Victor Naroditskiy, Richard Steinberg

Maximizing social welfare in congestion games via redistribution

Article (Accepted version) (Refereed)

Maximizing Social Welfare in Congestion Games via Redistribution

Victor Naroditskiy

University of Southampton

Richard Steinberg

London School of Economics and Political Science

Abstract

It is well-known that efficient use of congestible resources can be achieved via marginal pricing; however, payments collected from the agents generate a budget surplus, which reduces social welfare. We show that an asymptotically first-best solution in the number of agents can be achieved by the appropriate redistribution of the budget surplus back to the agents.

1. Introduction

Congestion games model situations in which multiple agents use shared resources, where each agent's value of a resource decreases with the total usage of the resource or, equivalently, the corresponding level of congestion, a negative externality. Thus, the higher the level of congestion, the less valuable the resource is to an agent. From the viewpoint of social welfare, the best use of resources occurs when the sum of the agents' values is maximized, i.e., when the use of resources is efficient.

In order to direct self-interested agents towards the efficient use of resources, one can resort to pricing. Specifically, requiring each agent to pay for the corresponding disutility she imposes on others[1] results in an efficient use of the resources (see, e.g., MacKie-Mason and Varian (1995); Kelly (1997)). While such pricing will maximize the agents' total value, each agent who

[1] These payments are known as *Pigouvian taxes* (Pigou, 1920), *marginal cost prices*, or *congestion prices*.

makes a payment will suffer a reduction in utility. In some contexts, the collected revenue is desirable, as it increases the utility of the party collecting it (e.g., the seller). However, in many congestion scenarios the main objective is the welfare of the agents, which is decreased by any payments collected (Cole et al., 2006). Indeed, congestion scenarios often arise in settings where the resources are intended for public use and not for the generation of revenue. In this work we ask: How can social welfare be maximized in congestion games? More specifically: *How can most of the revenue be redistributed back to the agents while ensuring that an efficient allocation is achieved?*

In the first part of the paper, after observing that an atomic congestion game can be modeled as the allocation of multiple copies of heterogeneous items, we show that as the number of agents increases, all of the revenue can be redistributed asymptotically while still achieving the efficient allocation. Specifically, we prove that a redistribution rule designed for non-congestion models by Bailey (1997) and generalized by Cavallo (2006) asymptotically achieves full budget balance in the presence of congestion. Thus, we identify a first-best solution to the problem of welfare maximization in atomic congestion games. It is interesting to observe that while in non-congestion settings (e.g., allocating multiple copies for an identical item to agents with unit demand) the revenue redistributed by Bailey-Cavallo can be arbitrarily low, we show here that in congestion settings (asymptotically) all of the revenue is redistributed.

In the atomic context, we also clarify the relationship between congestion prices and VCG payments by proving that, as the efficient level of congestion increases, the two asymptotically approach each other. Intuitively, both congestion prices and VCG payments charge the agent for the "externality" she imposes on others, and therefore, the connection between them is not entirely surprising. However, we have not seen a formal analysis of the relationship between the prices, while the two are normally treated separately (see, e.g., Sections 9.1 and 9.5 in Courcoubetis and Weber (2003)).

In the last section of the paper prior to the Discussion, we turn to nonatomic congestion games. We show that revenue can be redistributed to the agents in equal shares, resulting in a first-best solution. As the effect of an individual agent on the total congestion in an atomic congestion game becomes negligibly small, the redistributed amounts for atomic and nonatomic models coincide.

We now place our work within the existing literature. Revenue redistribution has received considerable attention in the work on mechanism design, mostly in the context of dominant-strategy implementation in allocation models. Specifically, the agenda of welfare maximization in models without a residual claimant has been pursued in the allocation of identical items (Moulin, 2009; Guo and Conitzer, 2008, 2009; de Clippel et al., 2014), the allocation of non-identical items (Guo, 2012), in public good settings (Naroditskiy et al., 2012), and for the general application of redistribution (Cavallo, 2006). However, the revenue redistribution literature has thus far not considered scenarios with congestion, even though redistribution is important there. Indeed, many congestion scenarios are characterized by the lack of a residual claimant (as argued above), and high amounts of revenue collected before redistribution (as argued below). In this paper, we apply the rule from (Cavallo, 2006) to redistribute revenue in a model with congestion.

Following this revenue redistribution literature, our results on atomic congestion games are derived in the context of centralized mechanisms. Congestion games are usually considered in a decentralized context. However, centralized mechanisms have been applied to the study of congestion games, e.g., in the context of computational complexity (Chakrabarty et al., 2005; Blumrosen and Dobzinski, 2007). In more detail, Blumrosen and Dobzinski (2007) derive computational complexity of finding welfare-maximizing use of resources. The model they adopt is also different. In our atomic case, agents have combinatorial preferences over the resources but share the same congestion function. In their case, congestion functions are player specific, and agents have non-combinatorial valuations over resources.

Well-studied in the field of congestion pricing are routing and traffic equilibria models (see, e.g., Roughgarden (2005)). In these models agents are to be routed along the edges of a network with each agent associated with a sink node and a source node. Edges on the network differ in their capacity to carry traffic as reflected by edge-specific *congestion functions*. A congestion function specifies the cost each agent routed along the edge experiences. The objective is to induce a congestion-minimizing flow when agents make their routing choices selfishly. However, the agents may arrive at a non-efficient Nash equilibrium. A seminal text (Beckmann et al., 1956) shows how edges can be priced in order to induce the efficient flow in Nash equilibrium. Specifically, each edge is associated with the *congestion price* equal to the marginal cost incurred by the agents using the edge.

The question of welfare-maximization in congestion domains has been considered before. Cole et al. (2006) investigated how congestion prices can be modified when no redistribution is possible. Redistribution of revenue was studied by Adler and Cetin (2001) within a model that had been introduced by Vickrey (1969). In Vickrey's model, congestion is manifested in waiting times to use a resource. Agents form a queue, and the waiting time of each is given by their position in the queue. Prices can be used to eliminate waiting times by encouraging each agent to deviate from her preferred departure time. Thus, agents have different costs. Deviation from the ideal departure time can be interpreted as the level of service provided—the greater the deviation, the lower the level of service. Payments of agents receiving a higher level of service can be redistributed to agents receiving a lower level of service. In contrast, in this paper we focus on a model where each allocated agent obtains the same level of service, i.e., experiences the same level of congestion, which is the case in congestion games.

The remainder of the paper is structured as follows. In Section 2 we show that a first-best solution to welfare-maximization in atomic congestion games is provided by applying the Bailey-Cavallo redistribution rule. We derive the result for the single-resource congestion game first, and then for general congestion games. This section also shows that congestion prices and VCG payments in congestion settings are fundamentally similar. Section 3 talks about nonatomic routing games. There we prove that revenue can be redistributed in equal shares. Concluding remarks appear in Section 4.

2. Centralized Solution: Atomic Congestion Games

This section focuses on games where the number of players is finite and each may have a non-negligible effect on congestion. We describe how congestion effects from these games can be represented in an allocation model, and apply a redistribution rule that was designed for allocation settings.

The mapping between congestion and allocation models will make it clear that the impossibility result regarding fully budget-balanced efficient mechanisms (Green and Laffont, 1977; Holmstrom, 1979) apply to our model.

Both allocation models and congestion models are concerned with allocating scarce resources. A difference is that in most allocation models an agent's utility depends only on what is allocated to him, while in congestion models an agent's utility also depends on how many other agents use the same resources. However, we can view each congestible resource as an item, multiple copies of which can be allocated. Each copy has a cost representing the total decrease in utility experienced by the agents due to sharing the resource with one more agent. We illustrate this in Section 2.1.

2.1. Single-Resource Atomic Congestion Game

We begin the study of welfare maximization in congestion settings by first considering the model of a single congestible resource. We would like to apply a redistribution rule designed for allocation settings, and begin by transforming the congestion problem into an allocation problem. We then evaluate the performance of the Bailey-Cavallo redistribution rule (Bailey, 1997; Cavallo, 2006) and prove that it is asymptotically optimal in the number agents. Finally, we extend the results to general multi-resource congestion games.

In the single-resource atomic routing model, n agents would like to travel along a single link. Congestion on the link when k agents are routed is denoted by the *individual congestion function* $g(k)$, which measures the congestion cost experienced by an agent using the link. Each agent has a private value $v_i \in \mathbb{R}$ for traveling along the link. Monetary transfers are allowed, and agents' utilities are quasi-linear.

To transform the congestion problem into an allocation problem, we first consider the allocation model where k identical items are allocated among n agents. The utility of an agent is

$$u_i(v; f, t) \;=\; f_i(v)\, v_i - t_i(v)$$

where $f_i(v) \in \{0, 1\}$ determines the allocation of agent i and $t_i(v) \in \mathbb{R}$ denotes the payment. The allocation function must satisfy the allocation constraint, $\sum_i f_i(v) \leq k$.

We modify the allocation model to include congestion. To do this, we remove the allocation constraint and express the agent's utility as

$$u_i(v) \;=\; f_i(v)\,[v_i - g(F)] - t_i(v) \quad \text{where } F = \sum_i f_i(v).$$

We make the standard assumption that g is convex (see, e.g., MacKie-Mason and Varian (1995) or Courcoubetis and Weber (2003) Chapter 9.1). Here

$$G(k) \;=\; k\, g(k)$$

measures the *total congestion cost*.

We consider the class of *efficient mechanisms*: i.e., mechanisms where the allocation rule maximizes the sum of the values of the allocated agents taking congestion into account, i.e.,

$$\max_f \; \sum_i f_i(v)\,[v_i - g(F)]. \tag{1}$$

The socially optimal or *efficient level of congestion* is easy to characterize: sorting the agents in decreasing order of values ($v_1 \geq \ldots \geq v_n$), the efficient allocation is $f_i(v)=1$ for all $i \leq m(v)$, where

$$m(v) \;=\; \arg\max_i \,\{v_i - g(i) \;>\; (i-1)\,[g(i) - g(i-1)]\}. \qquad (2)$$

To ease notation, we will often write m instead of $m(v)$ to denote the efficient level of congestion when agents' values are given by v.

The only mechanisms that implement the efficient level of congestion in dominant strategies are the Groves mechanisms (see, e.g., Mas-Colell et al. (1995)), which include the VCG mechanism. Under the VCG mechanism, unallocated agents pay nothing, while allocated agents pay

$$t_i^{\text{vcg}}(v) \;=\; \max\,\{v_{m+1} - g(m),\; (m-1)\,[g(m) - g(m-1)]\}. \qquad (3)$$

VCG payments have an intuitive explanation. The mechanism charges each allocated agent $i \leq m$ the "externality" her presence imposes on others. This externality is manifested in either forcing one agent out of the allocation (in which case, the level of congestion does not change) or in having an additional unit of traffic allocated (increasing the congestion experienced by the other agents). In the former case, agent i takes an item away from agent $m+1$ (this is the agent who would have been allocated had agent $i \leq m$ not been there). The externality is the loss of utility $v_{m+1} - g(m)$ incurred by agent $m+1$. Here, the maximum in (3) resolves to the first term, as the requirement for agent $m+1$ to be allocated in absence of agent i is $v_{m+1} - g(m) \geq (m-1)\,[g(m) - g(m-1)]$. In the latter case, only $m-1$ agents are allocated when agent i is not there, and her presence decreases their utility by the total of $(m-1)\,[g(m) - g(m-1)]$.

All mechanisms within the Groves class can be described through VCG payments together with a *redistribution* function $h(v_{-i})$, where v_{-i} refers to the vector v with the ith component removed. Under a Groves mechanism,[2] agent i pays

$$t_i(v) \;=\; t_i^{\text{vcg}}(v) \;+\; h(v_{-i}). \qquad (4)$$

[2]In fact, Groves mechanisms allow for non-anonymous payment functions: a different function h_i can be specified for each agent. However, the mechanism considered in this paper uses the same function h for each agent.

Definition 1. Grove mechanisms *are specified by an efficient allocation function (i.e., the efficient level of congestion) and payments of the form* (4).

Each Groves mechanism corresponds to a different redistribution function h. Our goal is to find h that maximizes social welfare. Since social welfare is reduced by the payments, we are looking for the function h that redistributes as much of the VCG revenue as possible.

The VCG revenue is the sum of the VCG payments

$$R(v) = \sum_i t_i^{\text{vcg}}(v). \tag{5}$$

Observing that the VCG payment of each allocated agent is the same, the revenue can be expressed as

$$R(v) = m \max \{v_{m+1} - g(m),\ (m-1)\,[g(m) - g(m-1)]\}\,. \tag{6}$$

The higher the traffic, the higher the VCG payments, and the more revenue that is collected. This revenue directly reduces the social welfare. As we show next, for some valuations of the agents the entire social welfare is lost.

The maximum social welfare possible without external subsidies is the value of the efficient allocation

$$\sum_{i=1}^{m}[v_i - g(m)] \;=\; \sum_{i=1}^{m} v_i - m\, g(m).$$

However, all of this welfare may go into payments. For example, consider the valuation profile where each allocated agent has the same value of $v_1 = \ldots = v_m = g(m) + (m-1)[g(m) - g(m-1)]$ and $v_{m+1} = \ldots = v_n = 0$. The VCG payment of each allocated agent is $(m-1)[g(m) - g(m-1)]$, and the value of the efficient allocation equals the VCG revenue (both are $m(m-1)[g(m) - g(m-1)]$), resulting in zero welfare. This example represents the worst case for social welfare. However, in comparison with the total congestion cost, a relatively high revenue will always be collected—not just in the worst-case. Specifically, we show in Lemma 7 that, for any profile of values, the VCG revenue is close to or exceeds the total congestion cost.

Following the literature on redistribution in allocation domains (see, e.g., Moulin (2009); Guo and Conitzer (2009)), we would like to redistribute as much of the revenue as possible. Following the literature, we adopt the strictest performance metric—that of worst-case guarantee. The performance of a redistribution rule is measured by the fraction of revenue guaranteed to be redistributed for any profile of agents' valuations:

$$r = \min_v \frac{H(v)}{R(v)} \quad \text{where } H(v) = \sum_i h(v_{-i}). \tag{7}$$

The redistribution ratio r equals zero when no redistribution occurs and equals one when the entire revenue is redistributed. Under this metric, the redistribution ratio of the VCG mechanism is zero, as the example above illustrates.

We will prove that applying the Bailey-Cavallo redistribution rule in the congestion setting results in the redistribution of all the revenue asymptotically in the number of agents. We begin with the definition of the Bailey-Cavallo rule.

Definition 2. *The Bailey-Cavallo rule redistributes to each agent an equal share of the revenue collected had the agent not been present:*

$$h(v_{-i}) = \frac{1}{n} R(v_{-i}) \tag{8}$$

$$\textit{where } R(v_{-i}) = \sum_{j \neq i} t_j^{vcg}(v_{-i}). \tag{9}$$

From (7), the ratio of Bailey-Cavallo is:

$$r = \frac{1}{n} \min_v \frac{\sum_i R(v_{-i})}{R(v)} .$$

A crucial requirement of a redistribution function is that it should be *no deficit*: i.e., it should not redistribute more revenue than has been collected, otherwise, an external subsidy will be needed to run the mechanism.

Lemma 1. *Bailey-Cavallo satisfies no deficit, $H(v) \leq R(v)$, for convex g.*

Proof From (7) and (8), we want to show

$$H(v) \leq R(v)$$

$$\sum_j \frac{R(v_{-j})}{n} \leq R(v).$$

It is enough to argue that for any excluded agent j, $R(v_{-j}) \leq R(v)$. Recalling (6)

$$R(v) = m \max\{v_{m+1} - g(m), (m-1)[g(m) - g(m-1)]\}.$$

Let m^{-j} denote the number of items allocated when agent j is not present, and observe that $m^{-j} \in \{m-1, m\}$. When $m^{-j} = m$

$$R(v_{-j}) = m \max\{(v_{-j})_{m+1} - g(m), (m-1)[g(m) - g(m-1)]\}$$

where $(v_{-j})_{m+1}$ refers to the $m+1$ element of vector v_{-j}. As $(v_{-j})_{m+1} \leq v_{m+1}$, it holds that $R(v_{-j}) \leq R(v)$.

Now we consider the case $m^{-j} = m - 1$. Observe that the excluded agent can only be one of the first m agents: $j \leq m$. Had agent $j > m$ been excluded, the first m agents would have been allocated contradicting $m^{-j} = m - 1$. Computing $R(v_{-j})$ when $m^{-j} = m - 1$ given that $j \leq m$, we obtain

$$R(v_{-j}) = (m-1) \max\{v_{m+1} - g(m-1), (m-2)[g(m-1) - g(m-2)]\}$$

By convexity of g, the second term of the max for $R(v_{-j})$, $(m-2)[g(m-1) - g(m-2)]$, is less than the second term of the max for $R(v)$, $(m-1)[g(m) - g(m-1)]$. Thus, it is enough to focus on the first term of the max and show that

$$(m-1)(v_{m+1} - g(m-1)) \leq R(v).$$

Observe that from (6)

$$m(m-1)(g(m) - g(m-1)) \leq R(v).$$

But

$$(m-1)(v_{m+1} - g(m-1)) \leq m(m-1)(g(m) - g(m-1)).$$

follows from

$$v_{m+1} - g(m-1) \leq m(g(m) - g(m-1))$$

which holds as only $m - 1$ agents are allocated (i.e., the agent with value v_{m+1}, who is the mth agent in the market without j, is not allocated). □

We make an observation that provides intuition for why the proof of revenue monotonicity is so involved. Counterintuitively, the VCG payment of an agent may increase when another agent is removed. We illustrate this on a problem with 3 agents: agents 1 and 2 have value 100 and agent 3 has value 4. The congestion function is $g(1) = 0$ and $g(2) = 2$. The VCG payment of agent 1 is 2 when all agents are present and 4 when agent 2 is removed. The total revenue collected, however, is 4 with and without agent 2.

Next, we bound the redistribution ratio by observing that the VCG payments can be bounded from below and above based on the efficient level of congestion.

Lemma 2. *The VCG payment of any allocated agent i is bounded by*

$$(m-1)\,[g(m) - g(m-1)] \;\leq\; t_i^{vcg}(v) \;\leq\; (m+1)\,[g(m+1) - g(m)]. \tag{10}$$

Proof The lower bound follows trivially from (3). We will now prove the upper bound. Since the efficient allocation includes m and not $m+1$ agents, we obtain $v_{m+1} - g(m+1) \leq m\,[g(m+1) - g(m)]$ and

$$\begin{aligned} v_{m+1} - g(m) \;&\leq\; m\,[g(m+1) - g(m)] + g(m+1) - g(m) \\ &=\; (m+1)\,[g(m+1) - g(m)]. \end{aligned} \tag{11}$$

Observe that since g is convex, it follows that

$$(m-1)\,[g(m) - g(m-1)] \;\leq\; (m+1)\,[g(m+1) - g(m)]. \tag{12}$$

The upper bound follows immediately from (3), (11) and (12), completing the proof. □

The upper bound is tight when the value of agent $m+1$ is just below the level that would result in an efficient allocation of $m+1$ items. This is the case when agent $m+1$ is forced out and the VCG payment is given by $v_{m+1} - g(m)$. The lower bound is tight when the value of the $m+1$ agent is below $(m-1)(g(m) - g(m-1))$.

Note that the lower and upper bounds on the VCG payment depend on the efficient level of congestion $m(v)$ but not directly on v. In particular, the bounds are the same for any allocated agent i. Similar to the bounds, congestion prices are defined for a given level of congestion m, but independent of the values of the agents.

Beckmann et al. (1956) showed how edges can be priced in order to induce the efficient congestion level in equilibrium. Specifically, an efficient congestion level m results in the Nash equilibrium of a routing game when each edge is priced at $mg'(m)$. We will call $mg'(m)$ the *efficient congestion price*.

By the convexity of g, the efficient congestion price $mg'(m)$ is bounded by

$$m\,[g(m) - g(m-1)] \;\leq\; mg'(m) \;\leq\; m\,[g(m+1) - g(m)].$$

These bounds fall within the bounds of VCG payments: i.e.,

$$(m-1)\,[g(m) - g(m-1)] \;\leq\; mg'(m) \;\leq\; (m+1)\,[g(m+1) - g(m)].$$

For the following results we introduce a restriction on the rate of growth of g. Specifically, we only consider *sub-exponential* functions. A function $g(m)$ is sub-exponential if $g(m) \in O(2^{h(m)})$ where $h(m) \in o(m)$. As the name suggests, the restriction rules out functions that grow exponentially. Such a restriction is reasonable, since an exponentially growing congestion function would mean that a single agent can have a multiplicative effect on the congestion of all agents, which is highly non-standard for most applications of congestion modeling, such as road congestion or packet routing.

Lemma 3. *For a sub-exponential function g, the relative difference between the VCG payment of agent i and the efficient congestion price approaches zero as the efficient level of congestion $m(v)$ increases.*[3] *This holds for any agent i. In symbols,*

$$\begin{aligned} &\textit{for all } v \in \mathbb{R}^n \mid m(v) \to \infty \\ &\frac{m(v)g'(m(v)) - t_i^{vcg}(v)}{m(v)g'(m(v))} \;=\; 0 \quad \forall\, i. \end{aligned} \tag{13}$$

Proof For ease of notation we write

$$\lim_{m\to\infty} \frac{mg'(m) - t_i^{\mathrm{vcg}}(v)}{mg'(m)}. \tag{14}$$

[3] In the asymptotic analysis involving $m \to \infty$, it is implied that $n \to \infty$ as $n \geq m$.

We need to show

$$\lim_{m\to\infty} \frac{t_i^{\text{vcg}}(v)}{mg'(m)} = 1 \quad \forall\, i. \tag{15}$$

It is enough to show that the lower bound on the VCG payment is at least 1 and the upper bound is at most 1:

$$1 \le \lim_{m\to\infty} \frac{(m-1)[g(m) - g(m-1)]}{mg'(m)} \le \lim_{m\to\infty} \frac{t_i^{\text{vcg}}(v)}{mg'(m)} \le \lim_{m\to\infty} \frac{(m+1)[g(m+1) - g(m)]}{mg'(m)} \le 1.$$

We start with the lower bound. We use convexity of g in the first inequality and sub-exponentiality of g in the second equality:

$$\lim_{m\to\infty} \frac{(m-1)[g(m) - g(m-1)]}{mg'(m)} \ge \lim_{m\to\infty} \frac{m-1}{m}\frac{g'(m-1)}{g'(m)} = 1.$$

Similarly, we obtain for the upper bound:

$$\lim_{m\to\infty} \frac{(m+1)[g(m+1) - g(m)]}{mg'(m)} \le \lim_{m\to\infty} \frac{(m+1)}{m}\frac{g'(m+1)}{g'(m)} = 1.$$

□

This result says that in the congestion model VCG payments are effectively the same as congestion prices. Specifically, VCG payments of all allocated agents are the same and approach the congestion price as m increases.

We now turn to analyzing the performance of the Bailey-Cavallo redistribution rule in the congestion model. In the next lemma, we use the bounds on the VCG payments derived above to provide performance guarantees that are based only on the level of congestion and are independent of the total number of agents. Efficient level of congestion depends on the congestion function and the valuations of agents. Congestion function is a parameter in our derivations. In our analysis, we provide guarantees for all possible valuations. To proceed, we need to subdivide the valuation space into subsets with constant efficient level of congestion: $V_x = \{v \in V \mid m(v) = x\}$, where $V = \{v \in \mathbb{R}^n \mid v_1 \ge v_2 \ge \ldots \ge v_n\}$ and $x \in \{1, 2, \ldots, n\}$.

Lemma 4. *The fraction of revenue redistributed by Bailey-Cavallo is at least*

$$\min_{v\in V_x} \frac{H(v)}{R(v)} \ge \frac{(x-1)(x-2)\,[g(x-1) - g(x-2)]}{x(x+1)\,[g(x+1) - g(x)]} \quad \forall x \in \{1, 2, \ldots, n\}. \tag{16}$$

Proof Recall that under Bailey-Cavallo the total redistribution is $H(v) = \frac{1}{n}\sum_i R(v_{-i})$. For any efficient level of congestion x, it is enough to show that for any agent j

$$\min_{v \in V_x} \frac{R(v_{-j})}{R(v)} \;\geq\; \frac{(x-1)(x-2)\,[g(x-1) - g(x-2)]}{x(x+1)\,[g(x+1) - g(x)]}. \tag{17}$$

Applying Lemma 2 to the market without agent j, we obtain:

$$\begin{aligned}(m(v_{-j})-1)\,[g(m(v_{-j})) - g(m(v_{-j})-1)] \;\leq\; t_i^{\text{vcg}}(v_{-j})\\ \leq\; (m(v_{-j})+1)\,[g(m(v_{-j})+1) - g(m(v_{-j}))].\end{aligned}$$

Excluding an agent from the market either leaves the efficient level of congestion the same, or decreases it by one. Thus, for $v \in V_x$, $x-1 \leq m(v_{-j}) \leq x$ and by convexity of g

$$(x-2)\,[g(x-1) - g(x-2)] \;\leq\; t_i^{\text{vcg}}(v_{-j}) \;\leq\; (x+1)\,[g(x+1) - g(x)]. \tag{18}$$

Expressions (10) and (18) allow us to bound, respectively, the total revenue, $R(v) = \sum_i t_i^{\text{vcg}}(v)$, and the total revenue with agent j excluded, $R(v_{-j}) = \sum_{i \neq j} t_i^{\text{vcg}}(v_{-j})$:

$$\begin{aligned} x(x-1)\,[g(x) - g(x-1)] \;&\leq\; R(v) \;\leq\; x(x+1)\,[g(x+1) - g(x)] \qquad (19)\\ (x-1)(x-2)\,[g(x-1) - g(x-2)] \;&\leq\; R(v_{-j}) \;\leq\; x(x+1)\,[g(x+1) - g(x)]. \end{aligned}$$

Equation (17) follows immediately. □

By convexity of g, the result in Lemma 4 can be stated as

$$\min_{v \in V_x} \frac{H(v)}{R(v)} \;\geq\; \frac{(x-1)(x-2)}{x(x+1)} \cdot \frac{g'(x-2)}{g'(x+1)} \quad \forall x \in \{1, 2, \ldots, n\}. \tag{20}$$

For a sub-exponential g, the fraction of revenue redistributed increases with allocated traffic x and approaches 1 asymptotically as $n, x \to \infty$. This bound can be complemented by the following bound, which approaches 1 asymptotically as $n \to \infty$ for $x < \infty$.

Lemma 5. *The fraction of revenue redistributed by Bailey-Cavallo is at least*

$$\min_{v \in V_x} \frac{H(v)}{R(v)} \;\geq\; \frac{n - x - 1}{n} \quad \forall x \in \{1, 2, \ldots, n\}. \tag{21}$$

Proof For any x, observe that agents $x+2, x+3, \ldots, n$ affect neither the congestion nor the VCG revenue: that is, $R(v_{-i}) = R(v) \quad \forall i \geq x+2$. The amount redistributed to each of these $n-x-1$ agents is exactly $\frac{R(v)}{n}$ and we obtain

$$H(v) = \frac{1}{n}\sum_{i=1}^{n} R(v_{-i}) \geq \frac{1}{n}\sum_{i=x+2}^{n} R(v_{-i}) = \frac{1}{n}\sum_{i=x+2}^{n} R(v) = (n-x-1)R(v)$$

□

Combining the bounds above, we obtain the following result.

Theorem 1. *The fraction of revenue redistributed by Bailey-Cavallo is at least*

$$\min_{v \in V_x} \frac{H(v)}{R(v)} \;\geq\; \max\left\{ \frac{(x-1)(x-2)}{x(x+1)} \cdot \frac{g'(x-2)}{g'(x+1)},\; \frac{n-x-1}{n} \right\} \quad \forall x \in \{1, 2, \ldots, n\}. \tag{22}$$

This bound provides a performance guarantee for any level of efficient congestion and for any number of agents. We will argue below that asymptotically the bound approaches 1: i.e., all of the revenue is redistributed. However, finite-case performance guarantees are non-trivial even for small values of m and n. We illustrate this in Figure 1, which shows the guarantee for two congestion functions: $g(m) = cm$, where c is any constant and $g(m) = m^2$. The two lines are obtained by plotting (20) for the corresponding function g. The decreasing solid line corresponds to (21) for $n = 20$. The performance guarantee for each of the two g functions are independent of n (clearly n has to be at least as high as m). For $n = 20$, the performance guarantee is given by the maximum of the corresponding g line and the solid line. So for a linear g, the worst ratio is around .6 and for the quadratic g, it is around .4. Note that these result does not depend on differentiability of g: we can plug in (16) instead of (20), which will provide a slightly better bound.

We now show that the fraction of revenue redistributed approaches one as the number of agents increases.

Theorem 2. *For a sub-exponential g, the fraction of revenue redistributed by Bailey-Cavallo approaches 1 asymptotically as the number of agents increases:*

$$\lim_{n\to\infty} \min_{v} \frac{H(v)}{R(v)} = 1.$$

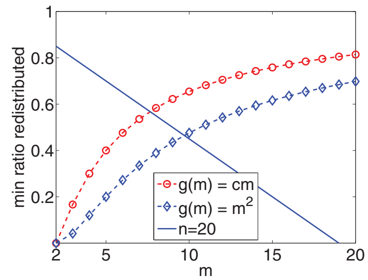

Figure 1: Performance guarantees

Proof

$$\lim_{n\to\infty}\min_{v}\frac{H(v)}{R(v)} = \lim_{n\to\infty}\min_{x\in\{1,2,\ldots,n\}}\min_{v\in V_x}\frac{H(v)}{R(v)} \geq$$
$$\lim_{n\to\infty}\min_{x\in\{1,2,\ldots,n\}}\max\left\{\frac{(x-1)(x-2)}{x(x+1)}\cdot\frac{g'(x-2)}{g'(x+1)},\ \frac{n-x-1}{n}\right\}$$

We need to show that

$$\lim_{n\to\infty}\max\left\{\frac{(x-1)(x-2)}{x(x+1)}\cdot\frac{g'(x-2)}{g'(x+1)},\frac{n-x-1}{n}\right\} = 1 \quad \forall x\in\{1,2,\ldots,n\}$$

For $x<\infty$, the equation holds as $\lim_{n\to\infty}\frac{n-x-1}{n}=1$. For $x=\infty$, the equation holds as $\frac{(x-1)(x-2)}{x(x+1)}\cdot\frac{g'(x-2)}{g'(x+1)}=1$. The last result holds by subexponentiality of g, which implies $\lim_{x\to\infty}\frac{g'(x-2)}{g'(x+1)}=1$. □

Thus, in this model, the Bailey-Cavallo rule asymptotically redistributes all revenue. In contrast, without congestion effects (i.e., in the model where m identical items are allocated), Bailey-Cavallo only redistributes $\frac{n-m-1}{n}$ of the collected revenue (Guo and Conitzer, 2009). Although Bailey-Cavallo is asymptotically optimal in the allocation model when the number of items is negligible relative to the number of agents, it is arbitrarily bad when the number of items is close to the number of agents.

We can provide some intuition for the improved performance of the redistribution rule when congestion is present. The VCG payment of each allocated agent in the allocation model is v_{m+1}. For value profiles where the first $m+1$ agents have the same non-zero value, say x, and the other agents have a zero value, the VCG mechanism collects $m\, v_{m+1} = mx$, but VCG in the market without agent $i \leq m$ collects $m\, v_{m+2} = 0$. As a result, agents $1, 2, \ldots, m{+}1$ receive no redistribution, and the overall fraction of redistributed revenue is low. This does not occur in the congestion model, as, by (18), a VCG payment in the market without agent i is at least $(m-2)\,[g(m-1) - g(m-2)]$.

We note that the optimal redistribution rule for allocating homogeneous items was derived by Moulin (2009) and Guo and Conitzer (2009). Their rule improves upon Bailey-Cavallo guaranteeing $1 - \frac{\binom{n-1}{m}}{\sum_{i=m}^{n-1}\binom{n-1}{i}}$, which asymptotically approaches one when fewer than half of the agents are allocated (Moulin, 2009). To complement these results, de Clippel et al. (2014) derived a non-efficient mechanism that provides asymptotic optimality when the ratio of allocated agents approaches 1, but leaves a positive gap when the ratio of allocated agents is strictly below 1. In contrast, our results for the congestion model prove that Bailey-Cavallo is asymptotically optimal regardless of the number of items allocated (or, the efficient level of congestion).

Finally, we motivate the need to redistribute by showing that collected revenue is high.

Lemma 6. *For a convex g, the revenue collected in congestion prices is at least as high as the total congestion cost $G(m) = mg(m)$.*

Proof The congestion price is $mg'(m)$ and the total revenue collected in congestion prices is $m^2 g'(m)$. Each agent incurs the congestion cost of $g(m)$ for the total congestion cost of $mg(m)$. We want to show

$$m^2 g'(m) \geq mg(m)$$
$$mg'(m) \geq g(m)$$

which follows immediately from the convexity of g.

Note that the congestion revenue can in fact be much higher than the total congestion cost. For $g(m) = m^5$, the congestion revenue is $5m^5$ while the total congestion cost is m^5.

Given that congestion prices are similar to VCG payments as we argued in Lemma 3, a similar results holds for the VCG revenue.

Lemma 7. *For a convex g, the VCG revenue is at least $mg(m-1)$, that is, above the congestion cost when the level of congestion is $m-1$.*

Proof From (19) we have

$$R(v) \geq m(m-1)[g(m) - g(m-1)]$$

From the convexity of g

$$(m-1)(g(m) - g(m-1)) \geq (m-1)g'(m-1) \geq g(m-1)$$

yielding

$$R(v) \geq mg(m-1).$$

□

2.2. Atomic Congestion Games

We now extend the results derived above to a general congestion games model, where K resources are available. Congestion on a resource $r \in K$ used by m_r agents is given by $g_r(m_r)$. Agent i has a value $v_i(L)$ for each subset of resources $L \subseteq K$. Formally, $v_i \in \mathbb{R}^{2^{|K|}}$, and $V = \mathbb{R}^{n2^{|K|}}$.

The VCG payment is now defined for the set of resources allocated to agent i, rather than for individual resources. However, we can bound the VCG payment of agent i based on a per-resource externality imposed on others as we did in (10). Letting $f_i(v)$ or f_i denote the set of resources allocated to agent i, the efficient allocation is

$$\begin{aligned} f &= \arg\max_{f'} \sum_i v_i(f'_i) - \sum_r m_r(f')\, g_r(m_r(f')) \quad \text{where} \\ m_r(f') &= \sum_i \mathbf{1}_{\{r \in f'_i\}}. \end{aligned}$$

Here $m_r(f)$ denotes the congestion on resource r under allocation f. Similar to f, we define f^{-j} as the efficient allocation in the market without agent j.

$$f^{-j} = \arg\max_{f'} \sum_{i \neq j} v_i(f'_i) - \sum_r m_r(f')\, g_r(m_r(f')).$$

We use $m_r = m_r(f)$ and $m_r^{-j} = m_r(f^{-j})$ to denote the efficient level of congestion on resource r in the market with and without agent j respectively. Let $S(v)$ and $S(v_{-j})$ denote the value of f and f^{-j}, respectively. Finally, let

$$\hat{S}_j(v) = S(v) - [v_j(f_j) - \sum_{r \in f_j} g_r(m_r(f))] \tag{23}$$

	v_i	a	b	c	d	e
agent 1	11	x	x	x	x	x
agent 2	4	x	x			
agent 3	4	x		x		
agent 4	4	x			x	
agent 5	4	x				x
agent 6	5	x				

(a) The desired set of resources and its value for each agent.

m_r	g_a	g_b	g_c	g_d	g_e
1	0	0	0	0	0
2	0	5	5	5	5
3	0				
4	1				
5	2				

(b) Congestion functions for resources

Table 1: The table on the left specifies values of agents 1-6 for resources a-d. For example, agent 3 has the value of 4 for the set of resources a and c. The table on the right specifies the congestion function for each resource. For example, the congestion on resource a is 1 when 4 units are used. The empty cells can be filled in with any values satisfying convexity of g_r.

denote the value of the efficient allocation not counting agent j (the amount in brackets is the contribution of agent j to the value of the efficient allocation f). The VCG payment of agent j is

$$t_j^{\text{vcg}}(v) \;=\; S(v_{-j}) - \hat{S}_j(v). \tag{24}$$

General atomic games are more complex than the single-resource games studied in the previous section. In particular, revenue-monotonicity does not hold (as, hence, Lemma 1 does not hold) and upper bounding the VCG payment is more complex than a straightforward generalization of the single-resource upper bound to $\sum_{r\in f_j}(m_r+1)(g_r(m_r+1)-g_r(m_r))$. However, the main result that almost all of the revenue can be redistributed remains. We start with an example illustrating non-monotonicity of revenue and showing that a most natural generalization of the upper bound does not apply.

There are 6 agents (agent 1 through agent 6), and 5 resources (a, b, c, d, e). Each agent is single-minded, i.e., has positive value for a single set of resources. The values of each agent for its corresponding set of interest appear in Table 1a, while the congestion functions for the resources appear in Table 1b. The efficient allocation is for agents 1 and 6 to be allocated their corresponding sets, a, b, c, d, e and a, respectively. The resource usage of this allocation is $m(v) = (2, 1, 1, 1, 1)$. The social welfare is $S(v) = (11+5) - 0 = 16$. When either agent 1 or agent 6 is not present, the efficient allocation is to agents 2 through 5. The resource usage is $m(v_{-1}) = m(v_{-6}) = (4, 1, 1, 1, 1)$. The social welfare is $S(v_{-1}) = S(v_{-6}) = 4 \cdot 4 - 4 = 12$.

The VCG payment of agent 6 is

$$t_6^{\text{vcg}}(v) = 12 - (16 - (5 - 0)) = 1.$$

A direct generalization of the upper bound incorrectly suggests that the payment cannot exceed

$$\sum_{r \in f_6} (m_r + 1)(g_r(m_r + 1) - g_r(m_r)) = (m_a + 1)(g_a(m_a + 1) - g_a(m_a)) = 0.$$

The revenue in the market with all agents is

$$R(v) = t_1^{\text{vcg}} + t_6^{\text{vcg}} = 7 + 1 = 8$$

where $t_1^{\text{vcg}}(v) = 12 - (16 - (11 - 0)) = 7$. The revenue in the market without agent 6 is

$$R(v_{-6}) = t_2^{\text{vcg}} + t_3^{\text{vcg}} + t_4^{\text{vcg}} + t_5^{\text{vcg}} = 4 \cdot 3 = 12$$

where $t_2^{\text{vcg}} = t_3^{\text{vcg}} = t_4^{\text{vcg}} = t_5^{\text{vcg}} = 12 - (12 - (4 - 1)) = 3$ as when agents 6 and $i \in \{2, 3, 4, 5\}$ are absent, the efficient allocation is to agents $\{2, 3, 4, 5\} \setminus \{i\}$, the resource usage is $(3, 1, 1, 1, 1)$ and social welfare is 12. The revenue without agent 6 is higher than the revenue with agent 6. Thus, revenue is non-monotone.

The non-monotonicity of revenue may result in violation of Lemma 1. To recover the no-deficit property, we can apply a generalization of redistribution rule in Equation 8 suggested by Cavallo (2006):

$$\tilde{h}(v_{-i}) = \frac{1}{n} \min_{v_i} R(v)$$

In words, an agent receives $\frac{1}{n}$ of the lowest possible revenue she can induce through her reported value. In the analysis below, we study the original redistribution function in (8) noting that a budget deficit may occur. Bounding the exact amount remains open for future work.

Lemma 8. *The VCG payment of agent j is bounded by*

$$\sum_{r \in f_j} (m_r - 1)\left(g_r(m_r) - g_r(m_r - 1)\right) \;\leq\; t_j^{vcg}(v) \tag{25}$$

$$\leq \sum_{r \in f_j} (m_r^{-j} + 1)(g_r(m_r^{-j} + 1) - g_r(m_r^{-j})) - \sum_{r \in f_j} g_r(m_r) + \sum_{r \in f_j} g_r(m_r^{-j}). \tag{26}$$

Proof Holding the allocation of other agents unchanged, the removal of agent j decreases the congestion cost on resources that she used by exactly $\sum_{r\in f_j}(m_r-1)\,[g_r(m_r) - g_r(m_r-1)]$. Thus, agents $i \neq j$ achieve the welfare of $\hat{S}_j + \sum_{r\in f_j}(m_r-1)\,[g_r(m_r) - g_r(m_r-1)]$ when each agent i is allocated according to f_i and agent j uses no resources. This provides a lower bound on the optimal welfare $S(v_{-j})$ without agent j, and the lower bound in the theorem follows immediately. We now turn to the upper bound.

$$S(v) \geq S(v_{-j}) + v_j(f_j) - \sum_{r\in f_j} g_r(m_r^{-j}) - \sum_{r\in f_j}(m_r^{-j}+1)(g_r(m_r^{-j}+1) - g_r(m_r^{-j})) \tag{27}$$

$$\hat{S}_j(v) + [v_j(f_j) - \sum_{r\in f_j} g_r(m_r)]$$
$$\geq S(v_{-j}) + v_j(f_j) - \sum_{r\in f_j} g_r(m_r^{-j}) - \sum_{r\in f_j}(m_r^{-j}+1)(g_r(m_r^{-j}+1) - g_r(m_r^{-j})) \tag{28}$$

$$S(v_{-j}) - \hat{S}_j(v) \leq \sum_{r\in f_j}(m_r^{-j}+1)(g_r(m_r^{-j}+1) - g_r(m_r^{-j})) - \sum_{r\in f_j} g_r(m_r) + \sum_{r\in f_j} g_r(m_r^{-j})$$

The left-hand side of (27) is the optimal solution while the right-hand side is a feasible solution: agents $j \neq i$ are allocated according to f^{-j}, while i is allocated f_i. Inequality (28) rewrites $S(v)$ as the sum of values of agents other than i and agent i. The final inequality is obtained by rearranging the terms. □

While valuations are combinatorial, we can break bounds down by resource. We define agent's bounds on payment for resource r as

$$\hat{t}_r^j = (m_r - 1)(g_r(m_r) - g_r(m_r - 1)) \tag{29}$$
$$\bar{t}_r^j = (m_r^{-j}+1)(g_r(m_r^{-j}+1) - g_r(m_r^{-j})) - g_r(m_r) + g_r(m_r^{-j}) \tag{30}$$

with the property that the VCG payment is bounded by the sum of bounds on the corresponding resources:

$$\sum_{r\in f_j} \hat{t}_r^j \leq t_j^{\text{vcg}} \leq \sum_{r\in f_j} \bar{t}_r^j \tag{31}$$

The bound is similar to (10), but the link between the efficient congestion levels with and without agent j is more difficult to characterize than in the single-resource case where $m - 1 \leq m^{-j} \leq m$ holds. However, the fundamental similarity remains: an agent pays marginal congestion costs for each of the resources that she uses. We introduce two mild restrictions on how m_r^{-j} relates to m_r.

Assumption 1. *As the efficient congestion level on resource r increases, removing a single agent cannot change the efficient congestion level by a non-vaninishing fraction. In symbols,*

$$\lim_{m_r \to \infty} \frac{m_r^{-j}}{m_r} = 1 \quad \forall r, j$$

Assumption 2. *As the efficient congestion level on resource r increases, removing a single agent cannot change the resource's congestion cost by a non-vaninishing fraction. In symbols,*

$$\lim_{m_r \to \infty} \frac{g_r'(m_r^{-j})}{g_r'(m_r)} = 1 \quad \forall r, j$$

Lemma 9. *Under Assumptions 1 and 2 and for a sub-exponential function g, the relative difference between the VCG payment of agent i and the efficient congestion price approaches zero as the efficient level of congestion m_r increases on all of the resources used by i. This holds for any agent j. In symbols,*

$$\begin{aligned} &\textit{for all } v \in V \mid m_r(v) \to \infty \quad \forall\, r \in f_j \\ &\frac{\left(\sum_{r \in f_j} m_r(v) g_r'(m_r(v))\right) - t_j^{vcg}(v)}{\sum_{r \in f_j} m_r(v) g_r'(m_r(v))} \;=\; 0 \quad \forall\, j. \end{aligned} \tag{32}$$

Proof For ease of notation we write

$$\lim_{\{m_r \to \infty\}_{r \in f_j}} \frac{\left(\sum_{r \in f_j} m_r g_r'(m_r)\right) - t_j^{\text{vcg}}(v)}{\sum_{r \in f_j} m_r g_r'(m_r)}. \tag{33}$$

We need to show

$$\lim_{\{m_r \to \infty\}_{r \in f_j}} \frac{t_j^{\text{vcg}}(v)}{\sum_{r \in f_j} m_r g_r'(m_r)} = 1 \quad \forall j. \tag{34}$$

It is enough to show that the lower bound on the VCG payment is at least 1 and the upper bound is at most 1:

$$1 \leq \lim_{\{m_r \to \infty\}_{r \in f_j}} \frac{\sum_{r \in f_j} \hat{t}_r^j}{\sum_{r \in f_j} m_r g_r'(m_r)} \leq \lim_{\{m_r \to \infty\}_{r \in f_j}} \frac{t_j^{\text{vcg}}(v)}{\sum_{r \in f_j} m_r g_r'(m_r)} \leq \lim_{\{m_r \to \infty\}_{r \in f_j}} \frac{\sum_{r \in f_j} \bar{t}_r^j}{\sum_{r \in f_j} m_r g_r'(m_r)} \leq 1.$$

These inequalities follow from Lemma 10, which proves the bound for each resources individually, and from Lemma 11, which shows how individual bounds combine into the result of this theorem.

Lemma 10. *Under Assumptions 1, 2 and for a sub-exponential function g, the relative difference between the lower bound on resource r and the efficient congestion price approaches zero as the efficient level of congestion $m_r(v)$ increases. The same holds for the upper bound on resource r and the efficient congestion price. This holds for any agent j.*

$$\lim_{m_r\to\infty} \frac{\hat{t}_r^j(v)}{m_r g_r'(m_r)} = 1 \quad \forall\, j \quad \forall\, r$$
$$\lim_{m_r\to\infty} \frac{\bar{t}_r^j(v)}{m_r g_r'(m_r)} = 1 \quad \forall\, j \quad \forall\, r$$

Proof We start with the first equality which holds by convexity and subexponentiality of g:

$$\lim_{m_r\to\infty} \frac{\hat{t}_r^j(v)}{m_r g_r'(m_r)} = \lim_{m_r\to\infty} \frac{(m_r-1)(g_r(m_r)-g_r(m_r-1))}{m_r g_r'(m_r)}$$
$$= \lim_{m_r\to\infty} \frac{m_r-1}{m_r} \cdot \lim_{m_r\to\infty} \frac{g_r(m_r)-g_r(m_r-1)}{g_r'(m_r)} = \lim_{m_r\to\infty} \frac{g_r(m_r)-g_r(m_r-1)}{g_r'(m_r)}$$
$$1 = \lim_{m_r\to\infty} \frac{g_r'(m_r-1)}{g_r'(m_r)} \le \lim_{m_r\to\infty} \frac{g_r(m_r)-g_r(m_r-1)}{g_r'(m_r)} \le \lim_{m_r\to\infty} \frac{g_r'(m_r)}{g_r'(m_r)} = 1. \tag{35}$$

For the second equality, we first show that the limit is at most one. Using convexity, subexponentiality and Assumtion 1 in the last step, we derive:

$$\lim_{m_r\to\infty} \frac{\bar{t}_r^j(v)}{m_r g_r'(m_r)} = \lim_{m_r\to\infty} \frac{(m_r^{-j}+1)(g_r(m_r^{-j}+1)-g_r(m_r^{-j}))-g_r(m_r)+g_r(m_r^{-j})}{m_r g_r'(m_r)}$$
$$\le \lim_{m_r\to\infty} \frac{(m_r^{-j}+1)g_r'(m_r^{-j}+1)-g_r(m_r)+g_r(m_r^{-j})}{m_r g_r'(m_r)}$$
$$\le \lim_{m_r\to\infty} \frac{(m_r^{-j}+1)g_r'(m_r^{-j}+1)+g_r'(m_r^{-j})(m_r^{-j}-m_r)}{m_r g_r'(m_r)} =$$
$$\le \lim_{m_r\to\infty} \frac{(m_r^{-j}+1)g_r'(m_r^{-j}+1)+g_r'(m_r^{-j}+1)\max(m_r^{-j}-m_r,0))}{m_r g_r'(m_r)} =$$
$$\lim_{m_r\to\infty} \frac{(\max(2m_r^{-j}-m_r+1, m_r^{-j}+1))g_r'(m_r^{-j}+1)}{m_r g_r'(m_r)} =$$
$$\lim_{m_r\to\infty} \frac{\max(2m_r^{-j}-m_r+1, m_r^{-j}+1)}{m_r} \cdot \frac{g_r'(m_r^{-j})}{g_r'(m_r)} = 1.$$

We now show that the same limit is at least one. Using convexity, subexponentiality and Assumtions 1 and 2 in the last step, we derive:

$$\lim_{m_r\to\infty}\frac{\bar{t}_r^j(v)}{m_r g_r'(m_r)} = \lim_{m_r\to\infty}\frac{(m_r^{-j}+1)(g_r(m_r^{-j}+1)-g_r(m_r^{-j}))-g_r(m_r)+g_r(m_r^{-j})}{m_r g_r'(m_r)}$$
$$\geq \lim_{m_r\to\infty}\frac{(m_r^{-j}+1)g_r'(m_r^{-j})-g_r(m_r)+g_r(m_r^{-j})}{m_r g_r'(m_r)}$$
$$\geq \lim_{m_r\to\infty}\frac{(m_r^{-j}+1)g_r'(m_r^{-j})+g_r'(m_r)(m_r^{-j}-m_r)}{m_r g_r'(m_r)} =$$
$$= \lim_{m_r\to\infty}\frac{m_r^{-j}+1}{m_r}\cdot\lim_{m_r\to\infty}\frac{g_r'(m_r^{-j})}{g_r'(m_r)}+\lim_{m_r\to\infty}\frac{m_r^{-j}-m_r}{m_r}=1$$

These two bounds yield

$$\lim_{m_r\to\infty}\frac{\bar{t}_r^j(v)}{m_r g_r'(m_r)} = 1.$$

□

Lemma 11.

$$\lim_{\{m_r\to\infty\}_{r\in f_j}}\frac{\sum_{r\in f_j}\hat{t}_r^j}{\sum_{r\in f_j} m_r g_r'(m_r)} = 1 \quad \forall\ j$$
$$\lim_{\{m_r\to\infty\}_{r\in f_j}}\frac{\sum_{r\in f_j}\bar{t}_r^j}{\sum_{r\in f_j} m_r g_r'(m_r)} = 1 \quad \forall\ j$$

Proof From Lemma 10, we have

$$\lim_{m_r\to\infty}\frac{\hat{t}_r^j(v)}{m_r g_r'(m_r)} = 1 \quad \forall\ r.$$

To complete the proof, we apply Lemma 12 to $\alpha_r(m_r) = \hat{t}_r^j(m_r)$ and $\beta_r(m_r) = m_r g_r'(m_r)$. The proof for $\bar{t}_r^j$ is analogous. □

Lemma 12.

$$\lim_{m_r\to\infty}\frac{\alpha_r(m_r)}{\beta_r(m_r)} = 1 \quad \forall r\in K$$

implies

$$\lim_{\{m_r\to\infty\}_{r\in K}}\frac{\sum_{r\in K}\alpha_r(m_r)}{\sum_{r\in K}\beta_r(m_r)} = 1.$$

Proof By definition of the limit we have: $\forall\ \epsilon_r > 0\ \exists c \mid \forall m_r > c$

$$(1-\epsilon_r)\beta_r(m_r) \leq \alpha_r(m_r) \leq (1+\epsilon_r)\beta_r(m_r).$$

Summing over r, we get

$$\sum_r (1-\epsilon_r)\beta_r(m_r) \leq \sum_r \alpha_r(m_r) \leq \sum_r (1+\epsilon_r)\beta_r(m_r).$$

Let $\bar{\epsilon} = \max_r \epsilon_r$:

$$(1-\bar{\epsilon})\sum_r \beta_r(m_r) \leq \sum_r \alpha_r(m_r) \leq (1+\bar{\epsilon})\sum_r \beta_r(m_r)$$

$$(1-\bar{\epsilon}) \leq \frac{\sum_r \alpha_r(m_r)}{\sum_r \beta_r(m_r)} \leq (1+\bar{\epsilon}).$$

The last line is equivalent to

$$\lim_{\{m_r\to\infty\}_{r\in K}} \frac{\sum_{r\in K}\alpha_r(m_r)}{\sum_{r\in K}\beta_r(m_r)} = 1.$$

□

Next we show that the Bailey-Cavallo redistribution rule is asymptotically optimal in atomic congestion games.

As in the previous section, performance bounds depend on a valuation profile only through the profile's effect on the efficient congestion. Our first bound depends on the efficient congestion without any excluded agent j as well as when all agents are present. Let $z_r \in \{1, 2, \ldots, n\} \quad \forall r$ and $x_r^{-j} \in \{1, 2, \ldots, n\} \quad \forall j, r$. The valuations space $V = \mathbb{R}^{n2^{|K|}}$ can be partitioned into regions $V_{z,x} = \{v \in V \mid m_r^{-j}(v) = x_r^{-j}\ \forall r, j \text{ and } m_r(v) = z_r\ \forall r\}$ with constant levels of efficient congestion in the markets with each of the agents excluded and in the market with all agents.

Lemma 13. *The fraction of revenue redistributed by Bailey-Cavallo is at least*

$$\min_{v\in V_{z,x}} \frac{H(v)}{R(v)} \geq \frac{\frac{1}{n}\sum_j \sum_{r\in K} x_r^{-j}(x_r^{-j}-1)\left(g_r(x_r^{-j}) - g_r(x_r^{-j}-1)\right)}{\sum_j \left(\sum_{r\in f_j}(x_r^{-j}+1)(g_r(x_r^{-j}+1) - g_r(x_r^{-j})) - \sum_{r\in f_j} g_r(z_r) + \sum_{r\in f_j} g_r(x_r^{-j})\right)} \tag{36}$$

$\forall z \in \{1, 2, \ldots, n\}^{|K|}, x \in \{1, 2, \ldots, n\}^{|nK|}$

Proof Using (25), for the numerator, we obtain:

$$\begin{aligned} H(v) &= \frac{1}{n}\sum_j R(v_{-j}) = \frac{1}{n}\sum_j \sum_{i\neq j} t_i^{\text{vcg}}(v_{-j}) \\ &\geq \frac{1}{n}\sum_j \sum_{i\neq j} \sum_{r\in f_i^{-j}} (x_r^{-j}-1)\left(g_r(x_r^{-j}) - g_r(x_r^{-j}-1)\right) \\ &= \frac{1}{n}\sum_j \sum_{r\in K} x_r^{-j}(x_r^{-j}-1)\left(g_r(x_r^{-j}) - g_r(x_r^{-j}-1)\right). \end{aligned}$$

Using (26), for the denominator, we have:

$$\begin{aligned} R(v) &= \sum_j t_j^{\text{vcg}}(v) \\ &\leq \sum_j \left(\sum_{r\in f_j} (x_r^{-j}+1)(g_r(x_r^{-j}+1) - g_r(x_r^{-j})) - \sum_{r\in f_j} g_r(z_r) + \sum_{r\in f_j} g_r(x_r^{-j}) \right). \end{aligned}$$

□

By convexity of g, the result in Lemma 13 can be stated as

$$\begin{aligned} \min_{v\in V_{z,x}} \frac{H(v)}{R(v)} &\geq \frac{\frac{1}{n}\sum_j \sum_{r\in K} x_r^{-j}(x_r^{-j}-1)\, g_r'(x_r^{-j}-1)}{\sum_j \left(\sum_{r\in f_j}(x_r^{-j}+1)g_r'(x_r^{-j}+1) - \sum_{r\in f_j} g_r(z_r) + \sum_{r\in f_j} g_r(x_r^{-j})\right)} \\ &\geq \frac{\frac{1}{n}\sum_j \sum_{r\in K} x_r^{-j}(x_r^{-j}-1)\, g_r'(x_r^{-j}-1)}{\sum_j \sum_{r\in f_j}\left((x_r^{-j}+1)g_r'(x_r^{-j}+1) + g_r'(x_r^{-j})(x_r^{-j}-z_r)\right)} \\ &\geq \frac{\frac{1}{n}\sum_j \sum_{r\in K} x_r^{-j}(x_r^{-j}-1)\, g_r'(x_r^{-j}-1)}{\sum_j \sum_{r\in f_j}\left((x_r^{-j}+1)g_r'(x_r^{-j}+1) + g_r'(x_r^{-j}+1)\max(x_r^{-j}-z_r, 0)\right)} \\ &\geq \frac{\frac{1}{n}\sum_j \sum_{r\in K} x_r^{-j}(x_r^{-j}-1)\, g_r'(x_r^{-j}-1)}{\sum_j \sum_{r\in f_j} g_r'(x_r^{-j}+1)\max(2x_r^{-j}-z_r+1, x_r^{-j}+1)} \qquad (37) \\ &\forall z\in\{1,2,\ldots,n\}^{|K|}, x\in\{1,2,\ldots,n\}^{|nK|} \end{aligned}$$

The bound above depends only on the level of efficient congestion. Next we derive a bound that depends on both the number of agents present and the efficient level of congestion. This is a generalization of Lemma 5.

Lemma 14. *The fraction of revenue redistributed by Bailey-Cavallo is at least*

$$\min_{v \in V_{z,x}} \frac{H(v)}{R(v)} \geq \frac{n - \sum_r z_r - \sum_j \sum_r x_r^{-j}}{n} \quad \forall p \tag{38}$$
$$\forall z \in \{1, 2, \ldots, n\}^{|K|}, x \in \{1, 2, \ldots, n\}^{|nK|}$$

Proof Recall the definition of agent j's VCG payment in (24). The only agents that affect the value of $S(v)$ (and therefore the value $\hat{S}_j(v)$ for each j) are the agents who are allocated (removing all other agents has no effect on $S(v)$). The maximum number of allocated agents is $\sum_r z_r$: it occurs when each allocated agent receives exactly one resource. Similarly, the number of agents that determine $S(v_{-j})$ is bounded by $\sum_r x_r^{-j}$. Any other agent i has no effect on VCG payments and receives $h_i(v) = \frac{1}{n}R(v_{-i}) = \frac{1}{n}R(v)$ in redistribution. □

Combining the bounds, we obtain the following result.

Theorem 3. *The fraction of revenue redistributed by Bailey-Cavallo is at least*

$$\begin{aligned} \min_{v \in V_{z,x}} \frac{H(v)}{R(v)} \geq \max\{ & \\ & \frac{\frac{1}{n} \sum_j \sum_{r \in K} x_r^{-j}(x_r^{-j} - 1)\, g'_r(x_r^{-j} - 1)}{\sum_j \sum_{r \in f_j} g'_r(x_r^{-j} + 1) \max(2x_r^{-j} - z_r + 1, x_r^{-j} + 1)}, \\ & \frac{n - \sum_r z_r - \sum_j \sum_r x_r^{-j}}{n} \} \\ & \forall z \in \{1, 2, \ldots, n\}^{|K|}, x \in \{1, 2, \ldots, n\}^{|nK|} \end{aligned} \tag{39}$$

We now establish asymptotic optimality of Bailey-Cavallo as the number of agents increases.

Theorem 4. *For a sub-exponential function g_r and under Assumptions 1 and 2, the fraction of revenue redistributed by Bailey-Cavallo approaches 1 asymptotically as the number of agents increases*

$$\lim_{n \to \infty} \min_v \frac{H(v)}{R(v)} = 1.$$

Proof

$$\lim_{n\to\infty}\min_{v}\frac{H(v)}{R(v)} = \lim_{n\to\infty}\min_{\substack{z\in\{1,2,\ldots,n\}^{|K|}\\ x\in\{1,2,\ldots,n\}^{|nK|}}}\min_{v\in V_{z,x}}\frac{H(v)}{R(v)} \geq$$

$$\lim_{n\to\infty}\min_{\substack{z\in\{1,2,\ldots,n\}^{|K|}\\ x\in\{1,2,\ldots,n\}^{|nK|}}}\max\{\frac{\frac{1}{n}\sum_j\sum_{r\in K}x_r^{-j}(x_r^{-j}-1)\,g_r'(x_r^{-j}-1)}{\sum_j\sum_{r\in f_j}g_r'(x_r^{-j}+1)\max(2x_r^{-j}-z_r+1,x_r^{-j}+1)},$$
$$\frac{n-\sum_r z_r-\sum_j\sum_r x_r^{-j}}{n}\}$$

We need to show that

$$\lim_{n\to\infty}\max\{\frac{\frac{1}{n}\sum_j\sum_{r\in K}x_r^{-j}(x_r^{-j}-1)\,g_r'(x_r^{-j}-1)}{\sum_j\sum_{r\in f_j}g_r'(x_r^{-j}+1)\max(2x_r^{-j}-z_r+1,x_r^{-j}+1)},$$
$$\frac{n-\sum_r z_r-\sum_j\sum_r x_r^{-j}}{n}\}=1$$
$$\forall z\in\{1,2,\ldots,n\}^{|K|}, x\in\{1,2,\ldots,n\}^{|nK|}$$

For $z_r<\infty \quad \forall r$ (note that by Assumption 1, $z_r<\infty\Rightarrow x_r^{-j}<\infty \quad \forall j$), the equation holds as $\lim_{n\to\infty}\frac{n-\sum_r z_r-\sum_j\sum_r x_r^{-j}}{n}=1$.

We now turn to the case when there is at least one resource r such that $z_r=\infty$ (and by Assumption 1, $x_r^{-j}=\infty \quad \forall j$). We rewrite the bound to allow us to consider each resource separately.

$$\lim_{n\to\infty}\frac{\frac{1}{n}\sum_j\sum_{r\in K}x_r^{-j}(x_r^{-j}-1)\,g_r'(x_r^{-j}-1)}{\sum_j\sum_{r\in f_j}g_r'(x_r^{-j}+1)\max(2x_r^{-j}-z_r+1,x_r^{-j}+1)}$$
$$=\lim_{n\to\infty}\frac{\sum_{r\in K}\frac{1}{n}\sum_j x_r^{-j}(x_r^{-j}-1)\,g_r'(x_r^{-j}-1)}{\sum_{r\in K}\sum_{j|r\in f_j}g_r'(x_r^{-j}+1)\max(2x_r^{-j}-z_r+1,x_r^{-j}+1)}$$

Only resources $\hat{K}=\{r\mid z_r=\infty\}$ will have an effect on the bound

$$\lim_{n\to\infty}\frac{\sum_{r\in K}\frac{1}{n}\sum_j x_r^{-j}(x_r^{-j}-1)\,g_r'(x_r^{-j}-1)}{\sum_{r\in K}\sum_{j|r\in f_j}g_r'(x_r^{-j}+1)\max(2x_r^{-j}-z_r+1,x_r^{-j}+1)}$$
$$=\lim_{n\to\infty}\frac{\sum_{r\in \hat{K}}\frac{1}{n}\sum_j x_r^{-j}(x_r^{-j}-1)\,g_r'(x_r^{-j}-1)}{\sum_{r\in \hat{K}}\sum_{j|r\in f_j}g_r'(x_r^{-j}+1)\max(2x_r^{-j}-z_r+1,x_r^{-j}+1)}.$$

We first show that the limit for each resource is one, and then apply Lemma 12 to complete the proof.

$$\lim_{n\to\infty} \frac{\frac{1}{n}\sum_j x_r^{-j}(x_r^{-j}-1)\, g_r'(x_r^{-j}-1)}{\sum_{j|r\in f_j} g_r'(x_r^{-j}+1)\max(2x_r^{-j}-z_r+1, x_r^{-j}+1)} \quad \forall r \in \hat{K}$$

Dividing by $z_r g_r'(z_r)$, and using Assumptions 1 and 2

$$\lim_{n\to\infty} \frac{\frac{1}{n}\sum_j x_r^{-j}}{\sum_{j|r\in f_j} 1} = \lim_{n\to\infty} \frac{\frac{1}{n}\sum_j x_r^{-j}}{z_r} = \lim_{n\to\infty} \frac{1}{n}\sum_j \frac{x_r^{-j}}{z_r} = \lim_{n\to\infty} \frac{1}{n}\sum_j 1 = \lim_{n\to\infty} \frac{1}{n} n = 1.$$

Finally, Lemma 12 allows us to combine the bounds for each $r \in \hat{K}$ into

$$\lim_{n\to\infty} \frac{\sum_{r\in\hat{K}} \frac{1}{n}\sum_j x_r^{-j}(x_r^{-j}-1)\, g_r'(x_r^{-j}-1)}{\sum_{r\in\hat{K}}\sum_{j|r\in f_j} g_r'(x_r^{-j}+1)\max(2x_r^{-j}-z_r+1, x_r^{-j}+1)} = 1.$$

□

We proved that Bailey-Cavallo fully redistributes the revenue collected in congestion domains as the number of agents increases. Specifically, the VCG mechanism with the Bailey-Cavallo rule provides an asymptotically optimal solution to the problem of welfare maximization in congestion games: the resources are used efficiently and only an asymptotically vanishing fraction of social welfare is lost in payments. Thus, in congestion domains, we are able to obtain a *first-best* solution: an efficient allocation and budget-balance.

3. Decentralized Solution: Nonatomic Routing

We now consider nonatomic routing: a class of routing games where the number of agents is large and a single agent has no effect on the efficient level of congestion. In these games, congestion prices are used to achieve the efficient level of congestion. The solution concept here is Nash equilibrium. The agents do not report their values to the center, but rather, the efficient level of congestion arises under the equilibrium behavior of the agents when congestion prices are charged. We show that redistributing the revenue collected in congestion prices in equal shares to all agents (regardless of which resources they use) provides the first-best solution.

We illustrate our results on a canonical example from routing (see, e.g., Roughgarden (2005)). In Pigou's example (see Figure 2) there are two parallel edges

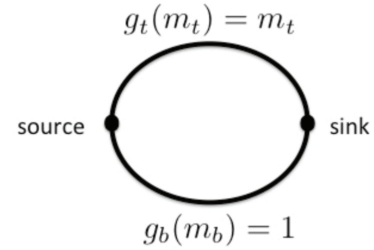

Figure 2: Pigou's example.

and all agents need to be routed from source to sink.[4] The objective is to minimize the routing cost. The cost of the top edge is a function of the traffic, which, in this example, we take to be linear $g_t(m_t) = m_t$ while the cost of the bottom edge is constant $g_b(m_b) = 1$. In nonatomic models, the total traffic can be normalized to 1: that is $m_t + m_b = 1$. The total congestion cost is minimized when half of the traffic is routed along the top edge and half—along the bottom. The congestion cost is $\frac{1}{2} + \frac{1}{2}\frac{1}{2} = \frac{3}{4}$. Without congestion prices, all of the agents take the top edge, and the resulting congestion cost is 1.

Charging congestion prices along each edge results in the efficient level of congestion arising in a Nash equilibrium. We denote by $p_r(m_r) = m_r g_r'(m_r)$ the congestion price along the edge r. For the congestion cost functions in Figure 2, we have $p_t(m_t) = m_t$ and $p_b(m_b) = 0$. It is easy to check that $m_t = m_b = \frac{1}{2}$ is a Nash equilibrium: an agent is indifferent between taking either edge at the equilibrium level of congestion. Let $c_r(m_r) = g_r(m_r) + p_r(m_r)$ denote this total cost which consists of the congestion disutility (or, congestion cost) and the congestion price. At equilibrium, each of the agents taking the top edge incurs a lower congestion cost of $\frac{1}{2}$ but has to pay an additional $\frac{1}{2}$ in congestion prices while the agents on the bottom edge only incur a high congestion cost of 1. From the point of view of the agents, the total congestion cost is the same as in the equilibrium without congestion prices: $C = m_t c_t(m_t) + m_b c_b(m_b) = 1$.

[4]This corresponds to a special case of the model studied in the previous section when each agent has a single value for being routed, and this value is larger than the congestion prices. In this special case, the efficient level of congestion is the one that minimizes the total congestion cost.

We improve on this by returning to the agents the cost incurred in congestion prices. The revenue collected is $R(m) = m_t p_t(m_t) = m_t m_t g_t'(m_t) = \frac{1}{4}$. Suppose each agent receives the rebate of $h(m) = \frac{1}{4}$ regardless of which edge she takes. Incentives are not affected by the rebate: in the equilibrium without rebates, agent's costs for taking either edge are equal, but now each of the costs is reduced by $h(m)$. So the agents taking either edge incur the cost of $\frac{3}{4}$ and the total congestion cost of $C = \frac{3}{4}$ which is the total congestion cost under the efficient level of congestion, and therefore, the first-best solution.

We extend the example above to a routing model. There is a network given by a graph G. Each pair of nodes is associated with the amount of traffic r_k that needs to be routed from the first (source) to the second (sink) node. The traffic r_k is composed of the demand of a large number of agents with contribution of any individual agent being negligible. Edges of the graph correspond to resources r and each is associated with a congestion function g_r. Congestion price $p_r(m_r) = m_r g_r'(m_r)$ is collected from everyone using resource r.

The total revenue collected is the sum of congestion prices paid by agents on each edge they take:

$$R(m) = \sum_r m_r p_r(m_r). \tag{40}$$

Let $h(m)$ denote the redistribution (or, subsidy) that an agent receives. We set $h(m)$ to be an equal share of the total revenue $\int_0^1 h(m)dx = R(m)$ or, simply, $h(m) = R(m)$ (recall that in nonatomic models the total traffic is set to 1; setting it to n, we would get $h(m) = \frac{R(m)}{n}$).

Theorem 5. *The efficient level of congestion is the Nash equilibrium of the mechanism that charges the congestion price $p_r(m_r) = m_r g_r'(m_r)$ for resource r and redistributes to each agent the amount $h(m) = \sum_r m_r p_r(m_r)$.*

Proof The proof follows from the result that the efficient level of congestion is the Nash equilibrium of the game without rebates (Beckmann et al., 1956) and the observation that the rebate an agent receives is independent of her actions. Indeed, the rebate is defined as $R(v)$ and the agents affect on $R(v)$ is negligible: $R(v) = R(v_{-i})$. Note that the mechanism is budget-balanced by construction. □

We link this result to the study of subsidies in (Maillé and Stier-Moses, 2009). The authors investigate the use of subsidies instead of congestion prices to regulate congestion. Indeed, subsidies can be viewed as negative congestion prices. For the example we covered, the authors suggest the subsidy of $\frac{1}{2}$ for the bottom edge. This will result in the efficient level of congestion, but will require an influx of money to subsidize the agents taking the bottom edge: $\frac{1}{2}\frac{1}{2} = \frac{1}{4}$. The authors leave the question of considering subsidies in combination with prices for future work. In particular, subsidizing the mechanism is not desirable. As the example above illustrates, subsidies can be used in combination with congestion prices to achieve full budget-balance.

The congestion games model discussed in Section 2.2 and the nonatomic model above coincide as the number of agents in the atomic model becomes large and the effect of a single agent becomes negligible. When a single agent has no noticeable effect on the revenue $R(v_{-i}) \approx R(v)$, the redistribution in (8) becomes $\frac{R(v)}{n}$.

4. Discussion

The redistribution problem arises in scenarios where payments are needed to achieve the efficient use of resources but the ultimate objective is the total welfare of the agents. Welfare maximization has been previously studied in the allocation of items. There, the VCG mechanism is typically used to choose the efficient allocation but it can collect arbitrarily high payments from the agents. The collected payments must leave the system in order to maintain dominant-strategy implementation (see, e.g., Moulin (2009); Guo and Conitzer (2009)). To this end, a number of recent papers have suggested ways of distributing much of the revenue in various allocation models (see Naroditskiy et al. (2013) and references therein). In contrast, the problem of redistribution in congestion scenarios has received little prior attention, despite the high revenue collected in congestion prices.

We provide positive results for a congestion model: an efficient outcome and asymptotic budget balance are obtained when the Bailey-Cavallo redistribution rule is applied. This finding is surprising given that, in the absence of congestion effects, no known mechanism provides an asymptotically first-best solution. In the absence of congestion effects, the optimal efficient mechanism cannot guarantee non-zero social welfare when the number of items is close to the number of agents (Moulin, 2009; Guo and Conitzer, 2009). Here inefficient mechanisms provide better performance (de Clippel et al., 2014), but still do not provide a first-best solution. While no impossibility results have been derived for allocating multiple-identical items to agents with unit demand, it is conceivable that there is no first-best solution for allocation without congestion. Our results show that the introduction of congestion effects ensures the existence of the first-best solution; further, we prove that the Bailey-Cavallo rule achieves it.

In addition to providing a first-best solution to welfare-maximization in the congestion model, this paper connects results on redistribution in centralized dominant-strategy VCG mechanisms and decentralized congestion-price mechanisms. In particular, we formally show that congestion prices and VCG payments are fundamentally similar and that so are the optimal solutions in both models.

References

Adler, J. L., Cetin, M., 2001. A direct redistribution model of congestion pricing. Transportation Research Part B: Methodological 35 (5), 447–460.

Bailey, M., April 1997. The demand revealing process: To distribute the surplus. Public Choice 91 (2), 107–26.

Beckmann, M., McGuire, C., Winsten, C., 1956. Studies in the economics of transportation. Research memorandum. Published for the Cowles Commission for Research in Economics by Yale University Press.

Blumrosen, L., Dobzinski, S., 2007. Welfare maximization in congestion games. Selected Areas in Communications, IEEE Journal on 25 (6), 1224–1236.

Cavallo, R., 2006. Optimal decision-making with minimal waste: Strategyproof redistribution of VCG payments. In: AAMAS. Hakodate, Japan, pp. 882–889.

Chakrabarty, D., Mehta, A., Nagarajan, V., 2005. Fairness and optimality in congestion games. In: Proceedings of the 6th ACM conference on Electronic commerce. EC '05. pp. 52–57.

Cole, R., Dodis, Y., Roughgarden, T., 2006. How much can taxes help selfish routing? Journal of Computer and System Sciences 72 (3), 444–467.

Courcoubetis, C., Weber, R., 2003. Pricing Communication Networks. John Wiley & Sons, Ltd.

de Clippel, G., Naroditskiy, V., Polukarov, M., Greenwald, A., Jennings, N., 2014. Destroy to save. Games and Economic Behavior 86 (C), 392–404.

Green, J., Laffont, J., 1977. Characterization of satisfactory mechanisms for the revelation of preferences for public goods. Econometrica 45, 427–438.

Guo, M., 2012. Worst-case optimal redistribution of VCG payments in heterogeneous-item auctions with unit demand. In: AAMAS. pp. 745–752.

Guo, M., Conitzer, V., 2008. Better redistribution with inefficient allocation in multi-unit auctions with unit demand. In: EC'08. pp. 210–219.

Guo, M., Conitzer, V., 2009. Worst-case optimal redistribution of vcg payments in multi-unit auctions. Games and Economic Behavior 67 (1), 69 – 98.

Holmstrom, B., September 1979. Groves' scheme on restricted domains. Econometrica 47 (5), 1137–44.
URL http://ideas.repec.org/a/ecm/emetrp/v47y1979i5p1137-44.html

Kelly, F., 1997. Charging and rate control for elastic traffic. European Transactions on Telecommunications 8 (1), 33–37.

MacKie-Mason, J., Varian, H., Sep 1995. Pricing congestible network resources. Selected Areas in Communications, IEEE Journal on 13 (7), 1141 –1149.

Maillé, P., Stier-Moses, N. E., 2009. Eliciting coordination with rebates. Transportation Science 43 (4), 473–492.

Mas-Colell, A., Whinston, M., Green, J., 1995. Microeconomic Theory. Oxford University Press, New York.

Moulin, H., 2009. Almost budget-balanced VCG mechanisms to assign multiple objects. Journal of Economic Theory 144 (1), 96–119.

Naroditskiy, V., Guo, M., Dufton, L., Polukarov, M., Jennings, N. R., 2012. Redistribution of VCG payments in public project problems. In: WINE. pp. 323–336.

Naroditskiy, V., Polukarov, M., Jennings, N., 2013. Optimal payments in dominant-strategy mechanisms for single-parameter domains. ACM Transactions on Economics and Computation 1 (1), 4.

Pigou, A. C., 1920. The economics of welfare. Macmillan, London :.

Roughgarden, T., 2005. Selfish Routing and the Price of Anarchy. The MIT Press.

Vickrey, W. S., May 1969. Congestion theory and transport investment. American Economic Review 59 (2), 251–60.