta,max}$ is an interior maximum, we have $O_\theta(z_{\theta,max}) - q = 0$, therefore both the left and right derivatives of $x_\theta \mapsto O_\theta(z^*_\theta(x_\theta)) - z^*_\theta(x_\theta)q$ at $z_{\theta,max}$ are zero.

Next, to show that the function $x_\theta \mapsto c_\theta(x_\theta)$ is strictly convex on $[0, d_p]$, we show that its derivative is increasing. By assumption that $P_\theta(\cdot)$ is strictly concave, the derivative of $x_\theta \mapsto -[P_\theta(d_p - x_\theta) - (d_p - x_\theta)q]$ is increasing. Moreover, the derivative of $x_\theta \mapsto -[O_\theta(z^*_\theta(x_\theta)) - z^*_\theta(x_\theta)q]$ is increasing on $[0, z_{\theta,max})$ and constant on $[z_{\theta,max}, d_p]$. Therefore the derivative of $x_\theta \mapsto c_\theta(x_\theta)$ is increasing.

Last, we show that $x_\theta \mapsto c_\theta(x_\theta)$ is increasing on $[\underline{x}_\theta, d_p]$. If $\underline{x}_\theta = d_p$, the result is trivial. If $\underline{x}_\theta < d_p$, by definition of $\underline{x}_\theta$, $c'_\theta(\underline{x}_\theta) \geq 0$. Hence, since we have shown that $c'_\theta(\cdot)$ is increasing, we have $c'_\theta(x_\theta) > 0, \forall x \in (\underline{x}_\theta, d_p]$ which gives the result.

## APPENDIX B
## PROOF OF THEOREM 1

Let $R$ and $r$ be fixed and let $j \in \{R, T\}$. For a given $G \in [0, D_p]$, the best response (solving the FOCs (20)) defines a measurable function $x^{(\text{resp})}(G) : \Theta \to [0, d_p]$ given for all $\theta \in \Theta$ by

$$x^{(\text{resp})}_\theta(G) = \begin{cases} 0, \text{ if } M^{j\prime}(G) - h(G) - c'_\theta(0) \leq 0, \\ d_p, \text{ if } M^{j\prime}(G) - h(G) - c'_\theta(d_p) \geq 0, \\ (c'_\theta)^{-1}\left(M^{j\prime}(G) - h(G)\right), \text{ otherwise.} \end{cases} \quad (25)$$

Due to assumption **(A2)**, $c'_\theta(\cdot)$ is strictly increasing, hence invertible and with an increasing inverse function. Therefore, (25) uniquely defines $x^{(\text{resp})}(G)$. Let

$$G^{(\text{resp})}(G) = \int_\Theta x^{(\text{resp})}_\theta(G) \mathrm{d}\mu(\theta) \quad (26)$$

be the aggregate best response. By definition and strict concavity of the utility function, a measurable function $x : \Theta \to [0, d_p]$ is a Nash equilibrium if and only if there exists $G \in [0, D_p]$ satisfying the fixed-point equation (21) such that $x = x^{(\text{resp})}(G)$. To conclude the proof, we show that (21) admits a unique fixed-point (see illustration on Fig. 1).

**Lemma 1.** *There exists a unique solution of* (21).

*Proof:* The r.h.s. of (21) ($G$) is clearly a strictly increasing continuous function of $G$, from $[0, D_p]$ to $[0, D_p]$.

For the l.h.s. ($G^{(\text{resp})}(G)$), firstly note that it is a continuous function of $G$. Indeed, due to assumption **(A2)**, $(c'_\theta(\cdot))^{-1}(\cdot)$ is strictly increasing continuous, hence $x^{(\text{resp})}_\theta(G)$ is continuous in $G$ for all $\theta \in \Theta$. Moreover, the function $x^{(\text{resp})}(G) : \Theta \to [0, d_p]$ is dominated by the constant function equal to $d_p$ (i.e., $|x^{(\text{resp})}(G)| \leq d_p$) which is integrable w.r.t. $\mu$. Therefore, for any $G \in [0, D_p]$ and for any sequence $(G_n)_{n\geq 0}$ which converges to $G$, we have $x^{(\text{resp})}(G_n) \xrightarrow[n\to\infty]{} x^{(\text{resp})}(G)$ pointwise (by continuity of $x^{(\text{resp})}_\theta(G)$ w.r.t to $G$ for all $\theta \in \Theta$) and by Lebesgue dominated convergence theorem,

$$\begin{aligned} \lim_{n\to\infty} G^{(\text{resp})}(G_n) &= \lim_{n\to\infty} \int_\Theta x^{(\text{resp})}_\theta(G_n) \mathrm{d}\mu(\theta) \\ &= \int_\Theta \left[\lim_{n\to\infty} x^{(\text{resp})}_\theta(G_n)\right] \mathrm{d}\mu(\theta) \\ &= G^{(\text{resp})}(G). \end{aligned}$$

Clearly, the l.h.s. ($G^{(\text{resp})}(G)$) is also a non-increasing function of $G$ taking values in $[0, D_p]$. Therefore, there is a unique fixed-point of (21). ∎

## APPENDIX C
## PROOF OF THEOREM 2

We first prove *(i)*. Let $\mathcal{X}$ be the set of functions $x:\Theta\to\mathbb{R}$ such that $\int_\Theta |x_\theta|\,\mathrm{d}\mu(\theta)<\infty$ and let $\mathcal{X}_0\subset\mathcal{X}$ be the set of functions $x:\Theta\to[0,d_p]$. Consider the aggregate welfare (5) as a functional on $\mathcal{X}_0$ taking values in $\mathbb{R}$:

$$W(x) = h\left(\int_\Theta x_\theta \mathrm{d}\mu(\theta)\right)\left(D_p-\int_\Theta x_\theta \mathrm{d}\mu(\theta)\right) - \int_\Theta c_\theta(x_\theta)\mathrm{d}\mu(\theta)+\int_\Theta \bar{u}_\theta \mathrm{d}\mu(\theta)-p\frac{D_p}{d_p}.$$

Since $\mathcal{X}_0$ is compact and the functional $W$ is continuous, it has a maximum (see Corollary 38.10 of [34, p. 152]). Let $x^*\in\mathcal{X}_0$ be such that $W$ is maximal and let

$$G^*=\int_\Theta x^*_\theta \mathrm{d}\mu(\theta).$$

Define the three subsets of $\Theta$: $\Theta_1$, $\Theta_2$ and $\Theta_3$ where $x^*=0$, $x^*\in(0,d_p)$ and $x^*=d_p$ respectively. We now derive necessary conditions for $x^*$ to maximize $W$ in each subset.

We start with the subset $\Theta_2$ corresponding to interior points. Let $y\in\mathcal{X}$ be such that $y_\theta=0$ for all $\theta\in\Theta\backslash\Theta_2$. We define the directional derivative (also called Gâteaux derivative) of $W$ around $x^*$ in the direction $y$ as

$$\mathrm{d}W(x^*,y)=\lim_{t\to0}\frac{W(x^*+ty)-W(x^*)}{t}.$$

Then, we have

$$\mathrm{d}W(x^*,y)=\int_{\Theta_2} y_\theta\cdot[h'(G^*)(D_p-G^*)-h(G^*)-c'_\theta(x^*_\theta)]\,\mathrm{d}\mu(\theta),$$

where the exchange between limit and integration in the last term (giving $-y_\theta c'_\theta(x_\theta)$) is justified by Lebesgue's dominated convergence theorem whenever $\int_\Theta |y_\theta\cdot c'_\theta(x_\theta)|\,\mathrm{d}\mu(\theta)<\infty$. This holds here due to assumption **(A3)**.

For $x^*$ to be optimal, it is necessary that $\mathrm{d}W(x^*,y)=0$, i.e.,

$$\int_{\Theta_2} y_\theta\cdot[h'(G^*)(D_p-G^*)-h(G^*)-c'_\theta(x^*_\theta)]\,\mathrm{d}\mu(\theta)=0.$$

For this to hold for any function $y$ such that $y_\theta=0$ for all $\theta\in\Theta\backslash\Theta_2$, it is necessary that we have

$$h'(G^*)(D_p-G^*)-h(G^*)-c'_\theta(x^*_\theta)=0, \qquad (27)$$

for almost-all $\theta\in\Theta_2$, i.e., for almost-all $\theta$ such that $x^*_\theta\in(0,d_p)$.

We now treat the case of subset $\Theta_1$, which corresponds the points of the lower boundary. Let $y\in\mathcal{X}$ be such that $y_\theta\geq0$ for all $\theta\in\Theta_1$ and $y_\theta=0$ for all $\theta\in\Theta\backslash\Theta_1$; that is $y$ is a direction that "pushes up" the values of $x^*$ that are at zero. The directional derivative of $W$ around $x^*$ in the direction $y$ is defined similarly to the previous case but with a limit $t>0$:

$$\mathrm{d}W(x^*,y)=\lim_{t\to0^+}\frac{W(x^*+ty)-W(x^*)}{t},$$

which gives

$$\mathrm{d}W(x^*,y)=\int_{\Theta_1} y_\theta\cdot[h'(G^*)(D_p-G^*)-h(G^*)-c'_\theta(x^*_\theta)]\,\mathrm{d}\mu(\theta).$$

For $x^*$ to be optimal, it is necessary that $\mathrm{d}W(x^*,y)\leq0$, i.e.,

$$\int_{\Theta_1} y_\theta\cdot[h'(G^*)(D_p-G^*)-h(G^*)-c'_\theta(x^*_\theta)]\,\mathrm{d}\mu(\theta)\leq0.$$

For this to hold for any function $y$ such that $y_\theta\geq0$ for all $\theta\in\Theta_1$ and $y_\theta=0$ for all $\theta\in\Theta\backslash\Theta_1$, it is necessary that

$$h'(G^*)(D_p-G^*)-h(G^*)-c'_\theta(x^*_\theta)\leq0, \qquad (28)$$

for almost-all $\theta\in\Theta_1$, that is for almost all $\theta$ such that $x^*_\theta=0$.

The case of subset $\Theta_3$ is handled similarly and yields the necessary condition:

$$h'(G^*)(D_p-G^*)-h(G^*)-c'_\theta(x^*_\theta)\geq0, \qquad (29)$$

for almost all $\theta$ such that $x^*_\theta=d_p$.

In summary, (27)-(29) show that for function $x^*$ to maximize $W$, it is necessary that $x^*$ is solution of the FOCs (20) where $M^{j'}(G)$ is replaced by $h'(G)(D-G)$. By assumption **(A1)**, this is a decreasing function of $G$. Therefore the same proof as for Theorem 1 shows that $x^*$ is uniquely determined almost-everywhere.

We now prove *(ii)*. From the proof of part *(i)*, it is clear that if $R=R^*$ (resp. $r=r^*$), then the FOCs (20) at a Nash equilibrium coincide with the optimality condition, which gives the result.

## APPENDIX D
## PROOF OF PROPOSITION 1

We provide the proof for the fixed-budget rebate mechanism ($j=R$). It is the same for the time-of-day pricing mechanism.

We first prove *(i)* using three cases.

*Case 1:* If $G^{(\mathrm{eq})}(R)=0$, then the result is obvious.

*Case 2:* If $G^{(\mathrm{eq})}(R)=D_p$, then we have $R/D_p-h(D_p)-c'_\theta(d_p)\geq0$ for almost-all $\theta\in\Theta$, which implies $R'/D_p-h(D_p)-c'_\theta(d_p)\geq0$. Hence $G^{(\mathrm{eq})}(R')=D_p$.

*Case 3:* If $G^{(\mathrm{eq})}(R)\in(0,D_p)$. For a given $G$, $x^{(\mathrm{resp})}(G)$ of (25) is non-decreasing when $R$ increases to $R'$, and strictly increasing for $\theta$'s s.t. $x^{(\mathrm{resp})}(G^{(\mathrm{eq})}(R))\in(0,D_p)$. Since the set of such $\theta$'s is of positive measure, the new fixed-point has $G^{(\mathrm{eq})}(R')>G^{(\mathrm{eq})}(R)$.

From *(i)*, *(ii)* follows clearly.

We finally prove *(iii)*. The existence of the threshold $\bar{R}$ follows from the fact that $M^{R'}(D_p)\xrightarrow[R\to\infty]{}\infty$, hence $M^{R'}(D_p)-h(D_p)-c'_\theta(d_p)\geq0$ for almost-all $\theta\in\Theta$ beyond $\bar{R}$ (due to assumption **(A3)**).

## APPENDIX E
## PROOF OF PROPOSITION 2

For the fixed-budget rebate mechanism with $R>0$, if $G=0$, the unit reward is infinite, hence each user wants to contribute positively. Therefore, $G=0$ is not an equilibrium.

## APPENDIX F
## PROOF OF PROPOSITION 3

We provide the proof for the fixed-budget rebate mechanism ($j=R$). It is the same for the time-of-day pricing mechanism.

Let $R\geq0$ and denote for simplicity $x=x^{(\mathrm{eq})}(R)$ and $G=G^{(\mathrm{eq})}(R)$. Using the notation of the proof of Theorem 2 (Appendix C), the derivative of the welfare around $x$ in the direction $y\in\mathcal{X}$ is

$$\mathrm{d}W(x,y)=\int_\Theta y_\theta\cdot[h'(G)(D_p-G)-h(G)-c'_\theta(x_\theta)]\,\mathrm{d}\mu(\theta).$$

Suppose first that $R<R^*$ and consider the direction $y\in\mathcal{X}$ such that

$$y_\theta=\frac{\mathrm{d}x^{(\mathrm{eq})}_\theta(R)}{\mathrm{d}R},\quad\forall\theta\in\Theta.$$

By Proposition 1, we have $G^{(\text{eq})}(R) \leq G^*$ and $x_\theta^{(\text{eq})}(R) \leq x_\theta^*$ for almost-all $\theta \in \Theta$. Therefore, we have $y_\theta \geq 0$ for all $\theta \in \Theta_2 \cup \Theta_3$ and $y_\theta = 0$ for $\theta \in \Theta_1$ (recall that $\Theta_1$, $\Theta_2$ and $\Theta_3$ are the subsets of $\Theta$ where $x^* = 0$, $x^* \in (0, d_p)$ and $x^* = d_p$ respectively). Moreover, we have $h'(G)(D_p - G) - h(G) - c'_\theta(x_\theta) \geq 0$ for all $\theta \in \Theta_2 \cup \Theta_3$. Indeed, in $\Theta_2$, $h'(G^*)(D_p - G^*) - h(G^*) - c'_\theta(x_\theta^*) = 0$ and $h'(G)(D_p - G) - h(G) \geq h'(G^*)(D_p - G^*) - h(G^*)$ (by concavity of $h(G)(D_p - G)$) and $c'_\theta(x_\theta) \leq c'_\theta(x_\theta^*)$. In $\Theta_3$, $h'(G^*)(D_p - G^*) - h(G^*) - c'_\theta(x_\theta^*) \geq 0$ and $h'(G)(D_p - G) - h(G) \geq h'(G^*)(D_p - G^*) - h(G^*)$ and $c'_\theta(x_\theta) = c'_\theta(x_\theta^*)$. We conclude that $\mathrm{d}W(x, y) \geq 0$, therefore $W^{(\text{eq})}(R)$ increases with $R$.

The case $R \in (R^*, \bar{R})$ is handled similarly and if $R \geq \bar{R}$, the equilibrium does not vary with $R$ by Proposition 1-*(iii)*.

## Appendix G
## Proof of Proposition 5

### A. Proof of the Proposition

From the proof of Theorem 1, we know that for any mechanism $j \in \{R, T\}$, $G_j^{(\text{eq})}$ is the fixed-point solution of (21). Here, we explicitly write the dependence in the mechanism, i.e., for mechanism $j$, we denote by $x_{j,\theta}^{(\text{resp})}(G)$ the individual best response (25), and by $G_j^{(\text{resp})}(G)$ the aggregate best response (26) to a given $G \in [0, D_p]$. We use the notation $r_j$ (rather than $M^{j'}$) for the unit reward (see (24)). From the proof of Theorem 2, we know that $G^*$ is found as the fixed-point solution of the same equation (21) with $r_j(G) = r_{SO}(G)$ defined by (24c). Therefore, we will use the notation $G^* = G_{SO}^{(\text{eq})}$ which emphasizes this similarity and helps shorten the proof's notation.

Before evaluating the variations of the equilibrium with $\epsilon$, note that when $\epsilon = 0$ (baseline), we have the same equilibrium:

$$G_R^{(\text{eq})}(0) = G_T^{(\text{eq})}(0) = G_{SO}^{(\text{eq})}(0), \tag{30}$$

and the same unit rewards at equilibrium:

$$r_R(G) = r_T(G) = r_{SO}(G) \quad \text{if} \quad G = G_R^{(\text{eq})}(0). \tag{31}$$

When $\epsilon \neq 0$, functions $c_\theta$ are perturbed. For a given $G \in [0, D_p]$, the aggregate best response is modified accordingly. We denote by $G_{j,\epsilon}^{(\text{resp})}(G)$ the new aggregate best response. Recall that we also denote by $G_j^{(\text{eq})}(\epsilon)$ the new equilibrium point which is the fixed point of $G_{j,\epsilon}^{(\text{resp})}(\cdot)$. The following lemma readily implies the result of Proposition 5.

**Lemma 2.** *For any $j \in \{R, T, SO\}$, we have*

$$G_j^{(eq)}(\epsilon) = G_j^{(eq)}(0) + \frac{J_\epsilon}{1 + \alpha_j} + o(\epsilon), \tag{32}$$

*where*

$$J_\epsilon = G_{j,\epsilon}^{(resp)}\left(G_R^{(eq)}(0)\right) - G_{j,0}^{(resp)}\left(G_R^{(eq)}(0)\right)$$

*is a first-order quantity in $\epsilon$ independent of the mechanism $j$, and*

$$\alpha_j = -\frac{dG_j^{(resp)}}{dG}\left(G_R^{(eq)}(0)\right). \tag{33}$$

*Proof:* First note that $G_{j,\epsilon}^{(\text{resp})}\left(G_R^{(\text{eq})}(0)\right)$ is continuously differentiable with respect to $\epsilon$, and by assumption, the first-order term in $J_\epsilon$ is non-zero. Moreover, due to (31), $G_{j,\epsilon}^{(\text{resp})}\left(G_R^{(\text{eq})}(0)\right)$ is independent of the mechanism $j$ (see (26) and (25)), and so is $J_\epsilon$.

Starting from the point $\left(G_R^{(\text{eq})}(0), G_{j,\epsilon}^{(\text{resp})}\left(G_R^{(\text{eq})}(0)\right)\right)$, at the first order, $G_{j,\epsilon}^{(\text{resp})}(G)$ decreases linearly when $G$ increases. Therefore it can be seen geometrically that it will cross again the first bisector at the new equilibrium point

$$G_j^{(\text{eq})}(\epsilon) = G_j^{(\text{eq})}(1) + \frac{J_\epsilon}{1 + \alpha_{j,\epsilon}} + o(\epsilon), \tag{34}$$

where $-\alpha_{j,\epsilon}$ is the slope of the curve $G_{j,\epsilon}^{(\text{resp})}(G)$ at $G = G_R^{(\text{eq})}(0)$, i.e.,

$$\alpha_{j,\epsilon} = -\frac{\mathrm{d}G_{j,\epsilon}^{(\text{resp})}}{\mathrm{d}G}\left(G_R^{(\text{eq})}(0)\right). \tag{35}$$

From (34), it is easy to see that since $J_\epsilon$ is first-order in $\epsilon$, the first-order term in the Taylor series of $\alpha_{j,\epsilon}$ will give a second-order term in the Taylor series of $G_j^{(\text{eq})}(\epsilon)$. Therefore, we can restrict the series of $\alpha_{j,\epsilon}$ (35) at the order zero: $\alpha_{i,\epsilon} = \alpha_j + o(1)$, which directly gives the desired result (32). ∎

### B. Reduction of *(C1-2)* to *(C1′-2′)*

We have

$$\begin{aligned}\frac{\mathrm{d}G_j^{(\text{resp})}}{\mathrm{d}G}(G) &= \frac{\mathrm{d}}{\mathrm{d}G}\left[\int_\Theta x_{j,\theta}^{(\text{resp})}(G)\mathrm{d}\mu(\theta)\right], \\ &= \int_\Theta \frac{\mathrm{d}}{\mathrm{d}G}\left[x_{j,\theta}^{(\text{resp})}(G)\right]\mathrm{d}\mu(\theta), \\ &= A_j(G) \cdot \left(r'_j(G) - h'(G)\right),\end{aligned}$$

where

$$A_j(G) = \int_{\Theta_{2,j}(G)} \left((c'_\theta)^{-1}\right)' \left(r_j(G) - h(G)\right)\mathrm{d}\mu(\theta), \tag{36}$$

and $\Theta_{2,j}(G)$ is the subset of $\theta$'s for which $x_{j,\theta}^{(\text{resp})}(G) \in (0, d_p)$.

At $G = G_R^{(\text{eq})}(0)$, $\Theta_{2,j}(G)$ is independent of the mechanism $j$, so that (31) shows that $A_j(G_R^{(\text{eq})}(0))$ is independent of the mechanism $j$. Denoting by $A = A_j(G_R^{(\text{eq})}(0))$ the common value, we have for all $j \in \{R, T, SO\}$

$$\alpha_j = A \cdot \left(r'_j\left(G_R^{(\text{eq})}(0)\right) - h'\left(G_R^{(\text{eq})}(0)\right)\right).$$

If we assume that for $j_1 \neq j_2$, $|\alpha_{j_1} - \alpha_{j_2}|$ is small, then

$$\begin{aligned}\left|\frac{1}{1 + \alpha_{j_1}} - \frac{1}{1 + \alpha_{j_2}}\right| &= \left|\frac{\alpha_{j_2} - \alpha_{j_1}}{1 + \alpha_{j_1}}\right| + o(|\alpha_{j_1} - \alpha_{j_2}|), \\ &= \frac{A}{1 + \alpha_{j_1}}\left|r'_{j_2} - r'_{j_1}\right| + o(\left|r'_{j_2} - r'_{j_1}\right|).\end{aligned}$$

With this, conditions **(C1′-2′)** are easily deduced from **(C1-2)**.

## Appendix H
## Proof of Theorem 3

We consider the aggregate welfare (5) as a function of $G$: $W(G) = W(x^{(\text{resp})}(G))$. We have $\frac{\mathrm{d}W}{\mathrm{d}G}(G^*(\epsilon)) = 0$. The result of Theorem 3 is then deduced from Proposition 5 using a taylor expansion around $G^*(\epsilon)$: for $j \in \{R, T\}$,

$$W(G_j^{(\text{eq})}(\epsilon)) = W(G^*(\epsilon)) + O\left((G_j^{(\text{eq})}(\epsilon) - G^*(\epsilon))^2\right).$$