



HAL
open science

Insights from economic studies of telecommunication networks

Patrick Maillé

► **To cite this version:**

Patrick Maillé. Insights from economic studies of telecommunication networks. Multimedia [cs.MM]. Télécom Bretagne; Université de Rennes 1, 2015. tel-01217488

HAL Id: tel-01217488

<https://hal.science/tel-01217488>

Submitted on 26 Oct 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

TÉLÉCOM BRETAGNE
UNIVERSITÉ RENNES 1

THÈSE

pour obtenir une

Habilitation à Diriger des Recherches

Présentée par

Patrick MAILLE

Insights from economic studies of telecommunication networks

Soutenue le 7 octobre 2015

Devant le jury composé de :

<i>Rapporteurs :</i>	Eitan ALTMAN	-	Inria
	Anna NAGURNEY	-	Univ. Massachusetts Amherst
	Samson LASAULCE	-	CNRS / Centrale-Supélec
<i>Examineurs :</i>	Dominique BARTH	-	Univ. Versailles-Saint Quentin
	Tijani CHAHED	-	Telecom SudParis
	Bruno TUFFIN	-	Inria

Preface

Network economics is a very hot topic, at the same time from a research point of view (with several conferences devoted to the theme, plus a devoted section in most of the other main telecommunication conferences), from a political point of view (as highlighted by the network neutrality debate, the increasing discussion on volume-based pricing, etc.), and of course from a business point of view (encompassing advertisement pricing definition, spectrum selling and sharing, bundling of offers, etc.).

This dissertation summarizes about ten years of research experience in the area after obtaining the PhD degree. While my PhD work only focused on designing auctions and applying them in telecommunication networks (basically, to sell bandwidth on communication links), I have had the opportunity since then to develop and study models for a variety of issues and settings, including the competition among providers at several levels (through price, technological investments, retention strategies), the economics of security and of peer-to-peer storage systems, optimal routing issues, and the debates regarding network neutrality and search neutrality. In all those cases the interactions among actors need to be described along with a solid scientific foundation towards a careful analysis.

Several more topics remain to be investigated, as highlighted in the last chapter of this document.

I would like to thank all my co-authors for the stimulating exchanges of ideas we have had, and in particular Bruno Tuffin for the very fruitful collaboration we have maintained since the end of my PhD work. This dissertation contains several excerpts from the book we have co-written, I also thank him for allowing me to use those parts.

I am very grateful to Anna Nagurney, Eitan Altman, and Samson Lasaulce, who gave me the honor of evaluating this document, providing me with some constructive feedback, and serving to my defense committee. This version has been improved based on their remarks; of course, all errors and omissions remain my sole responsibility. Finally, many thanks go to Dominique Barth, Tijani Chahed and Bruno Tuffin (again) for participating to my defense committee.

And to Maria, for her love and support.

Contents

1	Why study network economics?	1
1.1	A joint soaring growth of networks and demand	1
1.2	A (more and more) complex set of actors	2
1.3	Some economic problems to solve	4
1.4	The need for economic modeling and analysis	5
1.5	Methodology and tools	6
1.6	Are we the best community to address network economy?	7
1.6.1	Why should network economics be studied by non-economists?	8
1.6.2	What remains to do in network economy after a whole century of research?	8
1.6.3	Does research on networking done by mathematicians and engineers have an impact on economists? Is this research published in economics journals?	8
1.7	Outline of the dissertation	9
2	Modeling user behavior with nonatomic games	11
2.1	Nonatomic routing game models	11
2.1.1	Application: Influencing routing games through rebates .	13
2.1.2	Application to wireless cellular networks with heterogeneous users	15
2.1.3	Application to users choosing a security solution	19
2.2	Managing a peer-to-peer storage system	22
2.3	Summary	24
3	Competition among access providers	25
3.1	Association models based on user utility	25
3.1.1	Competition on sent packet prices	25
3.1.2	Discrete-choice models	30
3.2	Dynamic models	33
3.3	Providers competing in multiple-time-scale decision games	38
3.4	To license or not to license resources?	41
3.5	Summary	43

4	Interactions among content/app. providers	45
4.1	Introduction	45
4.2	Competition at the content level	46
4.2.1	General models	46
4.2.2	Illustrative model of competition between free CPs with advertisement	48
4.3	Economics of network security	52
4.4	Summary	53
5	Net and search neutrality	55
5.1	The network neutrality issue	55
5.1.1	Introduction and historical facts	55
5.1.2	Modeling content and network providers interactions and analyzing neutrality issues	57
5.2	Search neutrality	65
5.2.1	The debate	65
5.2.2	Do we need a regulatory intervention?	66
5.2.3	Neutral <i>versus</i> non-neutral search engine: a model	66
5.2.4	Comparison of the Neutral and Non-Neutral Ranking Policies	69
5.2.5	Vertical Integration and Investment	74
6	Conclusions and perspectives	77
6.1	A boundless research area	77
6.2	Some perspectives	78
6.2.1	The role and impact of Content Delivery Networks	78
6.2.2	Software-Defined Networking: principles, opportunities and threats for the Internet ecosystem	79
7	Bibliography	81
A	Main publications evoked in Chapter 2	91
	Eliciting coordination with rebates [75]	93
	Incentivizing efficient load repartition in heterogeneous wireless net- works with selfish delay-sensitive users [43]	115
	Interplay between security providers, consumers, and attackers: a weighted congestion game approach [73]	123
	Managing a peer-to-peer data storage system in a selfish society [76]	145
B	Main publications evoked in Chapter 3	153
	Price war in heterogeneous wireless networks [79]	155
	Price war with migrating customers [71]	197
	Technological investment games among wireless telecommunications service providers [86]	207

C Main publications evoked in Chapter 4	227
Influence of search neutrality on the economics of advertisement- financed content [28]	229
D Main publications evoked in Chapter 5	249
Impact of competition between ISPs on the net neutrality debate [29]	251
Revenue-maximizing rankings for online platforms with quality- sensitive consumers [64]	263

Chapter 1

Why study network economics?

1.1 A joint soaring growth of networks and demand

Telecommunication networks are occupying an increasing role in our daily life: almost everything is now available from the Internet (possibly from a mobile phone), and getting this kind of access has even become compulsory for many administrative operations, without mentioning the social pressure to be part of the trend. While the telephone network commercially started in 1877 following the birth of Bell Telephone Company, and its development slowly democratized, mobile networks and the Internet have quickly occupied a major place since the 70s: now more than 90% of the global population can reach a wireless network [51], and the number of broadband mobile subscriptions has exploded, reaching 2.3 billion in 2014, while the total number of Internet users is estimated to 3 billion. As an example of usage changes, Americans spend on average more time online (about 32 hours per week) than watching television. With an increasing number of subscribers but also because of more and more demanding applications in terms of bandwidth and resources, the Internet sees a tremendous increase of traffic worldwide, as illustrated in Figure 1.1. The networks themselves, and in particular the Internet (the key network of our analysis, as the one over which all networks converge), have considerably evolved to be able to cope with this increase, offering new possibilities and services [98]. The initial (actually, non-existing) business model of the Internet, with free interconnection between academics, and that is still partly in application, has therefore been more and more questioned: some ways to make revenues have to be defined to at least cover the capacity expansion costs.

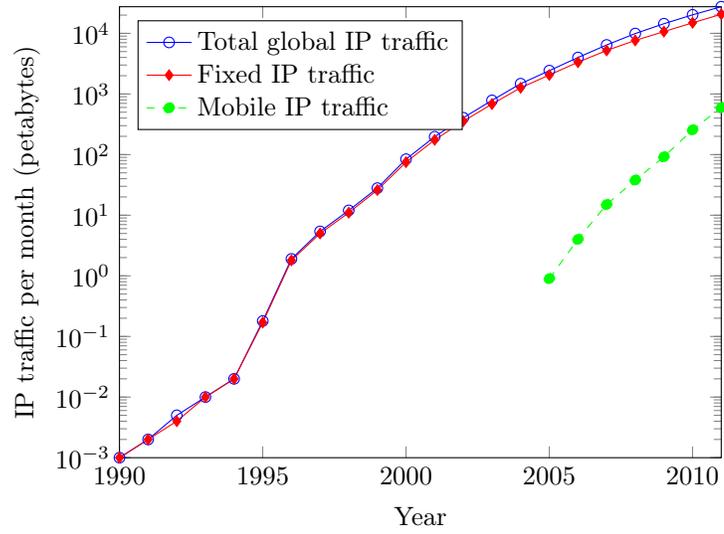


Figure 1.1: IP traffic evolution in petabytes per month, logarithmic scale (Cisco data).

1.2 A (more and more) complex set of actors

While the Internet was initially built by academics having the same purpose, its expansion has been accompanied with the involvement of new actors, most of them being commercial entities and therefore aiming at profit.

Nowadays telecommunication systems involve several categories of actors, as sketched in Figure 1.2; we briefly describe them here to fix the vocabulary used in this dissertation, and to highlight the complexity of the ecosystem. The various categories, and the presented examples of entities pertaining to those categories, are of course debatable.

1. *End users* are the actors to which services are delivered, and whose interest is in getting those services if the prices are “reasonable”. Modeling users’ behavior and level of acceptance in terms of price for a given service has always been hard; they are often represented through demand functions describing their reactions to the propositions of services and prices.
2. *Access network service providers*, also sometimes called *eyeball providers*, offer access to the end users, through a wired line at home or through radio links. Several access technologies coexist, and not all providers operate on all technologies: they have to strategically decide on which ones to make investments in terms of infrastructures, and licenses for wireless technologies. Initially, access operators always operated their network, but now *virtual operators* can lease the network of a competitor to serve

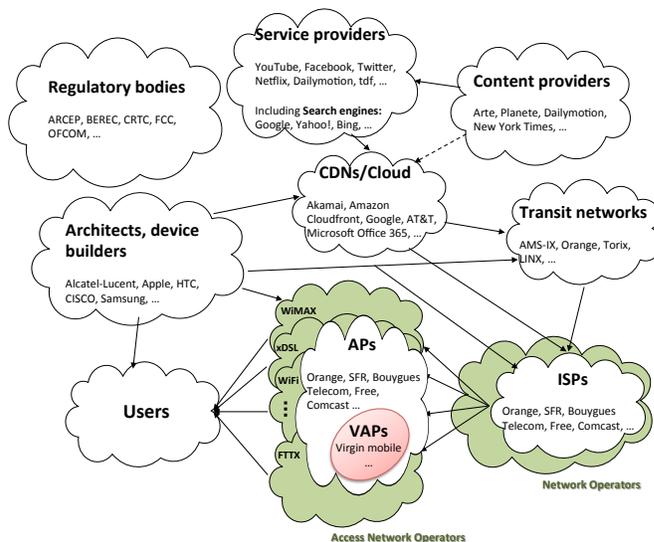


Figure 1.2: The main actor categories of the Internet.

customers; this has been encouraged by regulators in the mobile industry to foster competition.

3. *Transit providers* run the network resources at the core to transfer traffic between access providers and/or other transit providers. Transit providers are organized in a multi-tier hierarchy (even if the boundaries between levels tend to blur), with a clique of Tier-1 providers connected through peering links; more regional Tier-2 providers, customers of Tier-1 providers who can peer with some other providers but need to pay transit fees to reach some parts of the network; and smaller businesses Tier-3 providers, themselves customers of Tier-2 providers. Note that some access providers are also transit providers (Orange in France, British Telecom in the UK, etc).
4. *Content providers* create information, educational or entertainment content for the Internet, CD-ROMs or other software-based products. In the Internet, content mainly used to be produced by users, located usually on their individual web pages, but we now also have “bigger” content providers, i.e., companies selling goods, music (Sony or Studio Universal), displaying news (CBS, BBC, all newspapers such as the New York Times), providing videos (TV channels such as Arte or Planete in France). Some content providers make money by selling their content, while others provide the content for free but get their revenue by displaying advertisements banners.

5. *Service providers* (or more exactly application service providers) provide facilities or doorways to/for content providers. We can think of the portals of ISPs on which news and personal web pages can be published, or YouTube or Dailymotion on which you can publish your videos or music, but also social networks such as Facebook or Twitter (remark that Facebook could be seen as a content provider, since owning the copyright of all published content). In most of those cases, providers make money thanks to (targeted) advertisements on the displayed pages. Netflix, an Internet subscription-based service for watching movies and TV programs, being among the exceptions. Other specific categories of service providers we can think of are search engines and application stores.
6. *Content delivery networks (CDN) and cloud operators* are large distributed systems of servers deployed in the Internet, used by content and service providers for their infrastructures: while service providers provide the tools, the computer facilities are managed by CDNs. For example, Netflix's customer traffic is supported by Amazon's cloud services and Akamai. Basically, CDNs are paid by content providers and pay network providers for the traffic exchanged but more specific economic relationships can exist. Somehow similarly (the resource offered being computational power instead of storage), *clouds* provide facilities to perform operations for companies and individuals without having to buy their own costly resources.
7. *Network architects and device constructors*, such as Cisco or Alcatel-Lucent, provide the network infrastructures (devices) and softwares for performance-optimized transmissions with limited complexity. Device constructors such as Apple, Nokia, Samsung, Hewlett Packard sell mobile phones and computers, "physical tools" for an efficient use of applications provided by CDNs and CPs.
8. *Regulators* are agencies established by governments to control the market operations. They intervene to protect public interests, notions of fairness, etc, hence having an important role when some actors have a dominant position on the market. Regulators also ensure satisfying relationships between the different sets of actors.

1.3 Some economic problems to solve

Among the many questions that the telecommunication network actors need to answer, we can non-exhaustively mention:

- *Determining the most relevant and profitable access network pricing scheme for end users*, in a competitive context between access network providers, also called Internet service providers (ISPs). In particular, the fact that users are used to the so-called *flat-rate* pricing scheme has to

be kept in mind [37], but may not prevent providers from implementing usage-based pricing schemes.

- *Determining the best investments for network providers:* new technologies, capacity and infrastructure expansions, participation in spectrum auctions, are high-stakes strategic decisions to be taken very cautiously.
- *Managing economic relations between network operators:* transit operators, access operators, Content Delivery Network operators, all belong to a complex ecosystem, and need each other to run their business. The economic agreements among them are then complicated to analyze.
- *Understanding the relations between content providers and network providers.* The Internet now counts some large and bandwidth-consuming content and service providers (for example YouTube). Hence the ISPs to which they are not connected (but who have to forward their traffic) start to wonder why distant content providers should not be charged by them [66]. This ignited the *network neutrality* debate at the end of 2005, a still vivid debate where regulators play an important role.
- *Defining economic models for content and (application) service providers.* To get revenue, content and (application) service providers can either charge users for content access, or provide free access but insert advertisements. A similar issue occurs for search engines such as Google, Yahoo! or Bing. Search engines make money by presenting advertisement links, usually at the top and/or on the right of the page, in addition to the search results. Advertisement links are selected from an auction scheme whose rules have evolved since its creation. Note that the combined search engine revenue of Baidu and Google exceeded \$43 billion in 2014, this business still expanding.

Those items illustrate some of the stakes for all actors in the telecommunications business. Since telecommunication networks keep evolving, and with the complex set of actors involved, the question about the most favorable economic decisions always comes back.

1.4 The need for economic modeling and analysis

We now highlight that mathematical modeling and analysis is an important way to avoid pitfalls which can have dramatic consequences, through some simple examples.

The tragedy of the commons. An argument often advocated for changing the Internet pricing is the so-called *tragedy of the commons*, stating that several individuals acting independently, rationally and selfishly can actually deplete a shared limited resource. This type of counter-intuitive outcome can be understood thanks to modeling and analysis through game theory. In telecommunication networks, such situations can occur in several settings, for example

with regard to the use of unlicensed (wireless) radio spectrum, or of the Internet as a whole. Hence the need to carefully design mechanisms to incentivize an efficient use of the resources.

The Braess paradox also illustrates the need to be careful when taking decisions, especially at the governmental/regulation level. Like the tragedy of the commons, this undesirable outcome stems from user selfishness: adding extra capacity to a network when users selfishly choose their route, may actually harm all users. The possibility of experiencing the Braess paradox in the context of wireless networks has been highlighted in [4].

Spectrum auctions. The principle of spectrum auctions is that governments sell *licenses* to providers, allowing them to operate on specific bands of the radio spectrum. An *auction* conducted properly allows to allocate the resource to those who value it most, and to yield high revenues to governments. But the goals of regulators are also to ensure a fast deployment of services, good quality of service, and a fair competition among providers, to the benefit of end users. Though, some problems have been noticed in the auctions previously run, such as collusion among bidders, or final prices too high for providers. A careful design of the auction process and rules is therefore needed to end up with a spectrum allocation that is good for customers, and allows providers to get return on investment and governments to earn money. While spectrum auctions are not developed in this dissertation, it is another example where economic modeling is necessary (and indeed, the FCC and other organizations rely on renowned economists to design appropriate schemes).

Network Neutrality basically states that all packets should be treated equally by intermediate nodes in the network. Neutrality proponents argue that allowing differential treatment of Internet flows would harm innovation (since only those who could afford good quality may succeed), and could also impinge *freedom of speech* and the right to information. On the other hand, the opponents claim that ISPs currently have no incentive to invest in networking capacities because of their reduced revenues; also economic arguments plead for vertical integration (one actor being involved in several stages of the value chain) and service differentiation as ways to increase not only the revenues of providers, but also social welfare. Here again, some careful analysis is needed to weigh both sides and possibly draw recommendations to regulators.

1.5 Methodology and tools

This dissertation aims to show how I used modeling and analysis in telecommunication economics, for different settings, in order to understand the behavior of the current Internet and some of the evolutions that we are experiencing. The work presented here is in the stream of works on *revenue management* [119], which tries to maximize the revenue of the actors.

The tools involved are economics, in particular the branch of game theory and applied mathematics (optimization). Game theory is the study of interactions among several decision-makers (called actors, or players) having different

1.6. ARE WE THE BEST COMMUNITY TO ADDRESS NETWORK ECONOMY??

objectives [41, 100]. A key notion in game theory is the Nash equilibrium, that characterizes an outcome (a decision made by each player) from which no individual player can improve his objective through a unilateral move. Therefore, Nash equilibria represent stable situations, and are considered as good predictions for the result of the interaction.

My research methodology has evolved with the tools, theories and frameworks I have discovered through my readings, my teaching classes, and mainly the collaborations with some brilliant researchers. For example, I became acquainted with (atomic and non-atomic) routing games during my 6-month visit to Columbia University in 2006, during which I deeply studied them and worked on them with Nicolás Stier-Moses; since then I used that formalism for several different settings. Similarly, working with Bruno Tuffin and teaching stochastic processes helped me develop and study stochastic models for network economics, in settings going from user churn to search engine ranking strategies. Finally, my knowledge of the network infrastructure and protocols has considerably improved thanks to numerous exchanges with my colleagues in the RSM department of Telecom Bretagne, allowing to build models closer to the technical realities.

Among the contributions presented here, several consider different levels of decision making: for example a regulator will not make decisions at the same time as operators choosing whether to invest or not in a technology, or as end-users selecting an ISP for Internet access. For such settings, we use the classical *backward induction* method, where we assume at each stage that the decision-makers anticipate the resulting outcome from the stage below (e.g., that regulators anticipate how operators will behave).

Our methodology is also characterized by the willingness to obtain analytical results, in order to be able to interpret the influence of each model parameter. This is often at the cost of simplifications made to the models. When models become intractable analytically, we resort to numerical computations and/or to simulations.

1.6 Are we the best community to address network economy?

The “computer networking” community’s main field of study covers the technical aspects of computer networks. This means, to a great extent, the definition and analysis of communication protocols from a security, flexibility, implementation cost, or performance evaluation point of view.

Hence the study of economic relationships, and incentives, is not the specialty of this community, despite the appearance of conferences, special issues and journals devoted to economic problems among the networking community. This raises several important questions, highlighted in the next subsections.

1.6.1 Why should network economics be studied by non-economists?

Network economics is a highly interdisciplinary field, at the boundary between networking, applied mathematics, computer science, and economy. And several communities . But of course the main tools and reasoning principles are from economy. I happen to have a background in economy, from my education at Ecole polytechnique. I do not consider myself as a “pure” economist, but I am aware of some of the economic vocabulary and tools, mainly in microeconomics and in particular in game theory.

Besides, I have had the chance to collaborate with economists such as Galina Schwartz (UC Berkeley), which confirmed me with the idea that the gap between our communities is not so large. We can moreover be very complementary, for a good knowledge of the technical aspects can help define fine-grained models, and tools from economy can be needed to solve those models, together with an “economy-oriented” mind to interpret the results.

The models I try to develop in my work are quite specific, based on some particular scenarios, and while probably covered by the more general settings economics tend to analyze, the model specificities sometimes make them solvable when the general ones were not. In summary, I do not claim to be a researcher in economy (in the sense of improving economy tools and methodologies), but rather a researcher who *applies* economic reasoning and tools to networking situations.

In that context, the contribution from the “networking” community to network economics is, in my opinion, in the definition of models addressing specific questions appearing with the evolution of telecommunication networks. The risk is to “reinvent the wheel” by defining models that would be subcases of existing models studied by economists; it’s a pitfall we try to avoid. The tools necessary to solve those problems are not always specific to economics, but are part of the “common knowledge” of most communities studying such problems, involving mainly optimization methods.

1.6.2 What remains to do in network economy after a whole century of research?

As pointed out before, what remains to do is tackle very specific problems, that were not covered by general economic models. Hopefully, when those models are specific enough we can find some analytical solutions and get some insight from them.

1.6.3 Does research on networking done by mathematicians and engineers have an impact on economists? Is this research published in economics journals?

There are many links between communities: some research results from mathematicians/engineers are published in economics journals (like, for routing games,

in journals on game theory, or for models on network neutrality, in journals focusing on policy). In the other direction as well, some economists publish in journals/conferences from the engineering/mathematics community. As another sign of those links, some recent journals like *Netnomics*, or conferences/workshops like *NetEcon* are new convergence venues for the diverse communities working on network economics.

In summary, I think engineers, mathematicians and computer scientists can definitely bring interesting contributions to the field of network economics. But we would benefit from more interactions with economists.

1.7 Outline of the dissertation

The rest of this dissertation is organized as follows.

Chapter 2 focuses on noncooperative situations at the user level, and discusses how the outcome can be improved through outside interventions (classically, via monetary incentives). The models developed here will also serve as inputs for several parts of the other chapters.

Chapter 3 discusses the economic relations among ISPs. While initially, the telecommunication industry was a monopoly in most countries, the increasing competition dramatically changes the picture. The chapter presents my main contributions in the study of competition among access providers. When decisions other than prices have to be taken, we describe this competition as multi-level games, played at different time scales: at the largest ones, providers choose where and how much they invest (in terms of capacity, QoS provisioning, etc.), then they choose their price competitively, and at the shortest time scale, users choose their provider. Those games are usually solved by *backward induction* as described before: the games at a higher level are solved anticipating the solution of lower-level games. We present models of customer churn (i.e., switching providers), as well as provider retention strategies, and the regulation rules that can be imagined for a more efficient economic model. The question of licensing the resource spectrum versus sharing it is also discussed, from the regulator perspective.

Chapter 4 discusses competition at the content and service level. Competition between content providers, and between service (here, security) providers, and its potential impact on the providers decisions, are investigated.

Chapter 5 is about the interactions between content/service and network service providers, with a special focus on the so-called *neutrality* issues: we explain and propose models to study some aspects of the vivid net neutrality debate, as well as the more recent search neutrality debate.

Chapter 6 concludes the dissertation, and also highlights some new research directions opened by recent technologies, namely Content Delivery Networks—already widely used, but not studied a lot from an economic perspective—and the promising Software-Defined Networking paradigm.

The dissertation summarizes several models and their main results. Most of them have been published in international journals or conference proceedings,

the corresponding articles are provided in the appendices for completeness since the main body of the document does not include the mathematical proofs.

Chapter 2

Modeling user behavior with nonatomic games

This chapter focuses on the lowest-level interactions we consider in this dissertation, namely the competition among users for network resource. Several of the mathematical models described here will be re-used in later chapters, where we study higher-level interactions (e.g., competition among providers).

We consider in this chapter *nonatomic* games [114], i.e., games where individual users have a negligible influence: only fractions of the population have an impact on congestion. This framework is well adapted to the study of networks in many contexts, where the number of users is very large and externalities (such as congestion, or positive network effects) stem from the cumulated influence of many users.

We first present the nonatomic routing game framework, that allows to precisely model the behavior of a large number of users, each one having a negligible weight. Then we show some results obtained for specific network contexts, three having an underlying nonatomic routing game model (routing game with rebates, cellular networks, choice of security solutions), and another one using the nonatomicity assumption to gain some insight when comparing management schemes for a peer-to-peer storage system. The contributions summarized here can be found in the references [43, 44, 54, 73, 75, 76, 120].

2.1 Nonatomic routing game models

Routing games refer to situations where users try to select a cost-minimizing route between some origin and some destination on a network, whose links have load-sensitive costs. The typical applications are in transportation networks, however those types of games can also be found in other domains, like supply chain management [93], or telecommunication networks.

Non-atomic routing games in general may involve complex network topologies and multiple origin-destination pairs (hence, different types of players).

There, the description of the game consists in:

- a graph (nodes and arcs) describing the network topology, let us denote by \mathcal{A} the set of arcs;
- a set \mathcal{K} of origin-destination pairs and their respective demands (cost-independent or elastic) $(d_k)_{k \in \mathcal{K}}$,
- a cost function ℓ_a for each arc $a \in \mathcal{A}$ that depends on the load (generally, only on the load of link a).

Due to the non-atomicity of players, the conditions for a situation to be an equilibrium can be expressed in a simple way known as *Wardrop's first principle*, stating that each player selects one of the cheapest paths:

“The journey times [*the costs*] on all the routes actually used are equal, and less than those which would be experienced by a single vehicle [*a single player*] on any unused route.”

Wardrop, 1952 [124]

The key intuition is again that if there were an available route with a strictly lower cost, then part of the traffic would have an interest to switch to this one. That principle characterizes the outcome of non-cooperative behavior in non-atomic routing games, that is called a *Wardrop equilibrium*. The socially optimal outcome, on the other hand, is expressed in terms of aggregated (or equivalently, average over the population) cost, through *Wardrop's second principle*:

“The average journey time [*the total cost*] is a minimum.”

Wardrop, 1952 [124]

Mathematically, for non-elastic (i.e., fixed) demands, the user equilibrium as described by Wardrop's first principle, corresponds to the following mathematical program, where \mathcal{P}_k is the set of possible paths (sets of links) for an origin-destination pair $k \in \mathcal{K}$ and y_p is the flow on path p :

$$\forall k \in \mathcal{K}, \forall p, q \in \mathcal{P}_k, \quad y_p \geq 0 \quad (2.1)$$

$$\forall k \in \mathcal{K}, \quad \sum_{p \in \mathcal{P}_k} y_p = d_k \quad (2.2)$$

$$\forall k \in \mathcal{K}, \forall p, q \in \mathcal{P}_k, \quad y_p > 0 \quad \Rightarrow \quad \sum_{a \in p} \ell_a(x_a) \leq \sum_{a \in q} \ell_a(x_a), \quad (2.3)$$

where $x_a := \sum_{p \in \mathcal{P}: a \subset p} y_p$ is the load on link $a \in \mathcal{A}$. Relations (2.1) and (2.2) simply represent the constraints that the flow on each path be nonnegative and demand of each origin-destination pair is satisfied. Relation (2.3) represents the Wardrop condition, namely, only paths with minimum total cost (the sum of the costs on the path's links) are used. Solutions of this system are called Wardrop equilibria: when the cost functions are continuous and nondecreasing, a solution exists and is essentially unique, i.e., even if there are several solutions the cost on each path is unique.

On the other hand, a desirable outcome, that is often called *social optimum*, would be one satisfying Wardrop's second principle, which can be expressed mathematically as finding $y = (y_p)_{p \in \mathcal{P}_k, k \in \mathcal{K}}$ satisfying (2.1)-(2.2) and minimizing the total cost

$$C(y) = \sum_{k \in \mathcal{K}} \sum_{p \in \mathcal{P}_k} y_p \underbrace{\sum_{a \in p} \ell_a(x_a)}_{\text{cost on path } p} = \sum_{a \in \mathcal{A}} x_a \ell_a(x_a), \quad (2.4)$$

with x_a the load on link a as before.

The research community has been quite active in the study of those routing games in the early 2000's, an interest ignited by the seminal results of Roughgarden and Tardos [105, 109] on the so-called Price of Anarchy [60] of those games. Indeed the loss of performance due to user selfishness—the total cost (2.4) at Wardrop equilibrium *versus* social optimum—is often bounded, with bounds independent on the network topology and the demand structure, but only dependent on the cost functions. From there, much research has been carried out in several directions, mainly by teams in Stanford University (investigating the impact of controlling part of the traffic [106], of removing some edges to improve performance [108], of adding prices on links [23]), GeorgiaTech (extending the price of anarchy results to more general functions [101]), Columbia University (analyzing the case of atomic users [24]), MIT (considering capacity limits on links [26], providing new proofs of the initial results [25], studying fairness versus efficiency tradeoffs [27, 115]), and the university of Hong Kong (considering elastic traffic [21]). The activity on those models has decreased recently, due to the complexity of the untreated cases, but nonatomic routing game models are now more and more used as the underlying model to study higher-level interactions such as competition among providers, as we will see in the next Chapters.

The rest of this section contains our contributions in this domain, that can be classified into four categories, according to the setting considered:

- Specific case of negative prices applied on link (rebates), that can correspond to subsidies provided by cities on public transport but may also have interpretations in telecommunications (Section 2.1.1);
- Application of existing results to specific telecommunication cases (cellular networks) and derivation of analytic expressions for prices leading to cost-minimal outcomes (Section 2.1.2);
- Modeling of user choices for a security solution, taking into account the trade-offs between the price of each solution, the risk of an attack, and the likeliness of attacks being successful (Section 2.1.3).

2.1.1 Application: Influencing routing games through rebates

While the possibility of adding monetary incentives (tolls) to links had been quite studied [107], in [75] we investigate the impact of *negative tolls*, i.e., sub-

sidies or rebates set by regulators, for example to incentivize the use of public transportation.

With the notations above and denoting by $s_a \geq 0$ the rebate on link $a \in \mathcal{A}$, users will select paths p with minimal values of

$$\sum_{a \in p} [\ell_a(x_a) - s_a]^+,$$

with $[x]^+ := \max(x, 0)$. Then the objective function for the system owner is a weighted sum of the total cost experienced by users and the monetary cost of the rebate to the regulator. More precisely, the regulator will want to minimize

$$C_\rho(s) := \underbrace{\sum_{i \in \mathbb{L}} x_i^s [c_i(x_i^s) - s_i]^+}_{\text{participants' perceived cost}} + \rho \underbrace{\sum_{i \in \mathbb{L}} x_i^s \min(c_i(x_i^s), s_i)}_{\text{cost of rebates}}, \quad (2.5)$$

which can also be expressed as

$$C_\rho(s) = \underbrace{\sum_{i \in \mathbb{L}} x_i^s c_i(x_i^s)}_{\text{participants' real cost}} + (\rho - 1) \underbrace{\sum_{i \in \mathbb{L}} x_i^s \min(c_i(x_i^s), s_i)}_{\text{cost of rebates}}, \quad (2.6)$$

for some parameter ρ representing the reluctance to spend on rebates.

Our objective (social cost) function hence explicitly considers the transfer payments to capture the cost of providing rebates. Instead, most of the earlier articles that studied the coordinating power of tolls and prices consider a social cost equal to the sum of costs for all participants, thus ignoring the costs and benefits of payments because they are transfers that stay in the system (see [10, 12, 63] for classical references; [23] is a notable exception that considers transfer payments as part of the social cost).

We consider a *Stackelberg game* in which the system owner (e.g., the city or the transportation authority) is the *leader* and the participants are *followers* [122]. In a first stage, the leader offers rebates in each arc; in a second stage, participants selfishly select arcs that have minimal cost, taking rebates into consideration.

Focusing on the modal choice problem, we characterize the optimal rebates in the case of affine cost functions and networks with multiple arcs that connect two nodes (the alternative modes of transportation are substitutes). In [75], we prove that:

- if the system owner values the perceived cost more than rebates (i.e., $\rho < 1$), then an optimal strategy for the leader is to refund each participant the perceived cost at each arc under a system optimal solution;
- if the system owner is more sensitive to the investment in rebates than to the perceived cost ($\rho > 1$), it will offer rebates in the modes that are under-utilized. We establish that the proportion of participants that actually receive a positive rebate is then upper-bounded by $1/\rho$.

For the special case of affine cost functions on links, we use our characterization of Stackelberg equilibria to provide a polynomial-time algorithm that selects the arcs where rebates should be offered, and computes the optimal rebates for those arcs. This enables us to derive an explicit formula for the resulting social cost, from where we compute the price of anarchy, expressed as a function of the predisposition of the system owner to offer rebates. The main conclusion is that when the system owner is willing to offer rebates, the resulting solution has low social cost. Conversely, when the system owner cannot afford to provide significant rebates, the resulting outcome is close to a Wardrop equilibrium. This is represented by the ratio of the user cost experienced at the Stackelberg equilibrium compared to the minimum user cost, that we prove to be upper-bounded by $\frac{4\rho}{3\rho+1}$ for $\rho > 1$:

- when ρ tends to 1 the rebate cost is not too large, so that the system owner can almost perfectly coordinate users who then experience a cost close to the optimal one;
- for large values of ρ , the system owner cannot afford to use rebates to coordinate users, hence our measure converges to the Price of Anarchy [107] that estimates the loss of efficiency due to the lack of coordination.

While the initial motivation for rebate-based schemes was in the context of transportation networks, the two following subsections present applications of the nonatomic game framework to telecommunication services.

2.1.2 Application to wireless cellular networks with heterogeneous users

The selfish behavior of users in networks can be regulated through incentive tools, such as taxation or penalties. The idea being that users select the cheapest path from their position to their destination node in the network, taking into account the cost (latency, or delay, that is sensitive to congestion) of the paths but also possibly some additional (monetary) costs imposed by the network manager. So that a proper definition of the price levels influences user choices. In the homogeneous case, i.e., when all users have the same sensitivity to the taxation, Beckmann et al. [10] showed that the so-called Pigovian taxes—applied on each link, and computed using the derivative of the cost functions of the links—produce a minimum-latency (delay) traffic routing (see [102]).

Cole, Dodis and Roughgarden [22] consider the case when users may perceive differently the relative costs of delay and prices. The authors were the first to study this setting, for a situation when all users have the same source and destination, with any network topology in between. For that scenario, it is shown that there exist prices so that the resulting user flow minimizes the average latency. Those results are generalized to the multicommodity setting (i.e., several source-destination pairs) in [55,56]. A constructive proof is given to show that prices inducing the minimum average latency multicommodity flow exist for both the cases of elastic (i.e., cost-dependent) and nonelastic demands.

Our analysis falls in that framework (with nonelastic demand): we consider a system with n heterogeneous wireless networks covering the same area. The users situated in the common coverage area of these networks seek for an Internet connection. We assume that they can easily handover from one network to another, thus choosing at every moment the most suitable one. Users select which network to connect to based on the QoS they experience and the prices charged.

We assume a total user demand D , coming from different applications. Since QoS requirements can vary depending on the applications used and on user preferences, the trade-offs between QoS and monetary cost shall differ, which we model through the sensitivity to the monetary cost (or equivalently, the ratio of the price sensitivity to the latency sensitivity). To simplify notations, without loss of generality we will treat a user running q applications with different requirements as q separate users, each one running one application. Therefore from now on we only evoke users, each one having a given price sensitivity.

Let us express the problem mathematically, using index i to refer to network i , $1 \leq i \leq n$, and exponent j to refer to user class j , $1 \leq j \leq m$. Users in class j have price sensitivity $\alpha^j \geq 0$ and the total demand from class- j users is denoted by d^j , so that $\sum_{j=1}^m d^j = D$. Each network i has a QoS-related cost function

$$\ell_i(x_i) = \begin{cases} (c_i - x_i)^{-1} & \text{if } x_i < c_i, \\ \infty & \text{if } x_i \geq c_i, \end{cases} \text{ where } x_i \text{ is the flow (cumulated throughput)}$$

on network i , and c_i the network capacity (we assume $D < \sum_{i=1}^n c_i$, i.e., the aggregated capacity is enough to treat all demand). Finally, all networks are owned by the same provider, which is aiming to minimize the total QoS-cost experienced, and can influence users behavior through charging a price τ_i on each network i .

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

$$C_i^j(f) = \ell_i(x_i) + \alpha^j \tau_i. \quad (2.7)$$

The provider owning all considered networks is interested in minimizing the social cost (or total cost) expressed as:

$$C(f) = \sum_{i=1}^n x_i \ell_i(x_i), \quad (2.8)$$

where $f = (x_1, \dots, x_n)$ is the flow distribution vector, with $\sum_{i=1}^n x_i = D$. Note the difference with the previous subsection when monetary exchanges were part of the objective: here we just focus on the coordinating power of prices, ignoring their possible impact on social welfare.

The problem is that users selfishly select a network minimizing their own perceived cost expressed in (2.7), hence the provider does not directly control the flows but can only indirectly influence them through prices.

Routing game interpretation

The setting described above is actually a routing problem, with a common source for all users (the common network coverage area), and one common destination (the Internet). Each user forwards his flow through one of n routes, which are the n networks, with a routing cost equal to the cost in (2.7), as depicted in Figure 2.1.

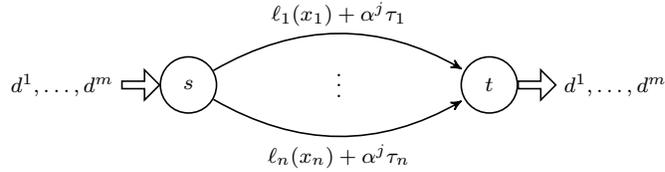


Figure 2.1: Logic representation of the network selection problem as a routing problem: the perceived cost on each route i depends on the load x_i and the price τ_i , but also on the user type j through the sensitivity α^j .

Computing optimal prices

We assume without loss of generality that $c_1 \geq c_2 \geq \dots \geq c_n$ and $\alpha_1 \leq \alpha_2 \leq \dots \leq \alpha_n$, i.e., network 1 is the most performant and class-1 users are the most QoS-sensitive (since they are the least price-sensitive).

The literature [22] already guarantees the existence of prices $(\tau_i)_{1 \leq i \leq n}$ leading to a user equilibrium minimizing the total cost 2.8, but no analytical expression is provided given the generality of the result. Our contribution [43, 44] is then in the analytical treatment of that specific game and the design of a simple algorithm to compute the optimal prices, as summarized below.

The expression for the optimal flows had been obtained previously [59], we recall it here for completeness.

Proposition 1 ([59]). The flows $(x_i^{\text{opt}})_{1 \leq i \leq n}$ minimizing (2.8) are unique and given by:

$$x_i^{\text{opt}} = \begin{cases} c_i - \frac{\sqrt{c_i}(\sum_{j=1}^k c_j - D)}{\sum_{j=1}^k \sqrt{c_j}} & \text{if } i \leq k, \\ 0 & \text{otherwise,} \end{cases} \quad (2.9)$$

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

$$c_i - \frac{\sqrt{c_i}(\sum_{j=1}^k c_j - D)}{\sum_{j=1}^k \sqrt{c_j}} \geq 0. \quad (2.10)$$

We then use that result to compute the prices leading to an efficient use of the network resources, even when users have different price sensitivities.

Proposition 2. The following prices are optimal:

$$\tau_{i+1} = \tau_i + \frac{\ell_i(x_i^{\text{opt}}) - \ell_{i+1}(x_{i+1}^{\text{opt}})}{\alpha^{s_i}}, \quad (2.11)$$

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

$$s_i := \min \left\{ j : \sum_{r=1}^i x_r^{\text{opt}} \leq \sum_{q=1}^j d^q \right\}. \quad (2.12)$$

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

We then have an algorithm to compute the optimal price: Proposition 1 should first be applied to obtain optimal flows, then (2.12) provides the value of s_i for each network i to be inserted into (2.11) so as to get the price value.

The intuition behind Proposition 2 and the algorithm is illustrated in Figure 2.2. We know how much (optimal) flow we want on each network, so we

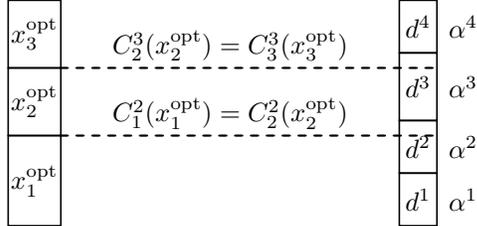


Figure 2.2: Example of user distribution among networks with optimal prices for the case $m = 4, n = 3$: class-1 (resp. class-4) users all attach to network 1 (resp. 3), while class-2 (resp. class-3) users are split among networks 1 and 2 (resp. 2 and 3).

use prices to drive users away from networks with too much flow. But we also know that the most price-sensitive users will be the first to react to prices, so it is easy to know which user classes will be ending using which network(s). On Figure 2.2, class-1 users would be the last ones to leave network 1 (which has the best QoS); but their demand is below x_1^{opt} , so some class-2 users should also select that network. The aggregated flow of classes 1 and 2 exceeds the optimal flow x_1^{opt} , so some class-2 users should use other networks. The Wardrop equilibrium conditions can then be used to express $\tau_2 - \tau_1$, since class-2 users should be indifferent between networks 1 and 2.

For this setting, we also carried out simulations in [44], showing that applying this static model (for which the analytical results are valid) to a dynamic setting (with users entering and leaving the system over time) performs quite well, even if the network owner has an imperfect knowledge of the repartition of classes in the total demand.

2.1.3 Application to users choosing a security solution

Another problem in telecommunications that we can represent as a nonatomic routing game is the choice of a solution to protect one's equipments against attacks. Cybercrime concerns colossal amounts of money, and is highly organized so that attacker efforts are rationalized to maximize the associated gains. This is why we model here an interesting negative externality effect of security architectures and systems, through the attractiveness for potential attackers: majority products are likely to be larger targets for hackers, and therefore become less attractive for consumers. Then, the choice of a particular system and security protection by the whole online population can now be considered as a congestion game.

The literature on network security involving game-theoretic models and tools is recent and still not very abundant. Some very interesting works have been published regarding the interactions between attacking and defending entities, where the available strategies can consist in spreading effort over the links of a network [16, 58] or over specific targets [32], or in selecting some particular attack or defense measures [15, 46]. In those references, the security game is a zero-sum game between two players only, and therefore no externalities among several potential defenders are considered.

Another stream of work considers security protection investments, through models that encompass positive externalities among users: indeed, when considering epidemic attacks (like, e.g., worms), the likeliness of being infected decreases with the proportion of neighbors that are protected. Since protection has a cost and users selfishly decide to protect or not without considering the externality they generate, the equilibrium outcome is such that investment is suboptimal [52] and needs to be incentivized through specific measures [65]. For more references on game theory applied to network security contexts, see [3, 84].

In contrast, the work presented here considers negative externalities in the choices of security software/procedures. Such situations can arise when attacks are not epidemic but rather direct, as are attacks targeting randomly chosen IP addresses. The interaction among users can then be modeled as a population game, that is a game where the user payoffs for a given strategy (here, a security solution) change as more users choose that same strategy [41].

In this model, we consider a very large population, where the extra congestion created by any individual user is negligible, hence again forming a nonatomic game. As stated before, nonatomic congestion games have seen recent advances for the case when all users are identical or belong to a finite set of populations [25, 56, 101, 107, 110], but we want here to encompass the larger attractiveness to attackers of "rich" users, compared to the ones with no valuable data online. Fewer results exist for those games [14, 91], even when user strategies only consist in choosing one resource among a common strategy set.

Moreover, in our model users undergo the *congestion cost* of the security solution they select (which depends on the congestion as well as on their particular data valuation) but also the *monetary cost* associated to that solution, which is the same for all users. As a result, following [90, 91] the game would

be called a *weighted congestion game with separable preferences*, and can be transformed into an equivalent *weighted congestion game with player-specific constants* [89] (i.e., the payoffs of users selecting the same strategy only differ through a user-specific additive constant). In general, the existence of an equilibrium is not ensured for such games when the number of users is finite [89–91]. In the nonatomic case, the existence of a mixed equilibrium is ensured by [114] and the loss of efficiency due to user selfishness is bounded [14], but the existence of a pure equilibrium in the general case is not guaranteed.

Our model considers a finite set \mathcal{I} of security solutions (each one on a given architecture). Users differ with the valuation for their data. When an attack is successful over a target user u , that user is assumed to experience a financial loss $v_u \geq 0$, which we call her data valuation. The distribution of valuations over the population is given by a cumulative distribution function F on \mathbb{R}^+ , where $F(v)$ represents the proportion of users with valuation lower than or equal to v .

We may not suppose that the support of F , that we denote by S_v , is bounded, but we assume that the overall value of the data in the population is finite, i.e.,

$$V_{\text{tot}} := \int_{S_v} v dF(v) < +\infty.$$

Security systems performance

In this model, we focus on direct attacks targeting some specific machines, which may for instance come from an attack-generating robot that randomly chooses IP addresses and launches attacks to those hosts. The attacks generated by such a scheme have to target a specific vulnerability of a given security system. As a result, the attacker has to select which security system $i \in \mathcal{I}$ to focus on.

If an attack is launched to a security system i , we consider that all machines protected by a system $j \neq i$ do not run any risk, while the success probability of the attack is supposed to be fixed, denoted by $\pi_i > 0$, on machines with protection system i . In other terms, the parameter π_i measures the effectiveness of the security defense.

The attacker point of view

Successful attacks bring some revenue to the attacker. Be it in terms of damage done to user data, or in terms of stolen data from users, it is reasonable to consider that for a given attack, the gain for the attacker is proportional to the value that the data had to the victim.

We then define for each provider $i \in \mathcal{I}$ the total value of the protected data, as

$$V_i := \int_{\text{users with prov. } i} v dF(v). \quad (2.13)$$

For an attacker, the expected benefit from launching an attack targeted at system i (without knowing which users are with provider i) is thus proportional to $\pi_i V_i$. We therefore assume that the likeliness of attacks occurring on system

i is a continuous and strictly increasing function $R_i(\cdot)$ of $\pi_i V_i$, and such that $R_i(0) = 0$.

User preferences

For a user u with data valuation v_u , the *total expected cost* at provider i depends on the risk of being (successfully) attacked, and on the price p_i charged by the security provider. That total cost is therefore given by

$$v_u \pi_i R_i(\pi_i V_i) + p_i, \quad (2.14)$$

that takes into account the price p_i for the security service, the valuation for the data to protect, the quality of the protection, and the likeliness of being attacked.

The variables of interest in the model are drawn in Figure 2.3, to summarize

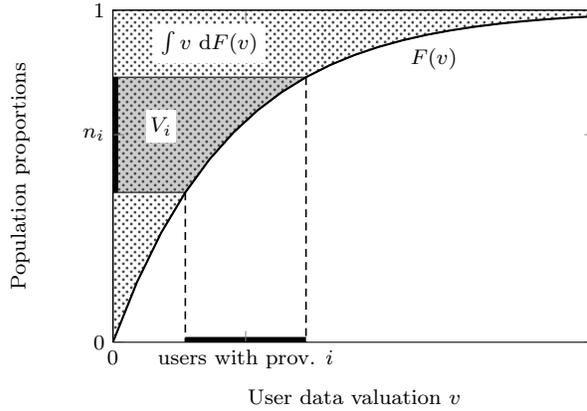


Figure 2.3: Values of interest in the security game model.

the interactions among the three types of actors in the model through those variables: for a particular provider $i \in \mathcal{I}$

- *users* care about the risk $(\pi_i R_i(\pi_i V_i))_{i \in \mathcal{I}}$ and the price $(p_i)_{i \in \mathcal{I}}$ when selecting a provider j with minimal total cost $v \pi_j R_j(\pi_j V_j) + p_j$;
- *attackers* focus on the target values (V_i) (balanced with the protection efficiencies (π_i));
- each *provider* i is interested in his market share n_i , through the product $p_i n_i$.

Results

In [73], we establish the *existence* and essential *uniqueness* (the corresponding repartition of value (V_i) among solutions is unique) of a pure equilibrium for our

model, as well as its *tractability* by proving that an equilibrium solves a strictly convex optimization problem.

Proposition 3. For any price profile $(p_i)_{i \in \mathcal{I}}$, there exists a user equilibrium, that is completely characterized by the valuation repartition \mathbf{V}^* , unique solution of the strictly convex optimization problem

$$\min_{\mathbf{V} \text{ feasible}} \sum_{i \in \mathcal{I}} \left(\int_{y=0}^{V_i} T_i(y) dy + p_i \underbrace{\left(\mathcal{N}(V_{[i]}) - \mathcal{N}(V_{[i-1]}) \right)}_{\text{Market share of prov. } i} \right), \quad (2.15)$$

where $V_{[i]} = \sum_{j=0}^i V_j$, and $\mathcal{N}(x)$ is the minimum proportion of the population whose aggregated value equals x (i.e., the proportion of “richest” users with total value x , that is computed using the distribution of valuations). Moreover, the value repartition at a user equilibrium is unique, and the user equilibrium is unique (unless for a zero-measure set of users) when all providers set different prices.

To the best of our knowledge, such proofs for nonatomic games had only been given for unweighted games [111, 112], with a finite number of different user populations; here we have a weighted game with possibly an infinity of different weight values, with the specificity that the differences in user congestion weights are directly linked to their user-specific valuations (in other terms, the valuation v of a user is both his sensitivity and his marginal contribution to congestion).

Those results can then be used to analyze the decisions made by security providers, in terms of pricing (how much to charge for their service, anticipating the resulting user equilibrium) and possibly of investment (how much to invest in improving the protection level—reducing the π_i), that we will present in Section 4.3.

2.2 Managing a peer-to-peer storage system

In this section, we focus on a specific telecommunication service, namely a *peer-to-peer storage system*, where participants can store data online on the disks of peers in order to increase data availability and accessibility. Due to the lack of incentives for peers to contribute to the service, we suggest and compare two approaches:

- the first one does not involve any money exchanges but enforces some reciprocity in providing the service: each peer’s use of the service is here limited to her contribution level;
- in the second one, we introduce a (monopolistic) system operator, who can buy storage space from some peers and sell it to other peers in order to maximize profit.

Using a noncooperative game model to take into account user selfishness, we study those mechanisms with respect to the social welfare performance.

Our model is not explicitly non-atomic since the number of users may be finite, but we make the reasonable assumption that individual users do not consider that they have any impact on the prices set by the monopolist, either because this impact is negligible (the case of nonatomic games) or because they are not aware of this impact. This assumption significantly simplifies the analysis, users then being simply price-takers.

In [76], we consider simple functions to describe users' willingness-to-pay for storing data on the system and perceived cost for offering storage space to the system, so that the resulting supply function (how much storage they are willing to offer for a given price) s_i and demand function (how much data they would store if the unit price is given) d_i for each user i are

$$s_i(p) = a_i[p - p_i^{\min}]^+, \quad (2.16)$$

$$d_i(p) = b_i[p_i^{\max} - p]^+, \quad (2.17)$$

where a_i, b_i, p_i^{\min} and p_i^{\max} are user-specific parameters. Figure 2.4 illustrates those functions, and shows the benefit for the user in the case of a manager buying resource at a unit price p^o and selling it at unit price p^s .

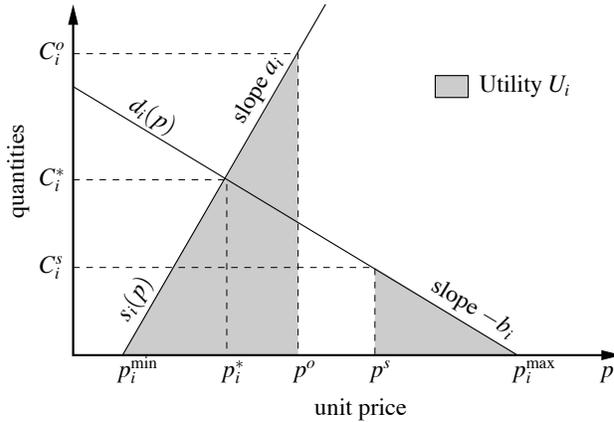


Figure 2.4: Reactions to prices and utility of a user i .

If the heterogeneity of the parameters p^{\min} and p^{\max} among users is limited, i.e., if

$$\max_i p_i^{\min} \leq \frac{p^* + \min_i p_i^{\min}}{2}, \quad (2.18)$$

$$\min_i p_i^{\max} \geq \frac{p^* + \max_i p_i^{\max}}{2}, \quad (2.19)$$

where $p^* := \frac{\sum_i a_i p_i^{\min} + b_i p_i^{\max}}{\sum_i a_i + b_i}$, then we show that symmetric schemes socially outperform profit-oriented pricing mechanisms if and only if

$$\frac{1}{4} \left(p^* - \sum_i \alpha_i p_i^{\min} \right) \left(\sum_i \beta_i p_i^{\max} - p^* \right) \geq \sum_i \omega_i (p_i^* - p^*)^2, \quad (2.20)$$

with the weights for all $i \in \mathcal{I}$: $\alpha_i := \frac{a_i}{\sum_j a_j}$, $\beta_i := \frac{b_i}{\sum_j b_j}$, $p_i^* := \frac{a_i p_i^{\min} + b_i p_i^{\max}}{a_i + b_i}$ and $\omega_i := \frac{a_i + b_i}{\sum_j a_j + b_j}$.

For the special case when all users have the same p_i^{\max} and the same $p_i^{\min} = 0$, then this gives the necessary and sufficient condition

$$\left(\frac{1}{\sum_i a_i} + \frac{1}{\sum_i b_i} \right) \sum_i \frac{1}{\frac{1}{a_i} + \frac{1}{b_i}} \geq \frac{3}{4}. \quad (2.21)$$

for symmetric schemes to socially outperform profit-oriented pricing mechanisms. Moreover, if the couples (a_i, b_i) are independently chosen for all users and identically distributed, then from the law of large numbers, symmetric schemes socially outperform profit-oriented pricing mechanisms if and only if

$$\frac{\mathbb{E}[f(a, b)]}{f(\mathbb{E}[a], \mathbb{E}[b])} \geq \frac{3}{4} \quad (2.22)$$

when the number of users tends to infinity, with $f : (x, y) \mapsto \frac{1}{1/x + 1/y}$.

Hence the economic analysis of both management solutions (without monetary exchanges but enforcing symmetry *versus* profit-driven mechanism) can help decide what scheme to prefer in terms of total value generated by the system.

2.3 Summary

This chapter highlights several types of contexts where a nonatomic game modeling is relevant in telecommunications. The nonatomicity assumption simplifies significantly the derivations, allowing us to reach analytical conclusions that give us general insights regarding the consequences of noncooperative behavior. Those results can then be used to study higher-level interactions, involving providers and possibly regulators, as will be seen in the next chapters.

Chapter 3

Competition among access providers

Access providers fight to attract users and make revenue. This chapter presents some of our contributions to the definition and mathematical analysis of such situations.

To study competition for customers, we need to model the decisions made by users regarding their choice of an access provider. Several ways to do that exist in the literature, we have defined and studied competition models using various approaches, that we summarize below, together with our main results.

3.1 Association models based on user utility

A first way to model the decision of users is to consider them as self-interested actors willing to maximize their individual utility. More precisely, for given provider choices made by the other users and given provider strategies (e.g., prices), each user has a utility associated to connecting to each provider, and is expected to select the provider yielding the highest utility.

We provide below examples with non-atomic users (i.e., individual users have negligible impact on congestion, as described in the previous chapter), so that the user equilibrium notion is the Wardrop equilibrium.

3.1.1 Competition on sent packet prices

Consider a set \mathcal{N} of providers implementing loss-based pricing: each provider sets a price per packet sent, and all sent packets are charged so that the perceived price per successfully transmitted packet increases when losses occur, as illustrated in Figure 3.1. Such a simple way to internalize congestion was initially proposed in [88]. In his model [88], Marbach deals with atomic users having to choose among several priority levels in a single network, and establishes the existence and uniqueness of an equilibrium among users. Here we

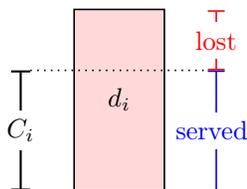


Figure 3.1: Loss model (during a slot): only C_i traffic units can be served among the d_i submitted. The $d_i - C_i$ remaining are lost, the lost units being chosen uniformly among the d_i .

rather consider non-atomic users facing several networks with different prices (but possibly different loss probabilities), and we add another stage, that of price competition among revenue-maximizing providers: instead of prices being fixed as in [88], here they are the result of strategic decisions by providers. In practice however, networking protocols rather tend to avoid losses than to send packets when the user wishes to pay for traffic. An alternative interpretation of the model, that also internalizes congestion without relying on losses, can involve a pricing per time unit (instead of per packet sent). If each user pays a given price per time unit, when there are too many users they have to stay longer and thus pay a higher price for the same data sent, which we represent with the model.

Let us denote by C_i the transmission capacity (in packets per time slot) of Provider i , and assume that losses occur when demand exceeds capacity on a network, in which case lost packets are randomly chosen (equal loss probability for all packets sent to a network). The perceived price \bar{p}_i per successful packet with Provider i is then

$$\bar{p}_i = p_i \max(1, d_i/C_i) \quad (3.1)$$

where p_i is the price charged per packet sent by Provider i , and d_i the total demand (number of packets sent per time slot) of that provider.

Users are assumed infinitesimal, and selfishly select the cheapest provider in terms of perceived price when they have a choice. Finally, demand is assumed elastic, i.e., the aggregated number of packets that all users wish to send (successfully) is a strictly decreasing function D of the perceived price. The function D is assumed continuous and such that $D(0) > \sum_{i \in \mathcal{N}} C_i$, and can stem from a distribution of valuations for the service among users and/or from individual elastic demands. We present two scenarios in terms of network topology.

Common coverage area. First consider that all users can reach all providers, as depicted in Figure 3.2. Then if we denote by d_i the total demand of Provider i , the Wardrop equilibrium conditions can be summarized by the following system

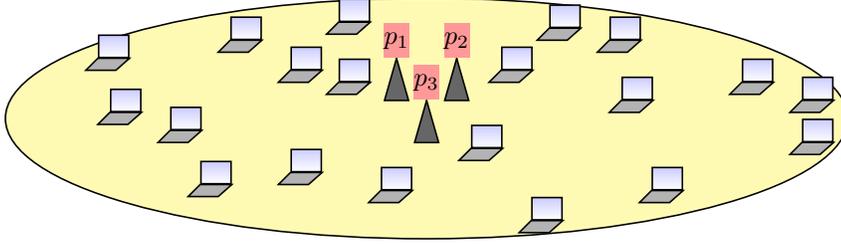


Figure 3.2: A network topology with all providers competing for the same set of users.

of equations:

$$\left\{ \begin{array}{l} \forall i \in \mathcal{N}, \quad \bar{p}_i = p_i \max \left(1, \frac{d_i}{C_i} \right) \\ \forall i \in \mathcal{N}, \quad \bar{p}_i > \min_{j \in \mathcal{N}} \bar{p}_j \Rightarrow d_i = 0 \\ \sum_{i \in \mathcal{N}} d_i = D(\min_{i \in \mathcal{N}} \bar{p}_i) \end{array} \right. \quad (3.2)$$

The first relation expresses the perceived price for each provider, based on (3.1); the second one comes from the Wardrop condition (only cheapest options are selected), and the last equation links total demand to the (common) perceived price of all chosen providers.

Assuming that providers set their price so as to maximize revenue, *anticipating* the outcome of the non-atomic game played by users (backward induction method), we established the following results (see [82] for proofs).

1. For any price vector $\mathbf{p} = (p_i)_{i \in \mathcal{N}}$ there exists a Wardrop equilibrium, and the corresponding perceived price of each provider is unique. This is illustrated in Figure 3.3, where the minimum perceived price (when users select the cheapest options) and the inverse demand function when the aggregated demand vary are plotted. A Wardrop equilibrium corresponds to an intersection point of those curves. If several providers set the same price \bar{p} and $D(\bar{p}) < \sum_{j: p_j \leq \bar{p}} C_j$ (case of Figure 3.3(b)), then users are indifferent between those providers, and any repartition of the demand $D(\bar{p}) - \sum_{j: p_j < \bar{p}} C_j$ among those providers such that none of them gets congested is a Wardrop equilibrium.
2. If among two outcomes leading to the same revenue, each provider prefers the one where he manages smaller demands, then when demand elasticity $\frac{p}{D(p)} \frac{dD(p)}{dp}$ is below -1 there is a unique Nash equilibrium of the price competition game: all providers set the same price $p^* := D^{-1}(\sum_{i \in \mathcal{N}} C_i)$ [82], where D^{-1} is the generalized inverse of D , i.e., $D^{-1}(y) := \min\{x \geq 0 : D(x) \leq y\}$.

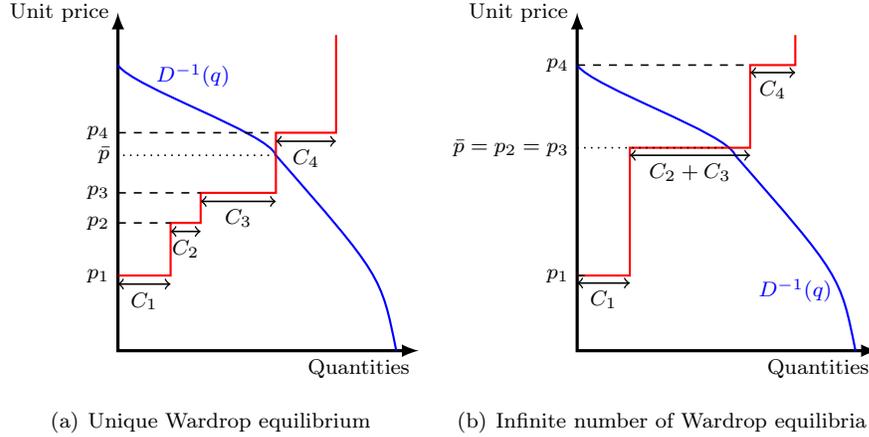


Figure 3.3: The common perceived price on all chosen providers when the total served quantity evolves is given by the stair-step function. At a Wardrop equilibrium, that perceived price is $\bar{p} = \min_{i \in \mathcal{N}} \bar{p}_i$ (intersection with the inverse demand curve).

3. The Wardrop equilibrium for that price vector $\mathbf{p}^* = (p^*, \dots, p^*)$ is unique and such that $d_i = C_i$ for each provider $i \in \mathcal{I}$. This situation is actually the one maximizing social welfare.

Different coverage areas. Consider now two wireless access points owned by competing operators 1 and 2 offering service over a given area, as depicted in Figure 3.4. Provider 2 operates in a subdomain of Provider 1, due to a

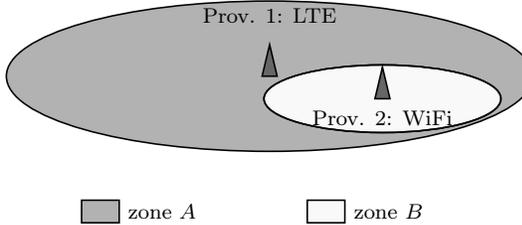


Figure 3.4: The competition situation: providers and topology [79].

limited range technology, e.g., a WiFi cell compared to an LTE one, covering a proportion $\alpha \in (0, 1)$ of the users in the LTE cell. Even if the number of providers has been reduced to two, this framework considerably complicates the analysis with respect to the common coverage area case treated just above. However we think the results provide interesting insights into the reasons why different prices can be observed.

Again, the reasoning at the pricing level is done by backward induction. A rigorous analysis for that model is carried out in [79], we summarize here the main results:

- For every price profile (p_1, p_2) with strictly positive prices, there exists a (non-necessarily unique) Wardrop equilibrium, and the corresponding perceived prices (\bar{p}_1, \bar{p}_2) are unique.
- If the demand elasticity $\frac{dD(p)/D(p)}{dp/p}$ is strictly smaller than -1 for all $p > 0$, then the pricing game played by providers has an infinity of Nash equilibria of the form $\{p_1 \in (0, p_1^*], p_2 \in (0, p_2^*]\}$, where

$$- \text{ if } \frac{C_1}{1-\alpha} \leq \frac{C_2}{\alpha},$$

$$p_1^* = D^{-1}\left(\frac{C_1}{1-\alpha}\right) \geq p_2^* = D^{-1}\left(\frac{C_2}{\alpha}\right); \quad (3.3)$$

$$- \text{ if } \frac{C_1}{1-\alpha} > \frac{C_2}{\alpha},$$

$$p_1^* = p_2^* = p^* = D^{-1}(C_1 + C_2), \quad (3.4)$$

where D^{-1} is again the generalized inverse of D . Despite the infinity of Nash equilibrium prices, all yield the same perceived prices $\bar{p}_i = p_i^*$, and the same revenues to providers, $R_i^* = p_i^* C_i$, $i = 1, 2$.

Let us take a step back here. Recall that the condition on elasticity implies that $pD(p)$ decreases with p , so providers should try to set low prices to maximize revenue. However, there is no benefit from lowering too much the price since providers can only “sell” their capacity (extra packets being lost, which affects the perceived price and reduces demand), so a natural objective is to sell all of one’s capacity. The case when $\frac{C_1}{1-\alpha} \leq \frac{C_2}{\alpha}$ corresponds to a *capacity per user* offered by Provider 2 alone in zone B larger than the one Provider 1 can offer in zone A only. Hence Provider 2 can sell all of his capacity C_2 in zone B for a perceived price $p_2^* = D^{-1}\left(\frac{C_2}{\alpha}\right)$, a perceived price that Provider 1 can never beat: assume $\bar{p}_1 < p_2^*$, then demand for successful traffic in zone A is $(1-\alpha)D(\bar{p}_1) > (1-\alpha)D(p_2^*) \geq C_1$, which Provider 1 cannot fulfill. As a result, Provider 1 cannot compete in zone B and leaves it to Provider 2, he then makes revenue by selling his capacity in zone A only, where he is a monopoly. On the other hand, when $\frac{C_1}{1-\alpha} > \frac{C_2}{\alpha}$ both providers compete in zone B , each one getting some demand.

Depending on the Nash equilibrium selected by providers, social welfare can go from 0 (when prices tend to 0) to its maximal possible value (when $p_i = p_i^*$ for $i = 1, 2$). That latter case is however more likely to be chosen by providers (e.g., if we add some small traffic management costs as we did for the common coverage area case).

It is worth noticing that in the examples above, a two-level game involving self-interested users and providers can lead to a socially optimal outcome. Because of the Wardrop conditions, a price war occurs when no provider is congested; but the price war stops because of congestion effects: decreasing further

one's charging price then does not bring any revenue improvement. Some more elaborate models of competition based on Wardrop user behavior can be found in [2, 53].

3.1.2 Discrete-choice models

In the previous (Wardrop) model, all users had the same comparative perception of the alternatives (here, the providers to select). Since this is not totally realistic, we can consider models that assume not all users are exactly the same: while some users are only looking for the cheapest option, others will consider it worth switching providers only if the price difference is large enough. Even more, some users can be willing to stay with the most expensive provider, even without clear evidence of his better quality: this can explain why incumbent providers often remain slightly more expensive than their competitors, without losing too much market share because of their better reputation. Such effects are due to rather subjective aspects, such as the perception of the overall service quality (e.g., through the reputation of the provider), and the attachment to the provider's brand (that can be affected by advertising).

The number of aspects other than price that can affect users' choices, and their somehow intangible nature, make it difficult to encompass each of them in a model. Therefore, a natural way to proceed is to *aggregate* all these unknown effects into one single value for each user and each provider, that represents the non-monetary benefit (or cost) that this particular user associates to this provider. Then we can still assume that each user makes a utility-maximizing choice, but now users are heterogeneous, so that they do not necessarily all prefer the same provider(s), even without congestion effects.

Such an approach is typical of *discrete choice* models, that are frequently used in economy [11]. To be more specific, the utility for a user u making the choice i (among a discrete set of options) is supposed to equal $v_i + \kappa_{u,i}$: the term v_i encompasses the objective aspects of the option (e.g., price, quality-of-service) and is the same for all users, while $\kappa_{u,i}$ is an unobserved user-specific value that is treated on the global level as a random variable. In most cases, it is assumed that the variables $(\kappa_{u,i})_{u,i}$ are all independent, and that for each option i the variables $(\kappa_{u,i})_u$ are identically distributed, so that the probability distributions of $\kappa_{u,i}$ for each option i completely characterize the model, and the subscript u can be omitted. Then from those distributions, one can compute the *probability* that a user selects option i for each i ; when the population is sufficiently large this corresponds to the proportion of users making that choice.

Let us now consider competing access providers, this time implementing a *flat-rate* pricing scheme, and assume congestion is not an issue (capacities are sufficiently large). The subscription price of Provider i is denoted by p_i , and each user has to decide which provider to subscribe to, if any. We suppose here that the average value for an ISP i is affected by his price, but also by some

average reputation factor $x_i > 0$, through the standard logarithmic relation

$$v_i = \alpha \log \left(\frac{x_i}{p_i} \right), \quad (3.5)$$

where $\alpha > 0$ is a sensitivity parameter. The logarithmic functional originally stems from psychophysics (the relationship between the magnitude of a physical stimulus and its perceived intensity is often logarithmic), and has recently been also observed in the context of telecommunications [104]. Remark that a null price p_i yields an infinite value v_i , so that a free option will always be preferred to one with charge. The average reputation factor can be interpreted as follows: if $x_1 = 1.2$ and $x_2 = 1$ for example, then on average customers accept to pay 20% more with Provider 1 than with Provider 2 (the reputation compensates for the larger price). Finally, we have to model the utility of the “no-provider” option, that we label by the index 0: we take the same form as for the other options, i.e., we denote by v_0 the average (negative) value of not having Internet access, and still consider a random part κ_0 in the individual utility of option 0.

Following the literature on discrete choice models [11], we assume that the user-specific random variables κ_i follow a Gumbel distribution of mean 0, i.e., their distribution satisfies $\mathbb{P}[\kappa_i \leq y] = \exp(-\exp(-y - \gamma))$, where $\gamma \approx 0.5772$ is Euler’s constant. Such an assumption is mainly made for mathematical convenience: it leads to a simple expression for the distribution of demand among providers. It can be seen as a way to justify the demand expression we obtain in (3.6) from a user utility model, instead of directly assuming its form.

It can then be proved (see [11]) that the probability of a user choosing the option $i \in \{0, 1, 2, \dots, n\}$ (n being the number of competing providers) equals

$$\sigma_i(\mathbf{p}) := \frac{(x_i/p_i)^\alpha}{\sum_{j=0}^n (x_j/p_j)^\alpha}, \quad (3.6)$$

with $\mathbf{p} = (p_0, p_1, \dots, p_n)$, $x_0 := 1$, and $p_0 := \exp(-v_0/\alpha)$ the “equivalent price” of option 0. We assume here that the number of users is large, so that the probability σ_i corresponds to the market share of Provider i for $i = 1, \dots, n$.

We can observe the effect of the sensitivity parameter α : when α tends to infinity users only focus on the cheapest option(s) (the one(s) with the smallest p_i/x_i). On the other hand, α going to 0 leads to a uniform repartition of users among all alternatives.

Now let us study the price competition among two providers, labelled 1 and 2: assuming without loss of generality that the total mass of users is 1, the revenue of Provider i equals

$$R_i(p_0, p_1, p_2) = p_i \sigma_i(p_0, p_1, p_2) = \frac{p_i (x_i/p_i)^\alpha}{(x_1/p_1)^\alpha + (x_2/p_2)^\alpha + 1/p_0^\alpha}. \quad (3.7)$$

Let us immediately treat the particular case of null prices: if a provider sets a null price, then he makes no revenue but also attracts all users unless his competitor sets a null price, hence preventing that competitor from making revenue.

Thus null prices $(p_1, p_2) = (0, 0)$ constitute a Nash equilibrium, however a peculiar one: as soon as the opponent sets a strictly positive price, it is strictly better for a provider to also fix a strictly positive price than to set a null price. Hence providers should try to avoid that equilibrium, when possible. As illustrated later in Figure 3.5, that equilibrium is actually unstable when $\alpha \in (1, 2)$. When $\alpha > 0$ it is the only equilibrium, i.e., a price war situation occurs.

We now look for equilibria different from $(0, 0)$. An equilibrium price profile (p_1, p_2) is such that each provider plays a best response to the price of his opponent; we therefore compute here the best-response functions.

From (3.7), we obtain for all strictly positive prices (p_1, p_2) ,

$$\frac{\partial R_i}{\partial p_i} = \sigma_i(1 - \alpha(1 - \sigma_i))$$

with σ_i given in (3.6). Since σ_i is a probability and is strictly positive when prices are non-zero, the case when $\alpha \leq 1$ leads to infinite best-response prices (i.e., maximizing R_i given $p_j, j \neq i$). This somehow corresponds to a small price elasticity of demand: a small α means that user choices are not much affected by prices, hence raising prices improves revenues since only a small proportion of users switch choices.

From now on, we make the more realistic assumption that $\alpha > 1$. For $i = 1, 2$ and $j \neq i$, the maximization of R_i in terms of p_i for a fixed $p_j > 0$ leads to $1 - \alpha(1 - \sigma_i) = 0$: indeed $\frac{\partial R_i}{\partial p_i}$ has the same sign as $1 - \alpha(1 - \sigma_i)$, which is decreasing in p_i from (3.6). We can then compute the best-response function of Provider i :

$$\text{BR}_i(p_j) = x_i \left((\alpha - 1) \left(\left(\frac{x_j}{p_j} \right)^\alpha + \left(\frac{1}{p_0} \right)^\alpha \right) \right)^{-1/\alpha}. \quad (3.8)$$

An example of those best-response functions is plotted in Figure 3.5 for two different values of the sensitivity parameter α . We remark that for $\alpha > 2$ there is no equilibrium with strictly positive prices. Indeed, such an equilibrium would correspond to $\sigma_1 = \sigma_2 = 1 - 1/\alpha$, which is only possible if $\alpha \leq 2$ (recall that $\sigma_1 + \sigma_2$ is the probability of users selecting no provider, thus $\sigma_1 + \sigma_2 \leq 1$). When $\alpha \in (1, 2]$, solving the system $\begin{cases} p_1 = \text{BR}_1(p_2) \\ p_2 = \text{BR}_2(p_1) \end{cases}$ yields a unique solution (as illustrated in Figure 3.5(a)). That solution (p_1^*, p_2^*) is such that $\sigma_1 = \sigma_2 = 1 - \alpha$, hence $x_1/p_1^* = x_2/p_2^*$, which gives

$$p_i^* = x_i p_0 \left(\frac{2 - \alpha}{\alpha - 1} \right)^{-1/\alpha}, \quad i = 1, 2.$$

We find here again the condition that $\alpha \in (1, 2]$. Interestingly, we also observe that equilibrium prices are proportional to the ‘‘reputation’’ factor x_i , i.e., $p_1^*/p_2^* = x_1/x_2$. Hence the relative operator prices directly reflect their relative reputations. Similarly, since equilibrium market shares are equal ($\sigma_1 = \sigma_2 = 1 - 1/\alpha$), providers’ revenues are also proportional to their reputation factors: $R_i = x_i \sigma_i = x_i(1 - 1/\alpha)$.

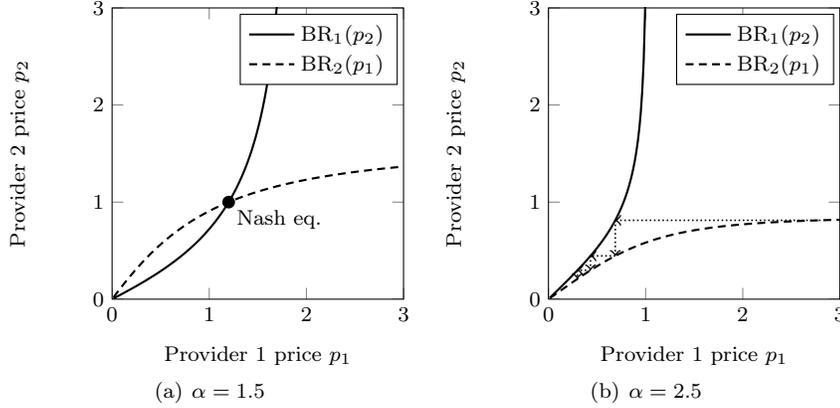


Figure 3.5: Best-response functions for the price competition, with $p_0 = 1, x_1 = 1.2, x_2 = 1$.

Finally, if $\alpha > 2$ we remark that $\text{BR}_i(p_j) < p_j \frac{x_i}{x_j}$, so that successive best responses lead to a strictly decreasing sequence of ratios p_i/x_i , and eventually to null prices as illustrated in Figure 3.5(b).

The discrete choice model presented above is developed in [30] to analyze the relations between ISPs in terms of traffic exchange. The main conclusion suggested by the numerical study in [30] is that regulating traffic exchange prices (or transit prices) is not necessary: it is sufficient to let ISPs agree on those transit prices, just imposing that an agreement be found (i.e., forbidding disconnection among ISPs).

To reach such a conclusion, we assume that ISPs decide the transit prices through negotiation, the result being the Nash Bargaining Solution [95]. The result depends on the so-called “disagreement point”, that is the outcome when no agreement is reached among ISPs. Our results in [30] suggest that when the disagreement point is a disconnection among ISPs, the negotiation results in a non-satisfying point (favoring the largest ISP over the smallest one), while the solution is near to the social optimum when the disagreement point is a “free peering” situation (i.e., null transit prices). Hence our suggestion that the regulation would be set to a minimum, just imposing free peering in case ISPs do not agree on transit prices.

3.2 Dynamic models

The telecommunication ecosystem is an extremely fast-changing environment. A salient example is the *churn* phenomenon (users switching providers) in mobile markets: yearly migration rates can often reach 25% [125]. In that context, it makes sense to take that dynamicity into account in the economic model.

We therefore propose here an analysis where user behavior over time (more precisely, the operator he subscribes to) is modeled as a random process. More precisely and to keep things simple, we consider a *finite-state continuous-time Markov chain* [97], as is done in [71]:

- the situation of the user is represented through a *state* (State i if the user is with Provider i , State 0 if the user foregoes the service);
- after some exponentially-distributed random time depending only on his current state i and the destination state j (through the parameter—or transition rate— λ_{ij}), the user switches providers (or goes with none) and finds himself in State j .

Let us consider a specific case, where two providers compete by playing on price only, price being indeed one of the most relevant churn determinants [19, 57, 103]. Then the churning rates are assumed to depend on the price vector $\mathbf{p} = (p_1, p_2)$, with p_i the price set by Provider i . We therefore model the churning behavior of a user with the Markov chain displayed in Figure 3.6. Using standard

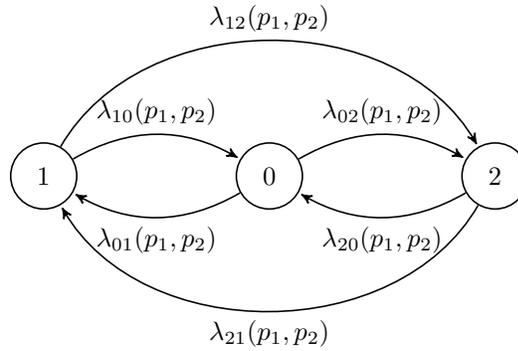


Figure 3.6: Continuous time Markov chain model for user switching behavior. State $i = 1, 2$ corresponds to the user being with Provider i , while State 0 means that the user chooses no provider.

Markov chain analysis, we can easily prove the existence and uniqueness of a *steady-state probability distribution*, i.e., a vector $\boldsymbol{\pi} = (\pi_i)_{i=0,1,2}$ of probabilities such that the proportion of time the user spends in State $i = 0, 1, 2$ tends to π_i as time passes. A sufficient condition is that $\lambda_{ij}(p_1, p_2) > 0$ for all i, j . When all users follow independently the behavior described in Figure 3.6, π_i also represents the stationary proportion of the population in State i (the average market share of Provider i , for $i = 1, 2$). Those steady-state probabilities then directly give the average per-user revenue that a provider $i = 1, 2$ can expect: that revenue simply equals $R_i := p_i \pi_i$. Since the probability vector $\boldsymbol{\pi}$ depends on the price vector $\mathbf{p} = (p_1, p_2)$, the revenue (utility) of each provider depends on both prices, hence a non-cooperative game on prices.

We study that game as a Stackelberg game, with providers as leaders, and users as followers (adapting to the leaders' actions through transition rates, the outcome being the steady-state probabilities). For this model, we cannot prove in general the existence or uniqueness of a Nash equilibrium. When transition rates are simple functions of prices, we may find the form of the Nash equilibria analytically, or decide to perform a numerical study, with some more realistic transition rates. Both approaches were presented in [71].

We only provide here some illustrative results from a numerical study (see [71] for details). The considered transition rates are of the form $\lambda_{ij} = e^{\beta p_i/p_j}/\gamma_i$: γ_i represents some reputation effect (reluctance to leave this provider), and β is a sensitivity to prices. Let us have a look at the impact of the parameter β (users' sensitivities to price differences). Nash equilibrium prices (p_1^*, p_2^*) and the resulting user repartition among the three states when β varies are displayed in Figure 3.7. We observe (similarly to the attraction

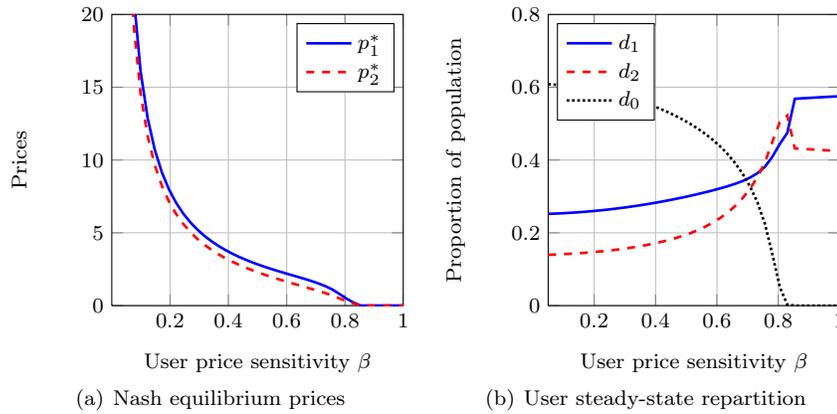


Figure 3.7: Influence of the price sensitivity factor β , with transition rates $\lambda_{ij} = e^{\beta p_i/p_j}/\gamma_i$ and $\gamma_0 = \gamma_2 = 1$, $\gamma_1 = 2$, $p_0 = 1$ [71].

model summarized by Relation (3.6)) that above a given threshold (around 0.85 here), user price sensitivity is such that providers engage in a price war and prices tend to 0. Below that threshold, prices decrease when sensitivity β increases, as expected, which results in more users selecting one of the providers (see Figure 3.7(b)).

Another aspect of the model worth considering is the asymmetry among providers, reflected by the likeliness γ_i to remain with Provider i . In Figure 3.8 we keep $\gamma_2 = 1$, and vary γ_1 from 1 (symmetric providers) to 5 (strong advantage for Provider 1). The Nash equilibrium p_1^* increases as expected, Provider 1 taking advantage of his higher reputation to set larger prices. The effect on price p_2^* and on market shares is less obvious; for the parameters chosen we observe that the better reputation of Provider 1 induces a price reduction of his

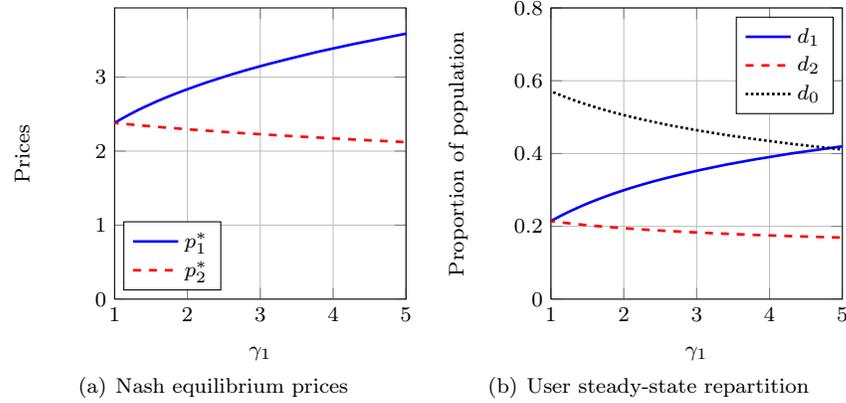


Figure 3.8: Influence of the asymmetry among providers.

competitor, but also a reduction in the competitor's market share, illustrating the importance of this parameter γ_1 for both providers.

We believe such analyses help understand the market: it is indeed observed that not all providers set the same price (even for comparable services), and in general the incumbent is slightly more expensive than the newcomers, benefiting from a better reputation.

Building more developed dynamic models. As we claimed before, one of the main challenges faced by mobile operators is to retain customers. This may imply the use of unfair practices such as delaying the migration process [94]. We include such practices in a dynamic model in [70], through a state i' representing the fact that the user has expressed the will to leave Provider i but is being delayed, as represented in Figure 3.9. From that state, three events can occur:

1. the provider finally releases the client (after some time depending on the retention policy of the provider);
2. the regulator intervenes (possibly when alerted by the client) to release the client, and imposes a sanction fee to the provider;
3. the client decides to give up his idea of leaving the provider, finally deciding to stay.

The two first alternatives lead to the user leaving the provider, but in the second case an additional fee is paid by the provider. Also, notice that while the client is in State i' , he continues to pay the subscription price to Provider i . From the point of view of Provider i , the trade-off in the retention policy is therefore between the gain in terms of subscription revenues due to retention (from users staying in State i' for a while, and deciding to stay with Provider i), and the

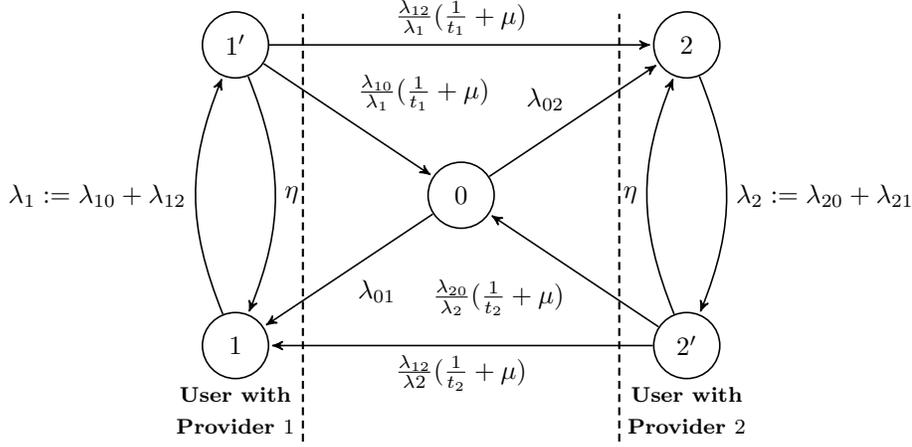


Figure 3.9: A dynamic model for user behavior with retention from providers (retention states: 1' and 2').

sanction fees imposed by the regulator when informed. If the sanction fee is not too large and users are quite patient, retention policies can be beneficial to providers, which can explain the increasing number of sanctions imposed by regulators to stimulate competition in the mobile market in the 2000's.

It is noticeable that the state 1' does not contain all the information about the user: the model “looses” the aimed state (0 or 2) by the user when he took the decision to leave Provider 1. However this is not a problem for the study, since that information is irrelevant for the metrics we compute: we could similarly consider that users decide to leave Provider 1 after some time (exponentially distributed with rate λ_1), and choose whether to go to Provider 2 or to the no-provider state only when released by 1. The same reasoning holds for Provider 2 and State 2'.

To analyze such a model, we first notice that the Markov chain has a unique steady-state distribution, and denote by π_s the steady-state probability of State s . Then it is easy to express the (per user and per time unit) revenue R_1 of Provider 1 in terms of those probabilities: that revenue equals

$$R_1 = \underbrace{p_1(\pi_1 + \pi_2)}_{\text{income}} - s_0 \underbrace{\pi_2 \mu}_{\text{sanction frequency}},$$

where p_1 is in monetary units per time unit, and s_0 (the sanction price) in monetary units. Considering the counterpart expression for Provider 2, one can study the non-cooperative game played on prices and retention times, after selecting the form of the transition rates in terms of those strategic variables. We performed that study numerically in [70], in order to estimate sanction levels that are sufficient to prevent providers from retaining users.

3.3 Providers competing in multiple-time-scale decision games

In practice, operators have a lot of decisions to make, that involve different time horizons. Let us consider a wireless access provider for example: the basic resource needed to provide some service are spectrum, access points, and backbone network.

Wireless licenses are typically sold for a decade-long period; similarly setting access points involves some long-term estate lease agreements and some significant equipment investments (we also include backhaul links—to ensure the connexions between access points and the backbone—in this time category). As regards backbone connectivity, the decision can be seen both as a long-term one (when the operator builds or upgrades his own backbone network) or a shorter-term one (when the operator outsources that part to another company).

Decisions on applicable policies are more flexible: operators can change—within the regulatory constraints—their prices, and/or the way they deal with QoS provisioning (e.g., by setting priority classes) on shorter time bases.

Finally, users base their decisions (to subscribe or not, and to which provider) on all those aspects: pricing policy, provided services (and their QoS), and coverage areas. As we saw before, the market of mobile users is very volatile, and regulation also exacerbates competition by favoring churn among providers. This therefore shall be the smallest-time-scale decision level.

Given that description, how should a provider bid in the spectrum auction? Also, what strategy should be set in terms of access points, or backbone transmissions? And finally, what QoS and pricing policies to implement? All those decisions also have to take into account the *competitive* context, i.e., providers are playing a noncooperative multi-level game.

Of course, there is a lot of uncertainty when making long-term decisions: about the acceptance of new services in the next 10 years, even about the nature of the services that might appear, etc. However those decisions can be based on the backward induction reasoning, where at each level the utility functions considered are given by the (Bayesian) equilibrium—if any—of the level below. The two last time scales of the game (price/QoS strategies, and association game for users) are the ones we have considered in the previous section, now the outcome from those interactions are assumed to be anticipated by providers making their higher-level investment decisions.

We summarize here a model we developed in [86], adding a third level on infrastructure and license investments to the price game and user decisions levels. The model relies on the Wardrop equilibrium notion for the user level, then on the Nash equilibrium notion for the pricing level (intermediate time scale) and for the investment level (highest time scale). The initial question addressed from the point of view of the providers takes the form “What are the technologies worth investing in?”, but answering it implies answering the following series of questions, getting lower and lower in the game levels:

- What technologies will my competitors propose?

3.3. PROVIDERS COMPETING IN MULTIPLE-TIME-SCALE DECISION GAMES 39

- Given the set of proposed technologies, what prices shall be set by each provider?
- For a given price vector and set of technologies offered, what are the user repartition and the corresponding revenue for each provider?

The capacity aspects are not treated here, since it is assumed that the transmission capacity is fixed for each technology. However, it is worth mentioning that in other works, capacity investments and pricing decisions are taken simultaneously, still considering some Wardrop equilibrium for the subsequent choices of users [53].

Let us denote by \mathcal{T} the set of technologies that are likely to be proposed to users (3G, 4G, WiFi, ADSL,...), and by \mathcal{N} the set of providers. The largest time scale decision for each provider $i \in \mathcal{N}$ therefore consists in choosing a subset $\mathcal{T}_i \subset \mathcal{T}$ of technologies to operate, weighing costs and expected revenues.

In terms of prices at the intermediate level/game, we assume a technology-transparent scheme, i.e., each provider i will offer a given price p_i (say, per month) to grant access to users on any of the available technologies in \mathcal{T}_i . This type of offer is more and more common, and goes in the direction of a simplification from the users' point of view, in a context of a multiplication of mobile devices and access technologies: users do not need (and in most cases, do not want) to know which specific technology is used for what usage and with what device, they just want their services to be available. The specific technology choice may be made by the device (remaining transparent to the user), or by the user himself; this does not affect our model as soon as the device tries to select the best option. Due to congestion effects, not all users shall select the same provider or technology: two subscribers of a provider will pay the same price but may be directed to two different technologies (WiFi and LTE for example, because one becomes more congested otherwise), in order to lead to a better price/QoS tradeoff for all of them. This is reflected by the Wardrop equilibrium conditions, implying that all chosen pairs (*operator, technology*) will have the same total cost, also below the total cost of the other options. That total cost, that we also call *perceived price* and denote by $\bar{p}_{i,t}$, is modeled as the sum of the subscription price p_i and a monetary-equivalent QoS-based cost $\ell_{i,t}$ for each operator i and technology $t \in \mathcal{T}_i$ depending on the load on that technology. To be more specific, we distinguish between technologies with *licensed spectrum* (such as 3G) from technologies with *shared spectrum* (such as WiFi): in the former case, congestion only comes from the demand level $d_{i,t}$ on the corresponding (*operator, technology*) pair (i, t) , while in the latter case interference comes from all users on the same spectrum (technology t), hence congestion depends on $\sum_{j \in \mathcal{N}: t \in \mathcal{T}_j} d_{j,t}$:

$\forall i \in \mathcal{N}, t \in \mathcal{T}_i,$

$$\bar{p}_{i,t} = \begin{cases} p_i + \ell_{i,t}(d_{i,t}) & \text{if } t \text{ is licensed,} \\ p_i + \ell_t(\sum_{j \in \mathcal{N}: t \in \mathcal{T}_j} d_{j,t}) & \text{if } t \text{ is a shared-spectrum technology,} \end{cases} \quad (3.9)$$

where $\ell_{i,t}$ and ℓ_t are continuous and strictly increasing functions reflecting the congestion cost of the technology t .

Finally, all users prefer the cheapest option—whose total price will be denoted by $\bar{p} := \min_{i \in \mathcal{N}, t \in \mathcal{T}_i} \bar{p}_{i,t}$, but the aggregated demand level is naturally assumed to decrease with \bar{p} , through some continuous demand function, as we previously did in Subsection 3.1.1.

Using very general results on nonatomic routing games [1], we can establish the existence of a Wardrop equilibrium for the lowest game level (users choosing a provider and a technology), and the uniqueness of the perceived price on each option (i, t) . The uniqueness of the Wardrop equilibrium (i.e., of all values $(d_{i,t})_{i \in \mathcal{N}, t \in \mathcal{T}_i}$) is guaranteed only if we fix a rule regarding demand repartition on each shared-spectrum technology among several providers with the same price; for example an even repartition can be assumed.

The study of the pricing game among operators at the intermediate level (using the above Wardrop equilibrium) is more involved, and in the general case the existence of a Nash equilibrium—or its uniqueness—cannot be proved. However the numerical study carried out in [86] (with two competing providers, linear decreasing demand, QoS-related costs given by the delay in an M/M/1 queue) exhibited unique Nash equilibria with positive prices for the pricing game with most technology choices $(\mathcal{T}_1, \mathcal{T}_2) \in \mathcal{T} \times \mathcal{T}$, actually for all cases where the shared-spectrum technology is proposed by only one provider. In the other cases, a price war occurs to attract all users of the shared-spectrum technology, leading prices to zero.

Plugging those equilibrium points into a matrix summarizing the total net benefit of each provider (including subscription revenues and license and infrastructure costs) for each combination $(\mathcal{T}_1, \mathcal{T}_2)$ of technologies then provides us with a two-player game in normal form for the choice of technologies at the largest time scale: we can indeed obtain results as exemplified in Table 3.1, where the terms in the payoff matrix are of the form “ $U_1; U_2$ ”, the value being in bold when the corresponding strategy is among the provider’s best-responses to the competitor’s choice (see [86] for the particular numerical values considered and detailed justification). Table 3.1 highlights the effect of regulatory measures on the outcome of the game: by appropriately setting the license prices for example (that act as additive constants for strategies involving each technology), the technological investment game can lead to a totally different Nash equilibrium. It is then up to the regulator to decide which outcome should be favored (for example, the one maximizing social welfare), and to set license prices accordingly. In the present case, reducing the 3G license price for Provider 1 may lead to both providers offering that technology, and an improved social welfare. Such a reasoning implies considering one additional game level, that of the regulator fixing license prices as the leader in a Stackelberg game.

1 \ 2	\emptyset	3G	4G	3G,4G	WiFi	WiFi,3G	WiFi,4G	WiFi,3G,4G
\emptyset	0;0	0;19	0;26	0;38	0;22	0;37	0;41	0; 48
3G	14;0	12;17	11;22	8;31	12;20	9;32	8;35	6; 41
4G	26;0	22;15	20;20	17;29	22;19	19;30	17;33	14;37
3G,4G	32;0	27;13	24;17	18;23	27;17	21;25	18;27	12; 29
WiFi	22;0	20;17	19;22	17; 32	0;0	-	-	-
WiFi,3G	32;0	27;14	25;19	20; 26	-	-	-	-
WiFi,4G	41;0	35;13	33;17	27;23	-	-	-	-
WiFi,3G,4G	43;0	36;11	32;14	24; 17	-	-	-	-

Table 3.1: Net benefits matrix when Provider 1 is positioned on WiFi only, while Provider 2 is already positioned on WiFi and 3G (hence has smaller 3G license and infrastructure costs) [86]. Values are in hundreds of euros per month and per cell, the symbol “-” indicates that the pricing game has no Nash equilibrium with strictly positive prices.

3.4 To license or not to license resources?

The previous example illustrates the importance of regulation (through license prices) on the technologies implemented by operators. But the regulator can take even more drastic decisions, by deciding *whether* to license some part of the spectrum, or to leave it unlicensed and accessible for free by operators. The tremendous success of WiFi, and the economic growth it allowed, talks in favor of developing license-free spectrum usage: this is among the objectives of the upcoming so-called *incentive auction* in the US. Moreover, with licenses some spectrum bands may be underused while others are congested; sharing the capacity for a more efficient use of the scarce resource may be beneficial to society.

Let us consider a model inspired from the one of Subsection 3.1.1, but mixing licensed and unlicensed spectrum usage. More precisely, consider two providers, each provider $i = 1, 2$ owning some licensed bandwidth to serve up to a demand C_i . In addition, some part C of the spectrum is unlicensed, and can be used by operators when their demand exceeds the capacity of their licensed bands. We assume that this unlicensed spectrum is shared among providers, proportionally to their excess demand as illustrated in Figure 3.10.

Consider the same pricing model as in Subsection 3.1.1, with users being charged for all sent packets and thus perceiving a congestion-sensitive cost per successful packet sent (see Equation (3.1)). One can then analyze the price competition among providers: the best-response correspondences, when plotted (see [78]), highlight a unique Nash equilibrium with non-zero prices. Also, the impact of the amount of unregulated spectrum can be studied; the main welfare metrics when the proportion of unlicensed spectrum increases are plotted in Figure 3.11.

It appears that unlicensing spectrum favors users at the expense of providers since the shared spectrum exacerbates price competition (lowering one’s price increases one’s demand and thus one’s share of the unlicensed capacity). For

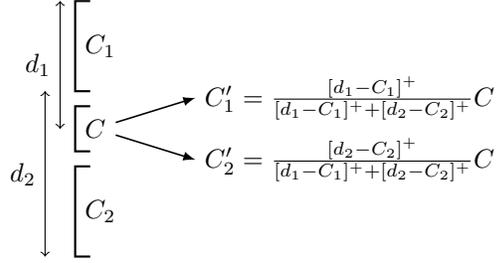


Figure 3.10: Sharing of the unlicensed spectrum capacity C [78]: d_i is the demand of Provider i , and $x^+ := \max(x, 0)$.

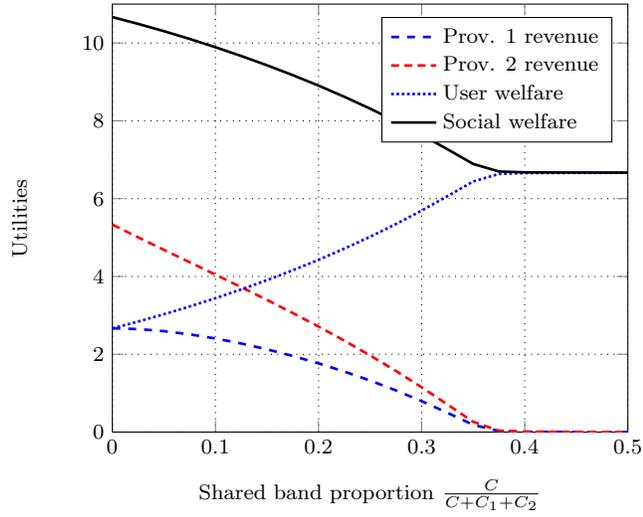


Figure 3.11: Utilities at Nash equilibrium when the share of unlicensed spectrum varies, the ratio C_2/C_1 being kept constant at 2 [78].

this model the overall effect of unlicensing in terms of social welfare is negative, but remember that the model is quite specific and does not encompass some WiFi-like services such as free Internet access, that are proposed more and more widely.

3.5 Summary

This chapter has highlighted several situations where service providers compete to attract users. My contributions in those contexts are in the modeling of those different settings (user churn among providers, shared versus licensed spectrum, technology investments) using various types of mathematical approaches for user decisions and the resulting demands. In each case, the models help gain insights about the outcome from selfish behavior (of providers and users), and about the need for regulation.

My publications related to this chapter are [9, 42, 45, 70–72, 77–79, 82, 85–87].

Chapter 4

Interactions among content or application service providers

4.1 Introduction

While the previous chapter was discussing the competition between network service providers, which could be access network providers, or transit providers needing (or required) to cooperate to deliver traffic to destination, the present chapter is focusing on the competition at the content and application service provider level. We are going to see that the models can be formulated in a very general way, hence closely related to what we have described in the previous chapter for access network providers. Indeed, the main driver for customer choices is the price, but some notions of quality (of service) and reputation, among others, can or need to be dealt with too, leading to similar models for customers' service provider choice (following Wardrop principle, or some discrete choice or stickiness models). Those choices will be briefly recalled in next section. We thus have multilevel Stackelberg games with providers playing first on price and sometimes at an even higher level content/service investment, anticipating the reaction of users to any strategy profile in terms of consumption.

The categories of content and application service providers which we have in mind are mainly (but are not limited to): (i) content providers such as news web sites for example competing on the content relevance and quality, design of the site, awareness and attractiveness through advertisement; (ii) online shops with similar characteristics; (iii) content delivery networks (CDNs) who have to attract content providers in terms of price too, but also in terms of the quality of service for users through investments in capacities strategically located; (iv) similarly cloud service providers; (v) service applications such as search engines competing for keyword searches through relevance of the results, leading to more revenues from sponsored links; we will describe security providers as an illustration. As said above, most of those competitive contexts can be encompassed in a general Stackelberg game which will be briefly recalled in the next section,

so we will afterwards focus on specific and arbitrarily chosen cases, by injecting some characteristic modeling properties in the model.

It is worth mentioning that competition at the content/application service level has risen with the so-called dot-com bubble in the late 90s, when Internet-based services or applications boomed. The dot-com crash occurred in 2000-2001, slowing the development of applications, bringing it back to a more realistic rate. It is interesting to note that the goal of companies at their start was (and still is for new entrants) to develop their base of customers as much as possible in order to get a chance to survive, even if this is at the expense of constant financial losses in the first years; typical examples are Google and Amazon. As a maybe unexpected consequence of the crash, some surviving companies managed to get a dominant position on their market, leading to some less competitive areas; Google on the advertisement and search engine markets is the immediate illustration. This type of dominant position and difficulty of emergence of competitors has to be kept in mind.

4.2 Competition at the content level

This section discusses general models of competition, but also (arbitrary) specific situations as illustrations.

4.2.1 General models

With full generality, the (abstract) models developed in previous chapters are also applicable to analyze competition at the application/content level. Basically, we can represent the competition by two-level games where:

1. At the largest time scale, providers compete on the quality and design of their content/application and their price (if any);
2. at the smallest time scale, users select their provider given the strategy profile of providers. Note that they can even choose several providers depending on the type of consumption, by splitting their usage between the providers. This is typically what happens for news web site readers, or often also for search engine users.

The games are here too solved by backward induction, the providers *anticipating* the reaction of users for a given profile of strategies at the largest time scale, making use of the subsequent users reaction.

Price was considered as the (or one of the) key decision variable for network providers in the previous chapter, since it is a parameter which has a major impact on users. This may still be the case for some specific service providers, such as cloud service providers, CDNs (in this case users are specific content providers themselves), e-commerce, content providers such as paid news sites, online video services, etc. Though there are situations where applications or contents are free, and even where free and paid services compete, hence other

parameters or variable decisions have to be emphasized. A few examples of such parameters:

- Content innovation and rejuvenation (or updating) for traditional web sites. This can be seen as a function of the level of financial investments made by the content provider towards activity on content. The utility (revenue) is then also a function of the investment level, as a function of the number of visits (through advertisement, but not necessarily only), itself dependent on the profile of investments for all providers, hence a game.
- Web site design and amount of incorporated advertisement. The more ads displayed, the larger the potential revenue, but displaying too many of them may deter users from consulting the content site. This has to be balanced, in relation with the choices made by competitors.

The models to analyze such a competition may thus again follow one of the three following approaches for customers' choices:

- Wardrop principle when customers are assumed non-atomic and choose the provider yielding the smallest *perceived cost*, or equivalently the largest utility. Here, cost can again be based on price, for paid services mentioned above, and on QoS (or congestion effects) such that more users implies a degraded service—the most relevant use case of this type of equilibrium notion. But other notions are usually involved, such that quality of content in addition to price, or instead of it in case of free services. On top of that game, a game is played by providers, making use of the subsequent Wardrop equilibrium. This type of model is of interest when congestion effects occur (a mass of users consuming the service creates a loss of utility for others), otherwise it ends up with only one provider with a positive demand.
- Discrete choice models such that the valuation of a customer u for a provider, say i , is

$$V_i = \sum_{j=1}^k v_{i,k} + \kappa_{u,i}$$

where $v_{i,k}$ represents the measured objective value of a characteristic/attribute k at Provider i , and $\kappa_{u,i}$ is an unobserved user-specific value that is treated on the global level as a random variable. In Subsection 3.1.2, a single attribute was considered, related to price, but more can be used, as predefined functions of a corresponding decision variable such as reputation, content investments, design of the application, advertisement level, etc. Customers will select the provider maximizing their utilities, hence the proportion of customers choosing i is $\mathbb{P}[V_i = \max_j V_j]$. Here again, providers may play on decision variables that impact the various $v_{i,k}$.

- Aggregated demand models, for instance if a decision vector $\mathbf{s} = (s_i)_{i \in \mathcal{N}}$ is given for a set \mathcal{N} of providers, with s_i the decision variable (or potentially vector) of Provider i , demand at i is given by

$$d_i(\mathbf{s}) = d_{i,0} - \alpha_i f(s_i) + \sum_{j \neq i} \beta_{ij} f(s_j)$$

where $d_{i,0}$, α_i and $(\beta_{ij})_{j \in \mathcal{N} \setminus \{i\}}$ are strictly positive parameters and f is a non-negative function. The decision variable can be the price as in Subsection 3.1.2 with f the identity function (typical of paid applications/content), or again content/QoS investments. Such models do not go down to the user level, and can then be seen as more arbitrary unless the demand function form is validated. Nevertheless they often lead to more tractable derivations.

The analysis can then be performed like in the previous chapter.

4.2.2 Illustrative model of competition between free CPs with advertisement

To give a mathematical model of a specific situation, we present here a simplified version of the one in [28]. The game is summarized in Table 4.1.

	CP i
Input: content quality	Q_i
Strategic choice: advertising level	a_i
Effect 1: perceived quality	$V_i = Q_i(1 - a_i)$
Effect 2: number of visits	$C_i = V_i / \sum_j V_j$ (competition)
Consequence: revenues $(C_i Q_i a_i)_i$	

Table 4.1: Game among CPs on advertising levels: a larger a_i impacts positively the revenue per visit, but negatively the number of visits because of competition (more ads means worse perceived quality, hence users may turn to competing CPs).

Consider a content, which could be a video sequence, a movie, or a TV show. This content is controlled (or offered) by a set \mathcal{I} of different *free* CPs, which can play with the amount of advertisement included in their webpage. CPs have different qualities *experienced* by users, depending on the number of clicks needed to reach the content, such as when ads are superimposed to the content and have to be clicked to be removed, or when there is a time before accessing the content if a video ad is displayed first. The level A_i of advertising at the CP i is thus a *nuisance* for users but an additional revenue through clicks for the CP. We assume that CP i earns A_i each time its content is accessed, and that the quality of experience (QoE) V_i that the user has with CP $i \in \mathcal{I}$ is of the form

$$V_i = Q_i - A_i, \quad (4.1)$$

where $Q_i > 0$ is the intrinsic quality of the content of CP i (i.e., the quality that the user would experience if there were no advertisement). To simplify notations, we will define and use $a_i := A_i/Q_i$ as the *relative* amount of advertisement introduced by CP i , such that $a_i \in [0, 1]$ because the amount of advertising will be reduced to ensure non-negative utility in (4.1). Note that we take the CP (or more precisely, publisher) point of view: the CP is not the one creating advertisements, it just displays them and gets paid per impression (thus, proportionally to the number of views), so we do not include advertisement creation costs in our model. We also assume, to simplify the model in [28], a stickiness model such that the proportion C_i of users choosing CP i is proportional to the valuation associated to that CP:

$$C_i = \frac{V_i}{\sum_{j \in \mathcal{I}} V_j} = \frac{Q_i(1 - a_i)}{\sum_{j \in \mathcal{I}} Q_j(1 - a_j)}. \quad (4.2)$$

The expected revenue per unit of time for a content provider $i \in \mathcal{I}$, through advertising, is

$$R_i = C_i A_i = C_i Q_i a_i, \quad (4.3)$$

where it is assumed linear in the amount of displayed advertisement. Each CP i 's revenue thus depends on its strategic choice a_i but also on the amount of advertising $(a_j)_{j \in \mathcal{I} \neq i}$ that the other CPs fix through the market share C_i .

For this game between CPs on the level of advertising, we established in [28] the existence of a (non-trivial) Nash equilibrium, using Brouwer's fixed-point result as the main tool.

Proposition 4. There exists a Nash equilibrium $a^{\text{NE}} \in (0, 1)^{|\mathcal{I}|}$, and any Nash equilibrium is such that

$$1/2 < 1 + \phi_i - \sqrt{\phi_i^2 + \phi_i} \leq a_i^{\text{NE}} < 1, \quad (4.4)$$

with $\phi_i = \sum_{j \in \mathcal{I} \setminus \{i\}} \frac{Q_j}{Q_i}$.

Let us briefly look at two particular cases, for which we can carry the analytical study further.

Symmetric case ($Q_i = Q \forall i \in \mathcal{I}$)

Looking more closely at the simpler symmetric situation, we get the expression

$$R_i = a_i \frac{1 - a_i}{\sum_{j \in \mathcal{I}} (1 - a_j)}. \quad (4.5)$$

The revenue optimization leads to

$$(1 - 2a_i) \sum_{j \in \mathcal{I}} (1 - a_j) + a_i(1 - a_i) = 0, \quad (4.6)$$

giving for any $i, k \in \mathcal{I}$

$$(1 - 2a_i) \sum_{j \in \mathcal{I}} (1 - a_j) + a_i(1 - a_i) = (1 - 2a_k) \sum_{j \in \mathcal{I}} (1 - a_j) + a_k(1 - a_k).$$

Stated otherwise, it gives

$$(a_k - a_i) \left(a_k + a_i - 1 + 2 \sum_{j \in \mathcal{I}} (1 - a_j) \right) = 0.$$

From our general result, $a_j > 1/2$ for all $j \in \mathcal{I}$ at a Nash equilibrium, so that the right part of the above expression is strictly positive, giving $a_i = a_k$. Nash equilibria are necessarily symmetric, of the form $a_i = a$ for all $i \in \mathcal{I}$. Plugging that condition into (4.6), we obtain a unique equilibrium, with

$$a_i^{\text{NE}} = \frac{n}{2n-1} \quad \forall i \in \mathcal{I}, \quad (4.7)$$

where n is the total number of CPs, i.e., $n := |\mathcal{I}|$, yielding the corresponding revenue

$$R_i = a \frac{1-a}{n(1-a)} = \frac{1}{2n-1}.$$

From this expression, we can remark that the more competition (that is, as n increases), the less advertisement at each CP at equilibrium, with an asymptotic value $1/2$. Moreover the sum of revenues $R = \sum_{i \in \mathcal{I}} R_i = \frac{n}{2n-1}$ is also decreasing up to an asymptotic value $1/2$.

Duopoly case

Consider now the case of an asymmetric duopoly. The best response functions then equal

$$\begin{aligned} a_1(a_2) &= 1 - \frac{Q_2}{Q_1}(1-a_2) \left[\sqrt{1 + \frac{Q_1}{Q_2} \frac{1}{1-a_2}} - 1 \right] \\ a_2(a_1) &= 1 - \frac{Q_1}{Q_2}(1-a_1) \left[\sqrt{1 + \frac{Q_2}{Q_1} \frac{1}{1-a_1}} - 1 \right]. \end{aligned} \quad (4.8)$$

We draw in Figure 4.1 an example of best-response functions. We can check that there are two Nash equilibria (the points where the curves intersect), first the unlikely dominated situation $a_1 = a_2 = 1$ that we disregard, but another point with $a_i > 1/2 \forall i$ (since in the symmetric case here, we get $a_i = 2/3$ in accordance with the previous results).

Figure 4.2 investigates the evolution of that Nash equilibrium point for a range of ratios Q_1/Q_2 between the qualities at CPs. It is interesting to note here that a content provider with higher intrinsic quality can increase its advertising load. For more numerical results on this type of model, the reader is advised to go to [28].

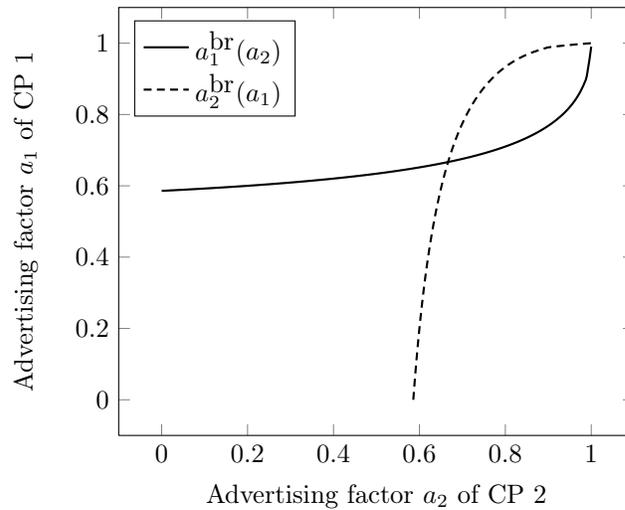


Figure 4.1: Best response functions in the duopoly case for the competition between free CPs.

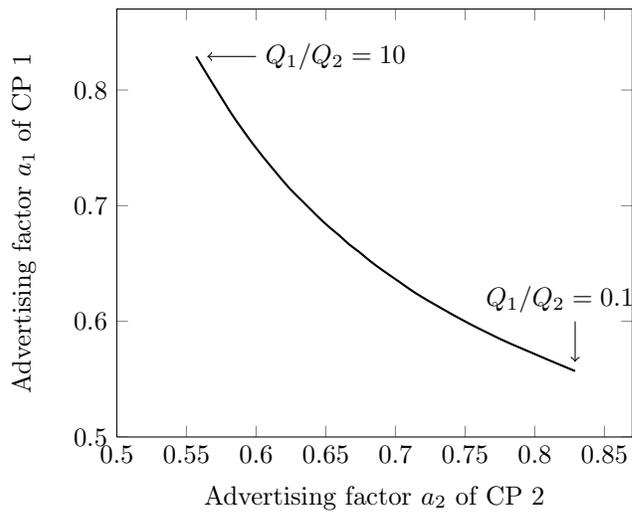


Figure 4.2: Nash equilibrium points in a duopoly for various ratios Q_1/Q_2 .

4.3 Economics of network security

Cybercriminality affects businesses (Cyber risk is considered the most critical risk by enterprises [118]) as well as private individuals. A recent illustration is the security breach that has been found in the PlayStation Network in April 2011, whose cost for Sony has been estimated in tens of millions of dollars. That breach exposed the personal information, and possibly the credit card data, of 77 million customer accounts.

Therefore, network service providers face new economic issues linked to the providing of communications that are protected against all potential types of attacks. That task is unfortunately impossible: with the appearance of new applications and services, the number of breaches to cover increases exponentially, so that many of them get discovered only when they are exploited by an attacker. The amounts of money that cybercriminality represent are enormous, be it in terms:

- of damage costs (imagine the financial loss of a 1-day service rupture for a company like Amazon),
- of investment costs (security represents about 10% of companies IT expenses),
- or even in terms of the underlying economy of attackers, i.e., attackers selling their “services”.

That last point is particularly striking: since it may become economically interesting for a company to harm a competitor’s IT system, some firms can be willing to pay to do so, which is now incredibly easy. Indeed, there now exists a (almost open) market for zombies renting, i.e., hackers who have managed to take control of a large number of machines rent them for a period of time, at a given price, so as to saturate the competitor’s servers, i.e., run a Distributed Denial of Service attack. In the same vein, but dealing more directly with people’s wealth, a highly competitive black market of stolen credit card numbers has risen in the last years [62, 128].

Those examples illustrate the fact that cybercrime is now highly organized and competitive. Building businesses such as credit card data selling or zombie renting takes some considerable effort, which has to be rationalized to maximize revenue. This is the reason why the interactions among all actors in the context of cybersecurity should be modeled and studied within the framework of game theory, that precisely considers the potential outcomes of situations where several self-interested agents are involved.

Let us come back to the setting described in Section 2.1.3 (with users choosing a security solution based on its price, performance, and likeliness of being targeted). We proved in [73] that for any price vector $\mathbf{p} = (p_1, \dots, p_{|\mathcal{I}|})$ set by the providers, a repartition $(V_i)_{i \in \mathcal{I}}$ of the total data value among the providers is the unique repartition minimizing a strictly convex function, which guarantees that a user equilibrium exists, is unique if all providers fix different prices, and is tractable (using classical convex optimization tools [13]).

Those results can then be used to study a higher-level game played by providers, consisting in fixing prices so that the resulting user equilibrium maximizes revenue. Again providers are leaders in a multi-level game, and are assumed to anticipate the user behavior. Remark that providers do not directly care about the *value* of the data they protect, but rather about their *market share*: denoting by n_i the proportion of users subscribing to Provider i , the revenue of that provider is indeed $p_i n_i$.

Actually, if a bounded-price alternative exists (e.g., a free security solution), then providers will not raise their prices to extremely high values since their revenue will decrease to 0 [73]. This is illustrated in Figure 4.3, where we consider two security providers competing on prices, but with a free alternative (denoted by Provider 0). Numerical results suggest it is quite likely that a price war situation arises, i.e., that successive price adaptations to the competitors' behavior lead to outcomes where providers make no revenue. However, when studying

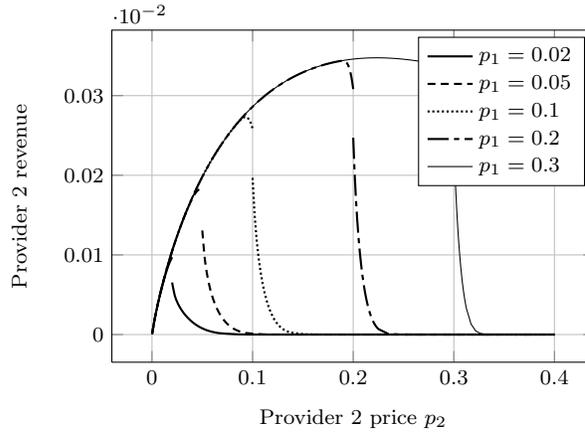


Figure 4.3: Revenue of provider 2 ($\pi_2 = 0.005$) when facing provider 1 ($\pi_1 = 0.01$) and free provider 0 ($\pi_0 = 0.05$), with $r_i(x) = 1 - e^{-x}$.

the interaction among security providers as a repeated game, i.e., a game played repeatedly over time, then equilibrium prices yielding positive revenues can be reached (due to the Folk Theorem).

4.4 Summary

In this chapter, we have summarized some of the results obtained for models where content (or service) providers are in competition. Given the numerous interactions among them, but also with users and possibly ISPs, those models can be very complex, hence the need for simplification to maintain tractability while capturing the main phenomena. The examples shown in this chapter illustrate the variety of strategic levers that actors can activate: not only prices

can be used, but also the amount of advertising or the level of performance (through investments) can be strategically determined. Another example, not presented here, is that of advertisers competing to attract users by bidding for keywords on search engines (e.g., via Google adwords). Advertisers' bidding strategies can then be quite elaborate, especially when advertisement budget constraints have to be taken into account [80, 81].

My publications related to this chapter are [9, 28, 35, 73, 80, 81].

Chapter 5

Net and search neutrality

This chapter focuses on debates that are still vivid, regarding the *neutrality* of intermediate actors between content providers and users. Neutrality is not easy to define, be it in technical or judiciary terms, but the underlying principle is quite intuitive: when providing users with access to content, intermediaries should not introduce any user-based or content-based bias.

The most famous specific cases are

- the *network neutrality* debate, where the behavior of ISPs (providing the “pipes” between content providers and users) is under scrutiny to prevent some contents from being (dis)avored over others;
- and the *search neutrality* debate, where search engines (providing links to content providers as a response to user requests) should present the most relevant results, independently of the interests they may have among the candidate results.

We present our contributions with regard to those two debates in the next two sections.

5.1 The network neutrality issue

5.1.1 Introduction and historical facts

Let us first briefly summarize the main facts and concerns regarding the debate.

- Especially due to vertical integration, some ISPs may favor their own content (or content paying for being favored) with respect to external one, by providing a better QoS for instance;
- With constantly decreasing transit prices (see Figure 5.1), ISPs are worried that revenues will become more hazardous and that investment in the network infrastructure will become more difficult. They are thus asking

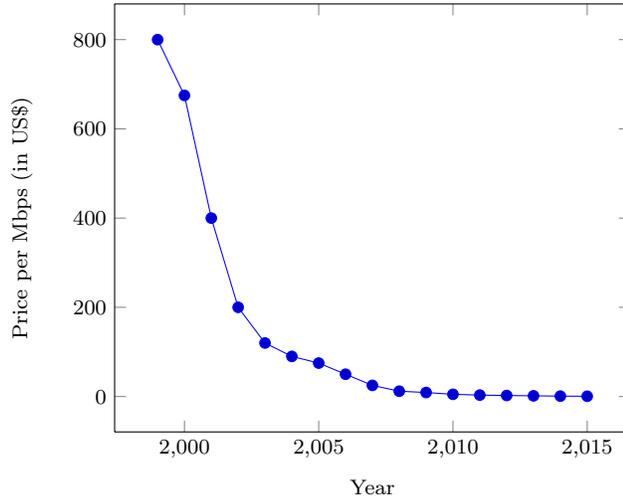


Figure 5.1: Internet Transit Prices per Mbps (the last values are *projected* prices). Data from drpeering.net.

for content providers to pay transit fees (i.e., to pay for traffic transferred by distant ISPs).

Those suggestions to discriminate traffic or to charge content providers seem in contradiction with the traditional vision of the Internet, providing a universal connectivity and serving all packets in the same way. This has raised the so-called *network neutrality debate*.

But discussing network neutrality first requires to define what a neutral (or a non-neutral) behavior is. Surprisingly, there is no well-defined notion of such neutrality. There is sometimes a distinction between *strong* and *weak* neutrality, where a strongly neutral network is such that it does not allow to manage packets differently in whatever way, while a weakly neutral network just prohibits user discrimination but allows discrimination between application types. In the former case, the network is understood to be as “dumb” as possible, just carrying packets, and the “intelligence”, if any, is rather placed at the source and destination. The idea in the latter case is that some applications may have more stringent QoS requirements (typically video needing limited delay and latencies with respect to email services), hence there is no harm but benefits for users to discriminate in case of congestion. A quite generally-used definition of neutrality was introduced by Professors Timothy Wu and Lawrence Lessig, saying that “Network neutrality is best defined as a network design principle. The idea is that a maximally useful public information network aspires to treat all content, sites, and platforms equally.” A kind of “official” definition has been summarized by the four following items and provided by the Federal Communications Commission (FCC) in the USA in 2005 [39]:

1. no content access can be denied to users;
2. users are free to use the applications of their choice;
3. they can also use any type of terminal, provided it does not harm the network;
4. they can competitively select the access and content providers of their choice.

The history and more details on the issues at hand, with the arguments developed by both sides, can be found in [83]. In our opinion, the debate is actually between two different worlds with two different goals:

- an idealistic (neutral or weakly neutral) network as imagined initially by scientists, with an organization in layers, a low cost, and for which end-to-end connectivity and universality are the key issues;
- a purely economic (non-neutral) view of the network, looking at an efficient economic management.

A strict comparison is thus difficult since it depends on the view of what the network should be. Network neutrality is thus rather a political question about whether we want a commercial or a non-commercial network, and the recent intervention of US President Obama indicated a clear stance toward the latter.

We can also remark here that whatever the goal, the need for regulation may differ depending on the considered country. Indeed, there are differences in the competition between ISPs which could mean a different output if a careful analysis is not performed: in the USA for instance, competition is not as fierce as in Europe for broadband Internet access. Moreover, the fact that the main content/service providers come from the USA may have an impact on the political decisions.

Regardless of the main stances toward the debate, economic modeling and analysis of the questions at hand should help decide what regulation (if any) to impose.

5.1.2 Modeling content and network providers interactions and analyzing neutrality issues

There exist many works trying to model network neutrality-related issues and perform an analysis thanks to game theory (see among others [5, 6, 18, 20, 40, 47, 67, 68, 92, 96] and the references therein). We illustrate in this section how modeling can be helpful to draw conclusions.

One content provider and one access provider

The model we present is based on the one in [29], with one CP and two ISPs, that we first simplify by considering one CP, whose parameters will be indexed by 1 and a single ISP, named *A*. The *flat rate* subscription fees prices charged

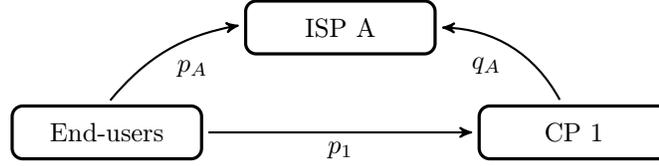


Figure 5.2: Charging interactions between stakeholders. Prices p_1 and p_A are positive flat rates, whereas q_A is a positive per-volume unit price.

to users to access the CP and ISP A are respectively denoted by p_1 and p_A . We are going to study and compare neutral and non-neutral outcomes, therefore we introduce a *side payment* $q_A > 0$ that the CP has to pay *per unit of volume* to ISP A . In the neutral case, q_A is just fixed to zero. The charges imposed by actors to other players are summarized in Figure 5.2.

We assume a continuum of end users, of mass one without loss of generality. We assume that users first choose whether to subscribe to the ISP (depending on price), and then in the affirmative case whether to subscribe to the CP. We separate this choice from the broadband access, because users also want to access the network for other reasons, such as email, web browsing, etc. We consider the discrete choice/stickiness model of Subsection 3.1.2 for the choice of subscribing or not to ISP A , with a “cost” p_0 of not subscribing to the Internet, leading to a mass (or proportion) of users subscribing to an Internet access (through A)

$$\sigma_A = \begin{cases} \frac{p_A^{-\beta}}{p_A^{-\beta} + p_0^{-\beta}} & \text{if } p_A > 0 \\ 1 & \text{if } p_A = 0. \end{cases} \quad (5.1)$$

Now, the willingness to pay a subscription to the CP is assumed to follow an exponential distribution with mean value $1/\alpha > 0$ over the population, independently of the ISP choice, leading to a proportion $e^{-\alpha p_1}$ of the ISP subscribers deciding to subscribe also to the CP, hence a total mass $\sigma_A e^{-\alpha p_1}$ subscribing to both. As previously with the Gumbel distribution for user preferences (Section 3.1.2), the choice of an exponential distribution (and hence an exponential demand) is mainly motivated by mathematical convenience. But when we study prices in a vicinity of a given level, what matters more than the actual distribution is the local *sensitivity* to price (or equivalently, the demand elasticity), which for our model can be tuned with the parameter α . We denote by D_0 the average volume the CP subscribers download from the CP, giving a data volume

$$D_A = D_0 \sigma_A e^{-\alpha p_1} \quad (5.2)$$

which will be needed to compute the *volume-based* transit costs for the CP to the ISPs.

The ISP revenue is then

$$U_A = p_A \sigma_A + q_A D_A$$

and the CP revenue equals

$$U_1 = (p_1 \sigma_A e^{-\alpha p_1} - q_A D_A) = (p_1/D_0 - q_A) D_A,$$

where we include the subscription gains and volume-based side payments.

The user welfare *associated with the existence of the CP* is then (looking at the "gain" $x - p_1$ for subscriber with willingness to pay x)

$$\begin{aligned} \text{UW}_{\text{CP}} &= \sigma_A \int_{p_1}^{\infty} \alpha e^{-\alpha x} (x - p_1) dx \\ &= \sigma_A \frac{e^{-\alpha p_1}}{\alpha} \\ &= \frac{D_A}{\alpha D_0}, \end{aligned}$$

User welfare can be decomposed into two components: the user welfare due to the existence of the CP (computed above), and the user welfare due to the presence of the ISPs. For that latter part, we take

$$\text{UW}_{\text{ISP}} = p_0 \left(\frac{p_0}{p_A} \right)^\beta.$$

(See [29] for a justification; remark that since user utility functions are not quasi-linear—i.e., expressed in a monetary-equivalent form—it is difficult to define UW_{ISP} , and other definition choices can be made.) The global user welfare generated by the system (ISPs and CP) is therefore

$$\text{UW} = \text{UW}_{\text{CP}} + \text{UW}_{\text{ISP}}. \quad (5.3)$$

Whatever the value of q_A assumed fixed first, CP 1 and ISP A choose respectively their price p_1 and p_A maximizing their revenue, but ISP A does it first at a larger time scale, anticipating the decision of CP 1. The ISP is therefore the leader in a Stackelberg game, hence an analysis of the interactions among providers using backward induction. The first-order condition then gives for CP 1 (with p_A and q_A fixed)

$$\frac{\partial U_1}{\partial p_1} = \frac{D_A}{D_0} - \alpha(p_1/D_0 - q_A) D_A = 0,$$

i.e.,

$$p_1 = 1/\alpha + q_A D_0.$$

We can remark here from the value of p_1 that inserting side payments induces a larger subscription fee for the CP, such that his revenue is $U_1 = D_A/(\alpha D_0)$ whatever the side payment, which is exactly the CP-related user welfare UW_{CP} . In other words, the interest of users connected to the Internet and that of the CP coincide.

To compute the optimal price p_A for ISP A , we also compute the derivative

$$\frac{\partial U_A}{\partial p_A} = \sigma_A + p_A \frac{\partial \sigma_A}{\partial p_A} + q_A D_0 e^{-\alpha(1/\alpha + q_A D_0)} \frac{\partial \sigma_A}{\partial p_A}.$$

In the neutral case ($q_A = 0$), p_A does not depend on p_1 , and we get $\frac{\partial U_A}{\partial p_A} = \sigma_A + p_A \frac{\partial \sigma_A}{\partial p_A} = p_A^{-\beta} (p_A^{-\beta} + (1 - \beta)p_0^{-\beta}) / (p_A^{-\beta} + p_0^{-\beta})$.

- If $\beta \leq 1$, this derivative is always positive, hence setting an infinite price is the “best solution”. Said differently, sensitivity to prices is not large enough to deter the ISP from increasing his price.
- If $\beta > 1$, the first-order condition gives

$$p_A = p_0(\beta - 1)^{-1/\beta}.$$

The non-neutral case is not tractable and we need to resort to a numerical evaluation. Figures 5.3 and 5.4 display the ISP revenue, CP revenue, user welfare and social welfare (sum of user welfare and provider revenues) in terms of q_A for the optimally chosen p_1 and p_A (this one being numerically determined), when $\alpha = 1$, $p_0 = 1$, $D_0 = 1$, and $\beta = 1.5$. To compare the outcome with the neutral case, we just need to compare with the point at the origin ($q_A = 0$). The

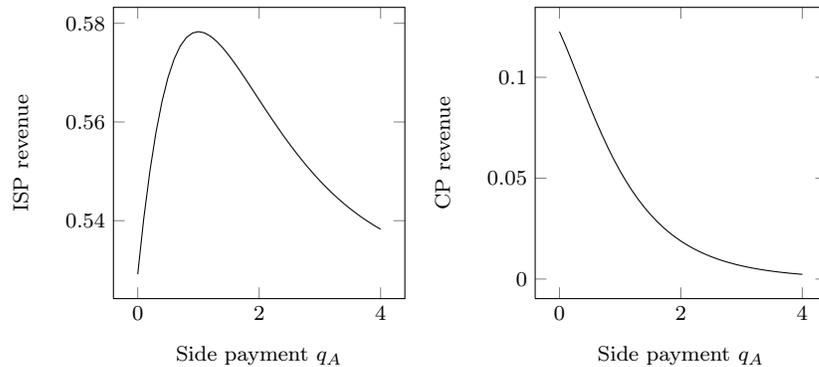


Figure 5.3: ISP and CP revenues, when the side payment q_A varies.

introduction of a side payment clearly increases the ISP revenue, user welfare and social welfare here, up to an optimal value above which demand decreases too significantly. Remark that the side payment optimizing the ISP revenue is slightly larger than the one optimizing user welfare. On the other hand, the impact of the side payment is negative on the CP revenue, explaining the strong reluctance of CPs to that “non-neutral” practice.

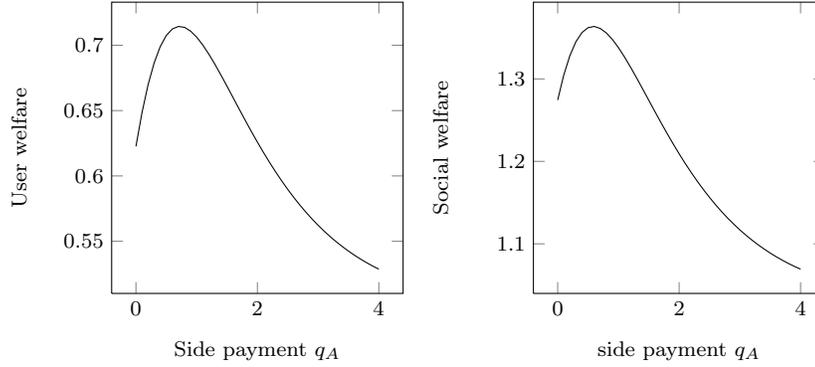


Figure 5.4: User and social welfare, when the side payment q_A varies.

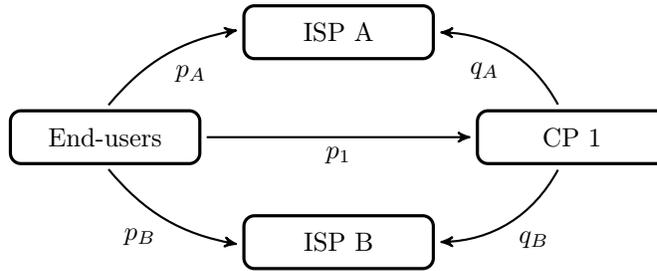


Figure 5.5: Charging interactions with two ISPs. Prices p_1 , p_A and p_B are positive flat rates, whereas q_A and q_B are positive per volume unit prices.

One content provider and two access providers

We now consider two ISPs in competition instead of just one, named A and B , all parameters for B being defined as for A above. Why considering competition between ISPs instead of between CPs? Actually ISPs complain that they endure competition (which is particularly true in Europe) at the network access level, while for most types of services there is often a dominant actor (Netflix, Google, etc.) and less competition, a reason why side payments are argued to become relevant. Because of that competition, ISPs say that they are forced to decrease their access prices and thus forced to get money from CPs. Our model can help to study the relevance of this argument. A new sketch of charges imposed among players for that setting is described in Figure 5.5. We assume again that users first choose their ISP, and then subscribe to the CP or not. We still focus here on a discrete choice/stickiness model for the ISP selection but notice that a model based on Wardrop principles has also been considered in [18], such that users simply select the cheapest ISP (leading to a Bertrand competition). With

the stickiness model, the proportion of users subscribing to ISP $i \in \{A, B\}$ is

$$\sigma_i = \begin{cases} \frac{p_i^{-\beta}}{p_A^{-\beta} + p_B^{-\beta} + p_0^{-\beta}} & \text{if } p_A > 0 \text{ and } p_B > 0 \\ 1 & \text{if } p_i = 0 \text{ and } p_j > 0 \\ 1/2 & \text{if } p_A = 0 \text{ and } p_B = 0 \\ 0 & \text{if } p_i > 0 \text{ and } p_j = 0. \end{cases} \quad (5.4)$$

Using exactly the same arguments as for a single ISP, user welfare associated with the existence of the CP is

$$UW_{CP} = \frac{D_A + D_B}{\alpha D_0}$$

and the providers revenues are for the ISPs ($i \in \{A, B\}$),

$$U_i = p_i \sigma_i + q_i D_i$$

and, for the CP,

$$\begin{aligned} U_1 &= (p_1 \sigma_A e^{-\alpha p_1} - q_A D_A) + (p_1 \sigma_B e^{-\alpha p_1} - q_B D_B) \\ &= (p_1/D_0 - q_A) D_A + (p_1/D_0 - q_B) D_B. \end{aligned}$$

The decisions on prices are still analyzed by backward induction, the decision at a given time scale being made anticipating the output at the later time scales (as before, ISPs play first—hence a game among them—and the CP adapts his price to the ISP prices).

We again first look at the smallest time scale (decision on p_1) for fixed other values. For convenience, we define $P_i := p_i^\beta$. A solution of the first-order condition gives (see [29] if details are required):

$$p_1^* = \begin{cases} \frac{P_A}{P_A + P_B} \left(D_0 q_B + \frac{1}{\alpha} \right) + \frac{P_B}{P_A + P_B} \left(D_0 q_A + \frac{1}{\alpha} \right) & \text{if } p_A > 0 \text{ or } p_B > 0, \\ D_0 \frac{q_A + q_B}{2} + \frac{1}{\alpha} & \text{if } p_A = 0 \text{ and } p_B = 0. \end{cases} \quad (5.5)$$

Here again, the CP's revenue when using this optimal price corresponds to the CP-related user welfare $\frac{D_A + D_B}{D_0 \alpha}$.

Knowing this reaction of the CP to ISPs' price, these ISPs play a game on their choice of p_A and p_B . In the neutral case (i.e., $q_A = q_B = 0$), it can be shown, using the following formulation for the revenue of ISP A when plugging the expression of the optimal price p_1 (a symmetric formulation being obtained for B), that

$$U_A = \begin{cases} \frac{P_0 P_B p_A}{P_0 P_A + P_0 P_B + P_A P_B} & \text{if } p_A > 0 \text{ and } p_B > 0 \\ 0 & \text{if } p_A = 0 \text{ or } p_B = 0, \end{cases}$$

so that $(p_A, p_B) = (0, 0)$ is a Nash equilibrium since no player can strictly increase his revenue by unilaterally changing his action. But such a player's strategy is strictly dominated by any other as soon as the adversary price is not zero. So it is not likely to be chosen by ISPs if another equilibrium exists. Actually, in this neutral case it can be shown (see [29] for a proof) that the (other) Nash equilibria can be described as presented in Table 5.1. In the case of

$\beta \leq 1$	$1 < \beta \leq 2$	$\beta > 2$
No equilibrium (prices tend to infinity)	Nash equilibrium $p_A = p_B = \left(\frac{2-\beta}{\beta-1}\right)^{1/\beta}$ $U_A = U_B := U^{\text{neutral}}$ $= (2-\beta)^{\frac{1}{\beta}} (\beta-1)^{1-\frac{1}{\beta}} \frac{p_0}{\beta}$	Nash equilibrium $p_A = p_B = 0$ (price war) $U_A = U_B = 0$

Table 5.1: Outcomes to expect from the ISP price competition game on p_A and p_B in the neutral case.

(positive) side payments, we are here too not able to get analytical results. But equilibria can be determined numerically and the resulting utilities compared with the neutral case. We present a part of the results in [29], still with $\alpha = 1$, $p_0 = 1$, $D_0 = 1$, and $\beta = 1.5$ (other values giving similar outcomes). Numerical computations show that the revenue of the CP and the user welfare he creates are always equal at equilibrium (which is easy to prove when $q_A = q_B$, but not in the general case). We display the revenues of providers in Figures 5.6 and 5.7.

Discontinuities can be observed, corresponding to situations when there is a price war: equilibrium subscription prices of both ISPs fall down to 0 for some side payments (this is for example the case when $q_A = q_B = 1$, but never when $q_B = 0$ or $q_B = 3$). We can briefly remark that the revenue of ISPs is not monotonic with the side payment, that the maximal revenue of an ISP, say A may be obtained for a null or positive side payment, and that the CP revenue has a tendency to decrease with side payments.

For this model with competing ISPs, we have not evoked yet the decision level corresponding to setting the side payment values. Side payments can be determined by the CP, or by ISPs, through a game. Conclusions from numerical investigations in [29] for these three cases are:

1. If side payments are decided by the CP: it is interesting to note that strictly positive side payments can be optimal for the CP, something counter-intuitive at first sight, especially since it was not the case when we had a single ISP. Actually, side payments exacerbate the competition between ISPs on access prices, reduced at equilibrium with respect to the neutral case; this is beneficial to end users, and finally to the CP who can reach more customers.
2. If side payments are decided by the ISPs, through a game: for the parameter values given above, $(0, 2.80)$ and the symmetric point $(2.80, 0)$ are

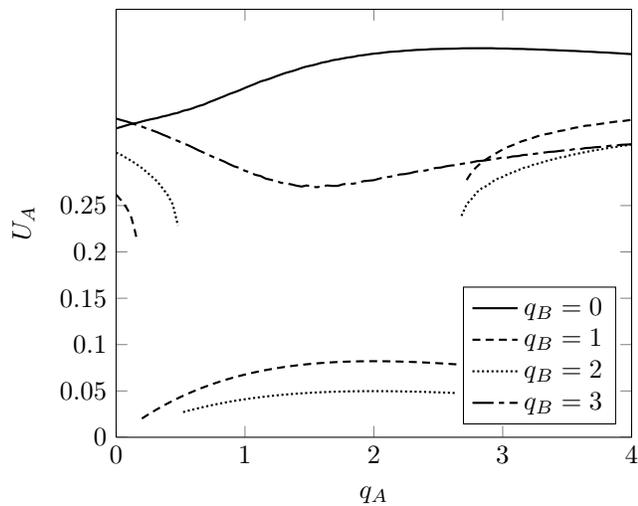


Figure 5.6: ISP A revenue at equilibrium as a function of the side payment q_A with $q_B \in \{0, 1, 2, 3\}$.

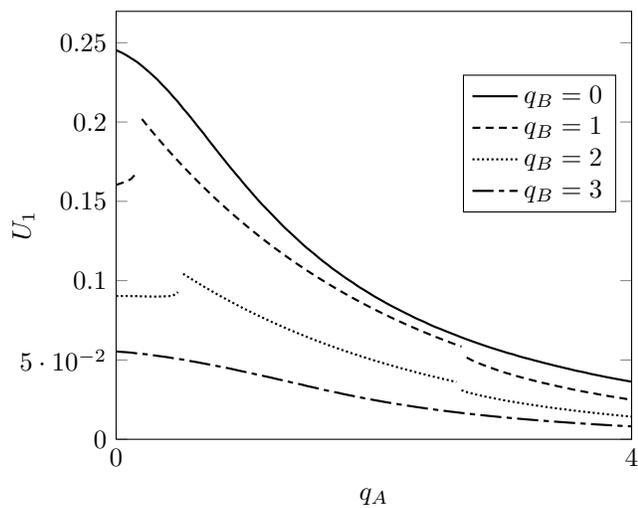


Figure 5.7: CP revenue at equilibrium as a function of the side payment q_A , with $q_B \in \{0, 1, 2, 3\}$.

Nash equilibria. Comparing with the neutral case, the ISPs total revenue increases by about 15% (hence in agreement with their call), while the CP revenue decreases by 75% of its value.

5.2 Search neutrality

5.2.1 The debate

Search engines play a pivotal role in the Internet economy, as the entry points to websites for most Internet users. As an illustration, in the US only about 20 billion requests from home and work desktops are treated by search engines each month¹. Search engines return a ranked list of links (the so-called organic results) to documents available on the World Wide Web given any keyword. The list is obtained from a link analysis algorithm which assigns a weight to documents [7, 8], the goal being to provide the most relevant results. In this section we will only focus on those organic results, ignoring the advertisement links aimed at yielding revenue to search engines.

But the ranking of organic links by search engines is now questioned by actors of the Internet and regulators, claiming that relevance is not its only factor, and that some revenue-making components are taken in consideration [34]. This question has become a vivid debate worldwide [34, 50]. The term *search neutrality* has been coined in 2009 by Adam Raff, co-founder of Foundem (a price-comparison company), after a vivid argument about Google penalizing his company in its ranking. The term is voluntarily inspired by *network neutrality*, because of the similarities in their stakes, namely the limitations on users' access to all relevant services on the Internet. Google actually acknowledged affecting a penalty to results such as Foundem's website, under the argument that it is a *vertical search engine*—i.e., a search engine focusing only on a part of the Internet, and that vertical search engines are perceived by users as spam. But Google also offers price-comparison and other specialized search services, and penalizing other companies in that same business can be seen as hindering competition; Google finally decided to whitelist Foundem manually, but kept its penalty policy towards other vertical search engines, leaving the debate open.

In addition (still focusing on Google because it represents more than 80% of the search market), Google offers many other services—e-mail, maps, calendar, video, shopping...—and would be naturally tempted to direct users towards them rather than towards their competitors. For example Google favors (or is accused to favor) YouTube content because money can be generated from those links². In this section, we therefore focus on the analysis of the reality of such biases, and discuss their impact on competition and on user welfare.

¹www.comscore.com

²See for instance <http://www.guardian.co.uk/technology/blog/2011/sep/21/eric-schmidt-google-senate-hearing> or, for measurements, <http://www.benedelman.org/searchbias/>

The main arguments in favor of search neutrality (i.e., imposing that ranking be based only on relevance) are in terms of user and social welfare. First, neutrality should benefit to users by providing them with the most relevant results, instead of driving them toward what yields the largest revenue to the search engine. For the same reason, neutrality benefits to the global economy by facilitating the access to the best-performing actors/services (and not those paying to be ranked well), so that new businesses can emerge more easily. As a side effect, it is also claimed that search neutrality elicits efforts from websites to improve their content quality, rather than to pay in order to be ranked better.

On the other hand, the opponents of search neutrality consider that users are interested in the differences among search engines, and select their preferred one, while such a differentiation would disappear if neutrality is enforced. Also, neutrality would prevent search engines from manipulating rankings, which they claim to do mainly to improve the results by avoiding spam: neutrality would then lead to worse results for users. Finally, imposing the transparency of the ranking algorithms (as often advocated by search neutrality proponents) raises issues in terms of intellectual property, and facilitates the job of spammers exploiting the working of those algorithms.

5.2.2 Do we need a regulatory intervention?

Investigations carried out in 2010 [38] and 2011 [126], both highlight some measurable biases in the results provided by search engines. The concern about non-neutral search engine is strongly shared by the European Union regulators, that are progressing toward an antitrust settlement deal with Google, after a complaint issued in 2010 by several specialized service companies. The European consumers' organization (BEUC) has acknowledged the risks of search bias, and suggested remedies. It has been agreed that Google search algorithm should follow some general principles to guarantee more fairness of its search results.

“Google must be even-handed. It must hold all services, including its own, to exactly the same standards, using exactly the same crawling, indexing, ranking, display, and penalty algorithms.”

BEUC, March 2013 [99]

In particular, measures consisting in labeling among organic results those pointing to own-content or sponsored content are not sufficient. This is in accordance with recent studies [49] showing that such labels have almost no effect on users' clicking behavior; the position of the link among the organic results appearing as the main factor.

5.2.3 Neutral *versus* non-neutral search engine: a model

Let us now investigate through a mathematical model (introduced in [64]) the incentives for a search engine to deviate from a ranking based only on relevance. Such a deviation will be called biased, or *non-neutral*.

To compare rankings based on *relevance* with other types of rankings, we first have to assume that an objective measure of relevance exists for each webpage potentially ranked by the search engine, and that the search engine is able to compute it. This assumption seems reasonable since search engines deploy complex algorithms to rank webpages, and should be able to fine-tune them for the best interests of their users (e.g., using the feedback on rankings that users make through their clicking decisions).

For each arriving request (i.e., a query sent to the SE by a user), different content providers (CPs) host pages that are relevant. Out of a universe of m_0 pages available online, we denote by $M \leq m_0$ the number of pages that match the arriving request. Each page $i = 1, \dots, M$ has a relevance value $R_i \in [0, 1]$, and an expected revenue per click $G_i \in [0, K]$ for the CP (here, K is a positive constant) of which the SE receives a fraction $\alpha_i \in [0, 1]$. Consequently, the SE's expected revenue per click from page i is $\alpha_i G_i$. The SE might sometimes also be the CP for a subset of the pages matching the request; in those cases $\alpha_i = 1$ because it receives all the revenue. Putting this all together, the instance of the ranking problem corresponding to a given request is encoded by a vector $Y = (M, R_1, G_1, \alpha_1, \dots, R_M, G_M, \alpha_M)$ that we assume to belong to a universe of admissible requests. After getting the request, the SE must select a permutation $\pi = (\pi(1), \dots, \pi(M))$ of the M pages and use it to display links to those pages in order. A stationary *ranking policy* μ is a function that assigns a permutation $\pi = \mu(Y)$ to each possible realization of Y . We shall only consider *deterministic* stationary policies, as opposed to *randomized* ones, which map each Y to a probability distribution over the set of permutations of M elements. The problem for the search engine will be to choose an optimal ranking policy, taking both into account the short-term effect (the immediate revenue) and the long-term effect (the number of visits, that depends on the user perceived relevance of results). That tradeoff is illustrated in Figure 5.8.

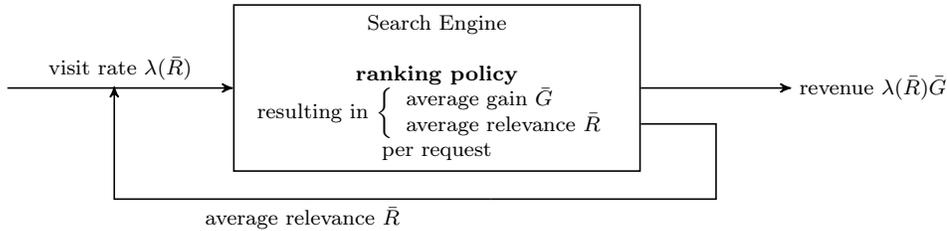


Figure 5.8: Search engine whose ranking policy produces and average relevance of results and an average gain. The number of visits (i.e., popularity of the engine) depends on the average relevance.

The *click-through-rate* (CTR) of a link that points to a page is defined as the probability that the user clicks on that link [48, Chapter 8]. This probability depends on the relevance of the content but also on the position number where

the link is displayed. We assume that the CTR of the link to page i placed at position $\pi(i)$ can be expressed as the (separable) product of a position effect and a relevance effect. That is, CTR is given by

$$\text{CTR}(i) = \theta_{\pi(i)} \psi(R_i),$$

where $1 \geq \theta_1 \geq \theta_2 \geq \dots \geq \theta_{m_0} > 0$ is a non-increasing sequence of fixed positive constants that describe the importance of each position in the ranking. The non-decreasing function $\psi : [0, 1] \rightarrow [0, 1]$ maps the relevance to the (position-independent) probability of the page. The assumption that the CTR is separable is pervasive in the e-Commerce literature [69, 121]. We will rely on it to derive simple optimality conditions. According to this assumption, to increase the CTR, we can either choose a more relevant page or we can choose a position closer to the top of the list.

Fixing a request Y and a permutation π , we now define the objective function we shall consider. The *local relevance* captures the attractiveness of the ordering from the consumer's perspective. It is computed by

$$r(\pi, Y) := \sum_{i=1}^M \text{CTR}(i) R_i = \sum_{i=1}^M \theta_{\pi(i)} \psi(R_i) R_i = \sum_{i=1}^M \theta_{\pi(i)} \tilde{R}_i, \quad (5.6)$$

where $\tilde{R}_i := \psi(R_i) R_i$. The expected total revenue arising from the request equals

$$g_0(\pi, Y) := \sum_{i=1}^M \text{CTR}(i) G_i = \sum_{i=1}^M \theta_{\pi(i)} \psi(R_i) G_i, \quad (5.7)$$

out of which the SE receives

$$g(\pi, Y) := \sum_{i=1}^M \text{CTR}(i) \alpha_i G_i = \sum_{i=1}^M \theta_{\pi(i)} \psi(R_i) \alpha_i G_i = \sum_{i=1}^M \theta_{\pi(i)} \tilde{G}_i, \quad (5.8)$$

where $\tilde{G}_i := \alpha_i \psi(R_i) G_i$.

To obtain an optimal ranking policy, we must consider that since customers are quality-sensitive, the choice of policy $\mu(\cdot)$ influences the future arrivals of customers. This has deep implications because a myopic policy for the SE (i.e., choosing $\mu(Y) \in \arg \max_{\pi} g(\pi, Y)$ for each Y) does not suffice to achieve optimality. To capture the dependence on future end-users that arrive to the SE, we consider the multivariate distribution of the input requests Y . Each request is then interpreted as a realization of Y according to that distribution.

We estimate the long-term value induced by a stationary ranking policy μ by taking expectations of the objectives presented earlier with respect to the distribution of input requests. Therefore, the expected relevance per request is

$$r := r(\mu) = \mathbb{E}[r(\mu(Y), Y)], \quad (5.9)$$

the expected total revenue per request is

$$g_0 := g_0(\mu) = \mathbb{E}[g_0(\mu(Y), Y)], \quad (5.10)$$

and the expected SE revenue per request is

$$g := g(\mu) = \mathbb{E}[g(\mu(Y), Y)]. \quad (5.11)$$

In the three previous definitions, the expectation is taken with respect to the random variable Y .

A non-myopic SE would be interested in the expected long-run revenue. This must depend on both the expected relevance per request r and on the expected SE revenue per request g . We capture the two dependencies through the general function

$$U_{\text{SE}} = \varphi(r, g), \quad (5.12)$$

where φ is an increasing function of r and g with bounded second derivatives over $[0, 1] \times [0, K]$. A natural particular case is, as suggested by the above discussion,

$$U_{\text{SE}} = \lambda(r)(\beta + g), \quad (5.13)$$

where $\lambda(r)$ is the expected number of searches per time unit when the expected relevance of the results is r . An *optimal policy* from the perspective of the SE is a stationary ranking policy μ that maximizes U_{SE} .

For that model, we characterized in [64] the optimal ranking policies:

Proposition 1. If the tuple (r, g) corresponds to an optimal policy, $\theta_k - \theta_{k+1} > 0$ for all k , then then this policy for almost all Y (with respect to the measure ν), μ sorts the pages by decreasing order of $\tilde{R}_i + \rho \tilde{G}_i$, with

$$\rho = \frac{\partial \phi(r, g) / \partial (g)}{\partial \phi(r, g) / \partial (r)}. \quad (5.14)$$

Note that then r and g depend on ρ , hence (5.14) is a fixed-point equation in ρ .

For the special case when $\phi(r, g) = \lambda(r)(\beta + g)$, we have sufficient conditions for such a policy to exist

Proposition 2. Assume that the distribution F has a density.

- If $x \mapsto \frac{\lambda(x)}{\lambda'(x)}$ is upper-bounded for all $x \in [0, r_0]$ then the fixed-point equation (5.14) has at least one solution.
- If $x \mapsto \frac{\lambda(x)}{\lambda'(x)}$ is non-decreasing then the fixed-point equation (5.14) has at most one solution.

In [64], we show how to approximate this fixed-point using simulation tools.

5.2.4 Comparison of the Neutral and Non-Neutral Ranking Policies

Let us now see on numerical examples how the theory developed earlier can be used to study the impact of different ranking policies on various performance indicators such as consumer welfare (captured by expected relevance), SE and CP revenue. In particular, we compare neutral ranking policies, where $\rho = 0$, with non-neutral ones, where the SE chooses the optimal ρ^* .

Table 5.2: CTR values used in the simulations of Section 5.2.4

θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9	θ_{10}
0.364	0.125	0.095	0.079	0.061	0.041	0.038	0.035	0.03	0.022

A Vertically Integrated SE with a CP

Example 1. We first focus on a specific type of request which can be served by either third-party CPs or by the SE itself. This is typical for many search categories where the SE also provides content (e.g., video, weather, finance, news, maps, flight information, and so on). In this case, a limited number of CPs compete with the SE, and the parameters r , g , and $\lambda(r)$ for the instance correspond to just this type of request. Let us assume that always ten pages match a request ($M = 10$). Nine of those pages are served by third-party CPs but one of them is served by the SE directly. Perhaps renumbering CPs, we have that $\alpha_1 = 1$, and $\alpha_2 = \dots = \alpha_{10} = 0$. In addition to the revenue coming from Page 1, the SE also receives an expected revenue of $\beta = 1$ per request from sponsored links. For $i = 1, \dots, 10$, R_i and G_i are all independent random variables uniformly distributed over $[0, 1]$, and $\text{CTR}(i) = \theta_i$ as specified in Table 5.2. Those numbers were taken from the first table in [36], which contains the observed relative numbers of clicks according to the position: the actual CTRs should therefore be proportional to those numbers, and the value of the multiplicative constant has no impact on our derivations (hence we take it equal to 1). Finally, we set $\lambda(r) = r$, and ψ to be the unit function.

The M pages are ranked by the SE by decreasing value of $\tilde{R}_i + \rho \tilde{G}_i$, for the correct constant $\rho \geq 0$. Note that for $i > 1$, $\tilde{G}_i = 0$ because $\alpha_i = 0$. To illustrate the dependence on ρ , Figure 5.9 shows the SE revenue $U_{\text{SE}}(\rho)$, as ρ varies, as well as the relevance $r(\rho)$, the revenue and the visit rate for CP 1 and for third-party CPs. All revenues are expressed as values per time unit. As discussed earlier, the more ρ increases, the more the SE favors CP 1, decreasing the overall relevance and increasing the visit rate to CP 1. The trade-off between short-term revenue and number of visits tells the SE to choose $\rho^* \approx 0.55$. Note that the bias affects only CP 1 and that the relative positions of all other CPs remain the same as in the neutral ranking. Consequently, the relevance $r(\rho)$ is only marginally affected by ρ in this case. If R_1 was stochastically much smaller than the other R_i 's (e.g., uniform over $[0, \epsilon]$ for a small ϵ), then the impact of ρ would be larger. When $\rho \rightarrow \infty$, CP 1 is always ranked first, so the relevance $r(\rho)$ becomes

$$r(\infty) = \left(\frac{\theta_1}{2} + \sum_{i=1}^9 \theta_{i+1} \mathbb{E}[U_{(10-i)}] \right) = \frac{\theta_1}{2} + \sum_{i=1}^9 \theta_{i+1} \frac{(10-i)}{10} \approx 0.517,$$

where $U_{(1)}, \dots, U_{(9)}$ are independent random variables uniformly distributed over $[0, 1]$ sorted by increasing order (the order statistics), and the CP 1 visit rate is $\theta_1 r(\infty) \approx 0.188$.

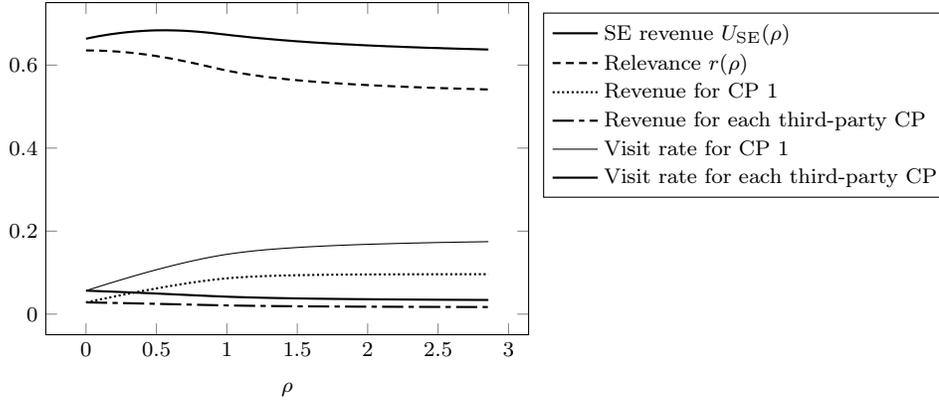


Figure 5.9: Performance measures as a function of the value ρ used in the ranking (simulation results)

To assess the sensitivity of the SE strategy to advertising, we now examine how results change for different values of β , i.e., depending on the level of advertisement revenues. This shows the tradeoff that the SE faces for different types of requests. For search keywords related to, e.g., airline tickets, hotel reservations, or retailer products, the SE may expect to make more profit by showing its own content among organic links than through sponsored search because requests of this kind may produce conversions, whereas for keywords that are appealing in the sponsored search market the SE may try to make the search as relevant as possible to boost that revenue stream. Figure 5.10(a) plots ρ^* as β varies while Figure 5.10(b) plots the ensuing revenue for CP 1 and for each third-party CP. The curves shown in the figures were estimated by simulation, using the iterative fixed-point method for ρ^* , with a fixed sample size of $n = 10^7$ at each step. When β grows, ρ^* tends to zero, because the revenue from sponsored links dominates, making it rewarding for the SE to improve quality to attract more users. In conclusion, the impact of non-neutrality is small because biasing the ranking only attracts limited additional revenue. Instead, when β is small, sponsored links do not pay off and it becomes worthwhile for the SE to sacrifice relevance to some extent to boost revenue from gains of CP 1. In the extreme case when $\beta = 0$, we have $\rho^* = \infty$, so CP 1 is always placed at the top regardless and the other CPs are sorted by decreasing order of relevance. This gives an average revenue of 0.09619 for CP 1 and 0.01695 for any other CP (even though all CPs have the same relevance and gain distributions). Although not shown in the figure, we remark that U_{SE} tends to grow linearly with β , which means that the increasing revenues of sponsored search dominate the possible revenue coming from CP 1. To illustrate the impact of non-neutrality, Table 5.3 reports the variations of the most relevant performance metrics when $\rho = \rho^*$ is

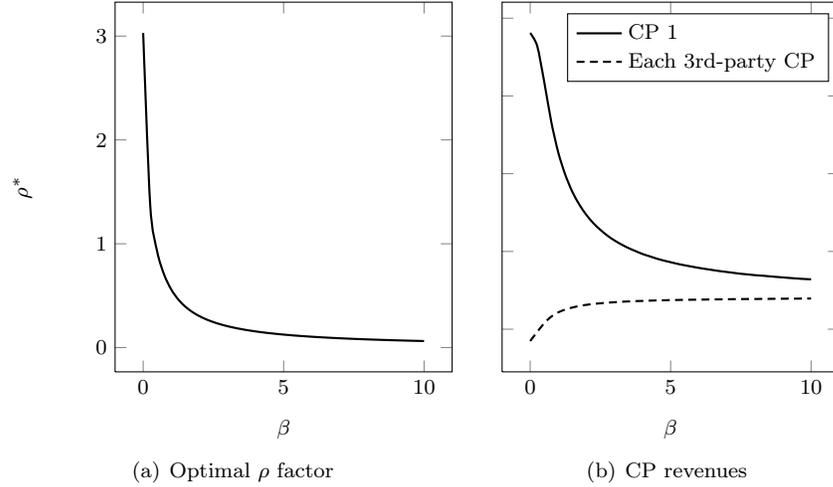


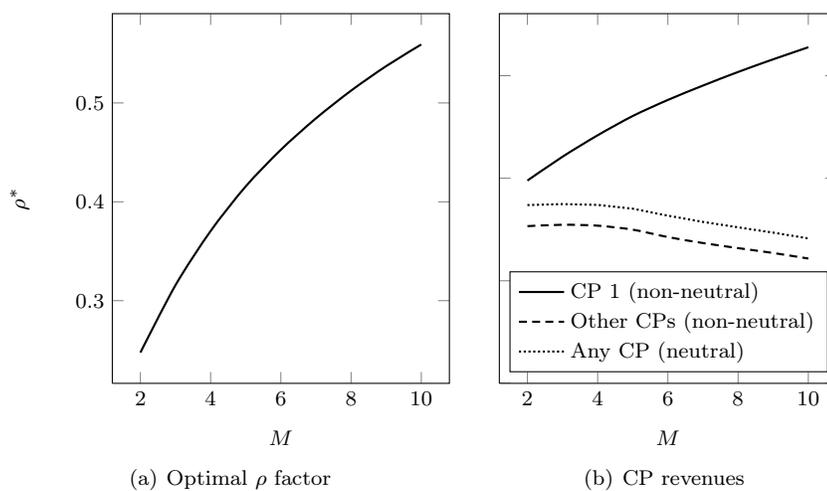
Figure 5.10: Optimal ρ factor to use for the ranking, and corresponding CP revenues per time unit as a function of the average advertisement revenue per visit β (simulation results)

used instead of $\rho = 0$ (neutral ranking), for different values of β . The table illustrates that while the impact on the perceived quality (relevance) remains small (around 10%), the impact on the visibility and the revenues of the SE-owned CP is substantial: by being non-neutral, the SE can multiply the revenues of its CP by a factor 2.8 and its visit rate by more than a factor of 3. On the other hand, the other CPs see their revenues and visit rates reduced by 14% to 32%, a significant loss that is likely to affect their possibilities of being profitable in the long term.

Finally, we explore the sensitivity of outcomes to the number of available results. Figure 5.11(a) plots ρ^* as a function of M while Figure 5.11(b) plots revenues as a function of the number of matching pages M . We include curves for both the neutral ($\rho = 0$) and non-neutral ($\rho = \rho^*$) regimes to compare both situations. As before, we estimate these values using the fixed-point algorithm with $n = 10^7$ at each step. As M increases, ρ^* increases too: The SE can give more weight to CP 1 and increase its revenue while making less damage to the relevance, because placing CP 1 higher has less impact on the overall relevance when M is larger. As a result, the revenue of CP 1 when $\rho = \rho^*$ increases with M , and so does the advantage of CP 1 over the other CPs. The loss of revenue of the other CPs seems close to constant as a function of M .

Table 5.3: Impacts of a non-neutral ranking for the scenario of Section 5.2.4

	Relevance	CP 1 revenue	other CP revenue	CP 1 visit rate	other CP visit rate
Neutral, $\rho = 0$ (reference case optimal for $\beta = \infty$)	0.635	0.028	0.0283	0.057	0.057
Non-neutral, $\rho =$ 0.559 (optimal for $\beta = 1$)	0.618 (-3%)	0.066 (+136%)	0.0243 (-14%)	0.112 (+96%)	0.049 (-14%)
Non-neutral, $\rho =$ 0.924 (optimal for $\beta = .5$)	0.592 (-7%)	0.084 (+200%)	0.0215 (-24%)	0.140 (+146%)	0.043 (-25%)
Non-neutral, $\rho =$ 1.374 (optimal for $\beta =$.25)	0.568 (-11%)	0.093 (+232%)	0.0193 (-32%)	0.158 (+177%)	0.039 (-32%)

Figure 5.11: Optimal ρ factor to use for the ranking, and corresponding CP revenues per time unit as a function of M (simulation results)

5.2.5 Vertical Integration and Investment

Example 2. Continuing with the example of vertical integration, we now assume that one of the nine third-party CPs, say CP 2, invests in quality and manages to improve the relevance distribution. More specifically, we assume that when it invests $z > 0$, the relevance of CP becomes uniformly distributed over $[0, 1+20z]$ (instead of over $[0, 1]$). The other parameters and distributions, including the distribution of its gain G_2 , are unchanged. Figures 5.12(a) and 5.12(b) show simulation results when the SE ranks CPs according to $\bar{R}_i + \rho \tilde{G}_i$, for varying values of ρ , and when $z = 2$. For a neutral ranking ($\rho = 0$), CP 2 logically makes more revenue than the other CPs, since it regularly gets higher ranking. However, when ρ increases and exceeds about 0.8, CP 1 becomes the one with highest revenue, despite its (stochastically) lower relevance.

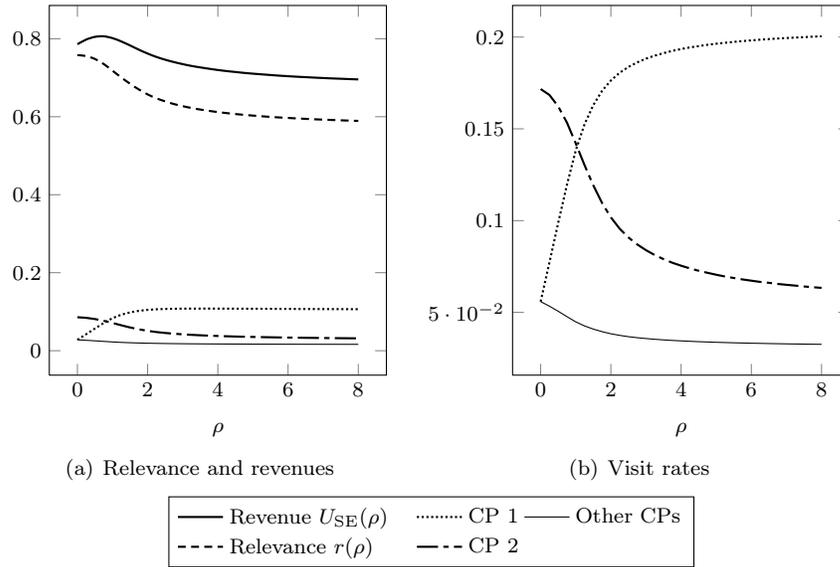


Figure 5.12: Average relevance, revenues and number of visits per time unit for the case of vertical integration with investment

We now take the perspective of CP 2, and compute its optimal decision. CP 2 invests z in quality to modify its relevance distribution to $[0, 1 + 20z]$, anticipating that the SE is going to rank requests according to ρ^* . (We assume that the SE can learn the distribution of relevance of all CPs quickly.) Therefore, CP 2's profit equals the revenue from the search market minus z . To find the optimal value of z we simulated the outcomes for $z \in [0, 0.45]$. Figures 5.13(a) and 5.13(b) plot the resulting curves. In both figures, we see that differences between neutral and non-neutral revenues are small, except for CP 1. This is particularly true for CP 2. This means that, at least in this case, non-neutrality

does not deter innovation. Actually, the optimal investment level under both regimes coincide and is equal to $z^* = 0.025$. Optimal profits, though, vary. They are 0.037 for the neutral case and 0.0296 for the non-neutral one; see Figure 5.14 where we show CP 2 profits as a function of the investment z .

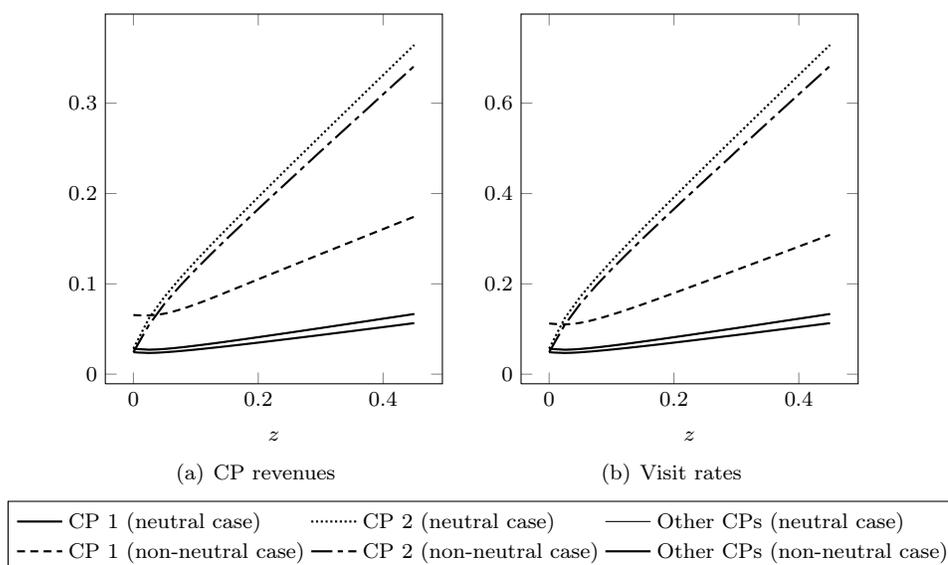


Figure 5.13: Revenues and visit rates to various CPs as a function of CP 2 investment

The results for this model might prove useful to platform owners (search engines, classified ads websites, online retailers) to navigate the tradeoff between short-term and long-term effects when defining their ranking strategies. They can also be of interest to regulators, seeking to understand the behavior of revenue-oriented platforms and to anticipate the impact of regulatory interventions, which is of particular importance with regard to the current *search neutrality* debate.

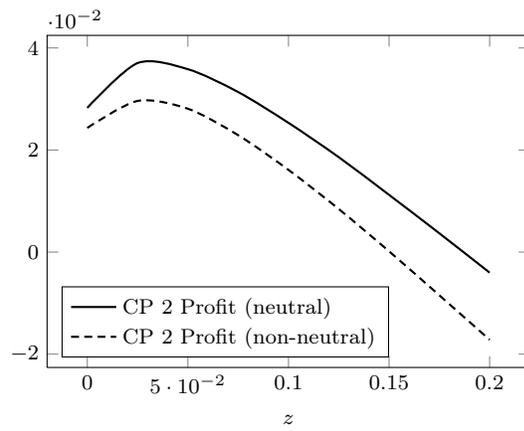


Figure 5.14: Profit per time unit as a function of CP 2 investment

Chapter 6

Conclusions and perspectives

6.1 A boundless research area

The Internet is an extremely complex ecosystem with enormous economic stakes, involving a lot of actors with non-aligned objectives. The balance among those actors, and in particular the consequences of their decisions on users and on newcomer companies are very carefully considered by regulators, in order to keep the Internet an innovation-fostering system, providing universal access to users with a sufficient quality.

The current debate on Net Neutrality epitomizes that attention from regulatory bodies (and the highest political levels) on the management of the Internet. The power equilibria among actors seem very sensitive to the (regulatory) context, and are also very likely to be dramatically affected by some recent technical changes in networks, when new functionalities enable new actors to enter the market and/or endanger the survival of existing actors.

In that context, economic studies can be of great help to understand and possibly anticipate the impact of new rules or new technologies on the ecosystem. Given the current speed of appearance of new services and technologies, there is definitely room (and need!) for research on the economics of telecommunications. I hope that the few contributions presented in this document illustrate this need, and the usefulness of such analyses.

From a scientific point of view, the telecommunication ecosystem is so complex and rich that potentially all kinds of interactions can be found, and with the flexibility offered by the transition to “virtualized systems”, new types of relationships among actors can even be imagined. Hence this field is an unbounded playground for scientific studies, to build and analyze the most complex game-theoretic models, design auction or regulation mechanisms (like the so-called *incentive auction* for spectrum in the US, whose rules are being defined by some of the most prominent economists in the world), or try to anticipate the user behavior when faced with new choices.

There are many new topics worth developing, I am presenting here only two,

but it is very likely that the coming months will witness more novelties in the telecommunication world deserving careful economic studies.

6.2 Some perspectives

6.2.1 The role and impact of Content Delivery Networks

The term *Content Delivery Network* (CDN) refers to both an *infrastructure* designed to deliver content at large scale over an underlying network, and the *economic actor* providing that service. We focus on the economic actor.

CDN have a huge economic weight (the annual revenues of Akamai, the CDN leading company, are over two billion dollars), and a growing impact on the Internet ecosystem: *i*) CDN activities affect the traffic exchanged between network providers, and consequently their economic relationships [61, 113]; *ii*) on many aspects (per-volume charging, connectivity service) CDN actors compete with transit providers, which explains why some major transit network operators such as Level 3 have shifted a fraction of their activities to CDN; and *iii*) other actors in the value chain of content delivery have started developing a CDN activity, including Internet Service Providers (ISPs), content providers, and equipment vendors [17, 116]. This fast-moving and business-driven environment exacerbates the concerns among user and regulation communities regarding service quality and economic fairness, epitomized by the net neutrality debate [33, 66, 74, 127].

The scientific literature provides models and analyses of the interactions between content providers and ISPs in order to address network neutrality, and sometimes to propose regulation remedies [5, 29, 30, 40, 96], but the role of CDNs is barely mentioned. To the best of our knowledge, the only official report mentioning CDNs is from the Norwegian regulator [117], where it is stated that “*the ordinary use of CDN servers is not a breach of net neutrality*”. More generally, the performance analysis community has barely considered the economics of CDN actors so far.

Currently, a few dominant players concentrate more than half of the Internet traffic on the behalf of millions of service providers. We would like to study the importance of CDNs in the ecosystem, with a particular focus on the net neutrality debate.

Our goal is to explore CDN management policies in economic terms, in particular investigating the consequences on the net neutrality debate. We therefore need to model the interplay between the aforementioned multiple parameters, determine general policies for actors, analyze today’s CDN policies and foresee possible evolutions. The finding would then help interpret the activities of CDN vis-à-vis the net neutrality debate and suggest possible regulation policies to mitigate disorders (if any) that CDNs introduce in the global management of the Internet.

We have started some preliminary work on those topics, by studying the behavior of revenue-driven CDN operators, for which the strategic decisions

are:

- Whose content to keep in the caches? Depending on the popularity of content, the transit costs in the network, the prices that content providers pays for the CDN service, and the cache capacity limits, the CDN has to decide if at all to keep some piece of content or not in its caches.
- Where to cache the content? CDNs own a whole network of servers, located at different distances from end-users. With tree-like topologies in the last miles, the trade-off is between caching content close to users (for better quality) and having it accessible to the largest number.
- How much to invest in cache capacities?

We are still investigating those questions in more and more complex (and realistic) scenarios, and hope to obtain results yielding insights regarding how to regulate (if any) the behavior of CDNs.

6.2.2 Software-Defined Networking: principles, opportunities and threats for the Internet ecosystem

Software-Defined Networking (SDN) solutions consist in separating the network control plan (e.g., making the routing decisions for some specific flows in the network) from the data plan (the actual material performing the data transfers). The behavior of the hardware components can then be modified and orchestrated in a simplified way, through standardized protocols such as OpenFlow. This allows a great flexibility: one can “build” a new network (based on existing hardware and infrastructure) or upgrade it very rapidly (in the order of minutes or seconds), compared to the extremely long delays to build new infrastructures or topologies (weeks when only commercial agreements need to be settled, or months when new physical communication lines have to be built).

From an economic point of view, the link between the network capacities offered and the physical resources used becomes less clear (the physical framework becoming transparent to the entity asking for some network service). As a result, many pricing schemes based on resource usage (as the ones surveyed in [31, 83, 123]) are harder to justify and should be redefined.

Moreover, the flexibility that SDN allows is very likely to strongly impact the actors involved, for example by rendering competition much fiercer. The resulting equilibria of forces may be totally different from what we currently know: for example we intend to investigate whether the current trend –where the market power of service/content providers increases at the expense of infrastructure providers– is likely to accelerate, or whether new types of actors (e.g., the SDN service providers) can take a preponderant place in the market. The impacts on the net neutrality debate are also worth considering.

Chapter 7

Bibliography

- [1] H.Z. Aashtiani and T.L. Magnanti. Equilibria on a congested transportation network. *SIAM Journal of Algebraic and Discrete Methods*, 2:213–226, 1981.
- [2] D. Acemoglu, K. Bimpikis, and A. Ozdaglar. Price and capacity competition. *Games and Economic Behavior*, 66:1–26, 2009.
- [3] T. Alpcan and T. Başar. *Network Security: A Decision and Game Theoretic Approach*. Cambridge University Press, 2010.
- [4] E. Altman, V. Kamble, and H. Kameda. A Braess type paradox in power control over interference channels. In *Proceedings of Physicomnet, WiOpt*, Berlin, Germany, April 2008.
- [5] E. Altman, A. Legout, and Y. Xu. Network non-neutrality debate: An economic analysis. In *Proc. of Networking*, pages 68–81, Barcelona, Spain, May 2011. LNCS 6641, Springer Verlag.
- [6] A. Arlandis and E. Baranes. Interactions between network operators, content producers and internet intermediaries: Empirical implications of network neutrality. *Intereconomics*, 46(2), 2011.
- [7] D. Austin. How google finds your needle in the web’s haystack. *American Mathematical Society Feature Column*, 10(12), 2006. <http://www.ams.org/samplings/feature-column/fcarc-pagerank>.
- [8] K. Avrachenkov and N. Litvak. Decomposition of the Google PageRank and Optimal Linking Strategy. Rapport de recherche RR-5101, INRIA, 2004.
- [9] D. Barth, J. Cohen, P. Maillé, and H. Pouyllau. Competition among online-gaming service providers. In *Proc. of International Conference on NETWORK Games, CONTROL and OPTimization (NetGCoop)*, Paris, France, 2011.

- [10] M. Beckmann, C. B. McGuire, and C. B. Winsten. *Studies in the economics of transportation*. Yale University Press, New Haven, Connecticut, 1956.
- [11] M. Ben-Akiva and S. Lerman. *Discrete choice analysis*. MIT Press, Cambridge, MA, 1985.
- [12] P. Bergendorff, D. W. Hearn, and M. V. Ramana. Congestion toll pricing of traffic networks. In P. M. Pardalos, D. W. Hearn, and W. W. Hager, editors, *Network Optimization*, volume 450 of *Lecture Notes in Economics and Mathematical Systems*, pages 51–71. Springer, Berlin, 1997.
- [13] D. P. Bertsekas. *Convex Optimization Theory*. Athena Scientific, 2009.
- [14] K. Bhawalkar, M. Gairing, and T. Roughgarden. Weighted congestion games: Price of anarchy, universal worst-case examples, and tightness. In *Proc. of 18th European Symposium on Algorithms (ESA)*, Liverpool, UK, 2010.
- [15] S. Bistarelli, M. Dall’Aglia, and P. Peretti. Strategic games on defense trees. In *Proceedings of 4th Intl Workshop on Formal Aspects in Security and Trust (FAST’06)*, LNCS 4691, pages 1–15, Hamilton, Ontario, Canada, Aug 2006.
- [16] N. Bohacek, J. P. Hespanha, J. Lee, C. Lim, and K. Obraczka. Game theoretic stochastic routing for fault tolerance and security in computer networks. *IEEE Transactions on Parallel and Distributed Systems*, 18(9):1227–1240, Sept 2007.
- [17] V. Bonneau. Evolution of the CDN market. In *CDN World Summit*, 2010.
- [18] F. BouSSION, P. Maillé, and B. Tuffin. Net neutrality debate: Impact of competition among ISPs. In *Proceedings of the Fourth International Conference on COMMunication Systems and NETWORKS (COMSNETS)*, Bangalore, India, 2012.
- [19] S. Buehler, R. Dewenter, and J. Haucap. Mobile number portability in Europe. *Telecommunications Policy*, 30(7):385–399, August 2006.
- [20] S. Caron, G. Kesidis, and E. Altman. Application neutrality and a paradox of side payments. In *Proceedings of the third International Workshop on Re-Architecting the Internet (ReArch 2010)*, Philadelphia, USA, Nov 2010.
- [21] C. K. Chau and K. M. Sim. The price of anarchy for non-atomic congestion games with symmetric cost maps and elastic demands. *Operations Research Letters*, 31:327–334, 2003.
- [22] R. Cole, Y. Dodis, and T. Roughgarden. Pricing network edges for heterogeneous selfish users. 2003.

- [23] R. Cole, Y. Dodis, and T. Roughgarden. How much can taxes help selfish routing? *Journal of Computer and System Sciences*, 72(3):444–467, May 2006.
- [24] R. Cominetti, J. R. Correa, and N. E. Stier-Moses. The impact of oligopolistic competition in networks. *Operations Research*, 57(6):1421–1437, 2009.
- [25] J. R. Correa, A. S. Schulz, and N. Stier-Moses. A geometric approach to the price of anarchy in nonatomic congestion games. *Games and Economic Behavior*, 64(2):457–469, Nov 2008.
- [26] J. R. Correa, A. S. Schulz, and N. E. Stier-Moses. Selfish routing in capacitated networks. *Mathematics of Operations Research*, 29(4):961–976, 2004.
- [27] J. R. Correa, A. S. Schulz, and N. E. Stier-Moses. Fast, fair, and efficient flows in networks. *Operations Research*, 55(2):215–225, 2007.
- [28] P. Coucheney, G. D’Acquisto, P. Maillé, M. Naldi, and B. Tuffin. Influence of search neutrality on the economics of advertisement-financed content. *ACM Transactions on Internet Technology*, 14(2-3), 2014.
- [29] P. Coucheney, P. Maillé, and B. Tuffin. Impact of competition between ISPs on the net neutrality debate. *IEEE Transactions on Network and Service Management*, 10(4):425–433, 2013.
- [30] P. Coucheney, P. Maillé, and B. Tuffin. Network neutrality debate and ISP inter-relations: Traffic exchange, revenue sharing, and disconnection threat. *Netnomics*, 2015. in Press.
- [31] C. Courcoubetis and R. Weber. *Pricing Communication Networks: Economics, Technology and Modelling*. Wiley & Sons, Inc., 2003.
- [32] M. Cremonini and D. Nizovtsev. Understanding and influencing attackers’ decisions: Implications for security investment strategies. In *Proc. of 5th Workshop on the Economics of Information Security (WEIS)*, Cambridge, UK, 2006.
- [33] P. Crocioni. Net neutrality in Europe: Desperately seeking a market failure. *Telecommunications Policy*, 35(1):1 – 11, 2011.
- [34] J. Crowcroft. Net neutrality: The technical side of the debate: A white paper. *ACM SIGCOMM Computer Communication Review*, 7(1), January 2007.
- [35] G. D’Acquisto, P. Maillé, M. Naldi, and B. Tuffin. Impact on retail prices of non-neutral wholesale prices for content providers. In *Proc. of 8th Euro-NF Conference on Next Generation Internet (NGI)*, Karlskrona, Sweden, 2012.

- [36] R. DeJarnette. Click-through rate of top 10 search results in Google, 2012.
- [37] R. Edell and P. Varaiya. Providing internet access: What we learn from index. *IEEE Network*, 13(5), 1999.
- [38] B. Edelman and B. Lockwood. Measuring bias in “organic” web search. <http://www.benedelman.org/searchbias/>, Jan 2011.
- [39] Federal Communications Commission. Appropriate framework for broadband access to the internet over wireline facilities. Technical Report FCC 05-151, CC Docket No. 02-33, Sept 2005.
- [40] R. Frieden. Internet Packet Sniffing and its Impact on the Network Neutrality Debate and the Balance of Power between Intellectual Property Creators and Consumers. Technical report, SSRN, 2007. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=995273.
- [41] D. Fudenberg and J. Tirole. *Game Theory*. MIT Press, Cambridge, Massachusetts, 1991.
- [42] V. Fux and P. Maillé. A rating-based network selection game in heterogeneous systems. In *Proc. of 8th Euro-NF Conference on Next Generation Internet (NGI)*, Karlskrona, Sweden, 2012.
- [43] V. Fux and P. Maillé. Incentivizing efficient load repartition in heterogeneous wireless networks with selfish delay-sensitive users. In *Proc. of 8th International Workshop on Internet Charging and QoS Technologies*, Zurich, Switzerland, 2013.
- [44] V. Fux, P. Maillé, J.-M. Bonnin, and N. Kaci. Efficiency or fairness: managing applications with different delay sensitivities in heterogeneous wireless networks. In *Proc. of IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, Madrid, Spain, 2013.
- [45] V. Fux, P. Maillé, and M. Cesana. Price competition between road side units operators in vehicular networks. In *Proc. of IFIP Networking*, Trondheim, Norway, 2014.
- [46] A. Ganesh, D. Gunawardena, P. Jey, L. Massoulié, and J. Scott. Efficient quarantining of scanning worms: Optimal detection and co-ordination. In *Proceedings of IEEE INFOCOM 2006*, Barcelona, Spain, April 2006.
- [47] P. Hand, M. Chiang, R. Calderbank, and S. Rangan. Network rate allocation with content provider participation. In *Proceedings of INFOCOM*, pages 990–998, 2009.
- [48] W.A. Hanson and K. Kalyanam. *Internet Marketing and E-Commerce*. International student edition. Thomson/South-Western, Mason, Ohio, 2007.

- [49] D. A. Hyman and D. J. Franklyn. Search neutrality and search bias: An empirical perspective on the impact of architecture and labeling. Illinois Program in Law, Behavior and Social Science Paper No. LE13-24; Univ. of San Francisco Law Research Paper No. 2013-15. Available at <http://ssrn.com/abstract=2260942>, May 2013.
- [50] Inria. Inria’s response to ARCEP consultation about network neutrality, 2012.
- [51] ITU. The world in 2014 - ict facts and figures. Available from <http://www.itu.int/>, 2014.
- [52] L. Jiang, V. Anantharam, and J. Walrand. Efficiency of selfish investments in network security. In *Proceedings of 3rd Workshop on the Economics of Networks, Systems, and Computation*, Seattle, WA, USA, Aug 2008.
- [53] R. Johari, G. Y. Weintraub, and B. Van Roy. Investment and market structure in industries with congestion. *Operations Research*, 58(5):1303–1317, 2010.
- [54] N. Kaci, P. Maillé, and J.-M. Bonnin. Performance of wireless heterogeneous networks with always-best connected users. In *Proc. of the 5th Conference on Next Generation Internet Networks (NGI)*, Aveiro, Portugal, 2009.
- [55] G. Karakostas, S. G. Kolliopoulos, and G. Stavros. Edge pricing of multi-commodity networks for heterogeneous selfish users with elastic demands. In *Proceedings of 12th annual international conference on Computing and Combinatorics*, 2006.
- [56] George Karakostas and Stavros G. Kolliopoulos. Edge pricing of multicommodity networks for heterogeneous selfish users. In *Proceedings of 45th Annual IEEE Symposium on Foundations of Computer Science (FOCS’04)*, pages 268–276, 2004.
- [57] H.-S. Kim and C.-H. Yoon. Determinants of subscriber churn and customer loyalty in the Korean mobile telephony market. *Telecommunications Policy*, 28(9-10):751–765, 2004.
- [58] M. Kodialam and T. V. Lakshman. Detecting network intrusions via sampling: A game theoretic approach. In *Proceedings of IEEE INFOCOM*, 2003.
- [59] Y. A. Korilis, A. A. Lazar, and A. Orda. Capacity allocation under non-cooperative routing. *IEEE Transactions on Automatic Control*, 42(3), 1997.
- [60] E. Koutsoupias and C. Papadimitriou. Worst-case equilibria. In *Proc of 16th Annual Symposium on Theoretical Aspects of Computer Science (STACS 1999)*, volume 1563 of *Lecture Notes in Computer Science*, pages 404–413, 1999.

- [61] A.M. Kovaks. Internet peering and transit. Technical report, Technology Policy Institute, 2012.
- [62] B. Krebs. I'll take 2 Mastercards and a Visa, please. <http://krebsonsecurity.com/2010/09/ill-take-2-mastercards-and-a-visa-please/>, Sept 2010.
- [63] M. Labbé, P. Marcotte, and G. Savard. A bilevel model of taxation and its application to optimal highway pricing. *Management Science*, 44(12):1608–1622, 1998.
- [64] P. L'Ecuyer, P. Maillé, N. Stier-Moses, and B. Tuffin. Revenue-maximizing rankings for online platforms with quality-sensitive consumers. <https://hal.inria.fr/hal-00953790>, 2014.
- [65] M. Lelarge and J. Bolot. Economic incentives to increase security in the internet: The case for insurance. In *Proc. of IEEE INFOCOM*, Rio de Janeiro, Brazil, 2009.
- [66] T.M. Lenard and R.J. May (Eds.). *Net Neutrality or Net Neutering: Should Broadband Internet Services be Regulated*. Springer, 2006.
- [67] R. T. B. Ma, D.-M. Chiu, J. C. S. Lui, V. Misra, and D. Rubenstein. Interconnecting eyeballs to content: A Shapley value perspective on isp peering and settlement. In *Proceedings of the International Workshop on Economics of Networked Systems (NetEcon)*, pages 61–66, 2008.
- [68] R.T.B. Ma, D.-M. Chiu, J.C.S. Lui, V. Misra, and D. Rubenstein. On cooperative settlement between content, transit and eyeball internet service providers. In *Proceedings of the ACM International Conference on Emerging Networking EXperiments and Technologies (CoNEXT)*, 2008.
- [69] P. Maillé, E. Markakis, M. Naldi, G. Stamoulis, and B. Tuffin. Sponsored search auctions: an overview of research with emphasis on game theoretic aspects. *Electronic Commerce Research*, 12:265–300, 2012.
- [70] P. Maillé, M. Naldi, and B. Tuffin. Competition for migrating customers: a game-theoretic analysis in a regulated regime. In *Proc. of IEEE GLOBECOM*, New Orleans, LA, USA, 2008.
- [71] P. Maillé, M. Naldi, and B. Tuffin. Price war with migrating customers. In *Proc. of 17th IEEE International Symposium on Modelling, Analysis and Simulation of Computer and Telecommunication Systems (MASCOTS)*, London, UK, 2009.
- [72] P. Maillé, M. Naldi, and B. Tuffin. Understanding and preventing tacit collusion among telecommunication operators. In *Proc. of NET-COOP*, Eindhoven, The Netherlands, 2009.

- [73] P. Maillé, P. Reichl, and B. Tuffin. Interplay between security providers, consumers, and attackers: a weighted congestion game approach. In *Proc. of the 2nd Conference on Decision and Game Theory for Security (GameSec 2011)*, College Park, MD, USA, 2011.
- [74] P. Maillé, P. Reichl, and B. Tuffin. Internet governance and economics of network neutrality. In Antonis Antoniadis and Burkhard Stiller, editors, *Telecommunications Economics - Selected Results of the COST Action IS605 Econt@Tel*, pages 108–116. Lecture Notes in Computer Science 7216, Springer Verlag, 2012.
- [75] P. Maillé and N. Stier-Moses. Eliciting coordination with rebates. *Transportation Science*, 43(4):473–492, 2009.
- [76] P. Maillé and L. Toka. Managing a peer-to-peer data storage system in a selfish society. *IEEE JSAC, Special Issue on "Game Theory in Communication Systems"*, 26(7):1295–1301, 2008.
- [77] P. Maillé and B. Tuffin. Analysis of price competition in a slotted resource allocation game. In *Proc. of IEEE INFOCOM*, Phoenix, AZ, USA, 2008.
- [78] P. Maillé and B. Tuffin. Price war with partial spectrum sharing for competitive wireless service providers. In *Proc. of IEEE GLOBECOM*, Honolulu, Hawaii, USA, 2009.
- [79] P. Maillé and B. Tuffin. Price war in heterogeneous wireless networks. *Computer Networks*, 54(13):2281–2292, 2010.
- [80] P. Maillé and B. Tuffin. Adword auction bidding strategies of budget-limited advertisers on competing search engines. In *Proc. of the 7th International Conference on Network and Service Management (CNSM 2011)*, Paris, France, 2011.
- [81] P. Maillé and B. Tuffin. Sponsored search engines in competition: Advertisers behavior and engines optimal ranking strategies. In *Proc. of 19th IEEE International Symposium on Modelling, Analysis and Simulation of Computer and Telecommunication Systems (MASCOTS)*, Singapore, 2011.
- [82] P. Maillé and B. Tuffin. Competition among providers in loss networks. *Annals of Operations Research, Special Volume on Operations Research and Systems (ALIO-INFORMS)*, 199:3–22, 2012.
- [83] P. Maillé and B. Tuffin. *Telecommunication Network Economics: From Theory to Applications*. Cambridge University Press, 2014.
- [84] P. Maillé, B. Tuffin, and P. Reichl. Of threats and costs: A game-theoretic approach to security risk management. In N. Gulpinar, P. Harrison, and B. Rustem, editors, *Performance Models and Risk Management in Communication Systems*, volume 46, pages 33–54. Springer “Optimization and Its Applications” series, 2010.

- [85] P. Maillé, B. Tuffin, and J.-M. Vigne. Economics of technological games among telecommunication service providers. In *Proc. of IEEE GLOBECOM*, Miami, FL, USA, 2010.
- [86] P. Maillé, B. Tuffin, and J.-M. Vigne. Technological investment games among wireless telecommunications service providers. *International Journal on Network Management - Special Issue: Economic Traffic Management*, 21(1):65–82, 2011.
- [87] P. Maillé, B. Tuffin, and J.-M. Vigne. Competition between wireless service providers sharing a radio resource. In *Proc. of 11th IFIP Networking Conference*, Prague, Czech Republic, 2012.
- [88] P. Marbach. Priority service and max-min fairness. In *Proceedings of IEEE INFOCOM*, 2002.
- [89] M. Mavronicolas, I. Milchtaich, B. Monien, and K. Tiemann. Congestion games with player-specific constants. In *Proc. of 32nd International Symposium on Mathematical Foundations of Computer Science (MFCS)*, pages 633–644, Český Krumlov, Czech Republic, 2007.
- [90] I. Milchtaich. Congestion games with player-specific payoff functions. *Games and Economic Behavior*, 13(1):111–124, Mar 1996.
- [91] I. Milchtaich. Weighted congestion games with separable preferences. *Games and Economic Behavior*, 67(2):750–757, Nov 2009.
- [92] J. Musacchio, J. Walrand, and G. Schwartz. Network neutrality and provider investment incentives. In *Signals, Systems and Computers, 2007. ACSSC 2007. Conference Record of the Forty-First Asilomar Conference on*, pages 1437–1444, nov. 2007.
- [93] A. Nagurney. *Supply Chain Network Economics: Dynamics of Prices, Flows And Profits*. Edward Elgar Publishing, 2006.
- [94] M. Naldi. A simple model for the effectiveness of delaying strategies for telecommunications churn reduction. In *Proceedings of 10th International Conference on Computer Modelling and Simulation (EUROSIM)*, Cambridge, UK, 2008.
- [95] J. Nash. Two-person cooperative games. *Econometrica: Journal of the Econometric Society*, pages 128–140, 1953.
- [96] P. Njoroge, A. Ozdaglar, N. Stier-Moses, and G. Weintraub. Investment in two sided markets and the net neutrality debate. *Review of Network Economics*, 2014. Forthcoming.
- [97] J. R. Norris. *Markov Chains*. Cambridge University Press, 1998.
- [98] A. Odlyzko. The history of communications and its implications for the Internet. Technical report, AT&T Labs, 2000.

- [99] BEUC The European Consumer’s Organization. Fair internet search - remedies in google case. Position paper, available from www.beuc.org, March 2013.
- [100] M. J. Osborne and A. Rubinstein. *A Course in Game Theory*. MIT Press, 1994.
- [101] G. Perakis. The “Price of Anarchy” under nonlinear and asymmetric costs. *Mathematics of Operations Research*, 32(3):614–628, Aug 2007.
- [102] A. C. Pigou. *The Economics of Welfare*. Macmillan, 1920.
- [103] J. Qi, Y. Zhang, Y. Zhang, and S. Shi. Treelogit model for customer churn prediction. In *Proceedings of IEEE Asia-Pacific Conference on Services Computing*, pages 70–75, Washington, DC, USA, 2006.
- [104] P. Reichl, R. Schatz, and B. Tuffin. Logarithmic laws in service quality perception: Where microeconomics meets psychophysics and quality of experience. *Telecommunication Systems*, 2013. To appear, DOI: 10.1007/s11235-011-9503-7.
- [105] T. Roughgarden. The price of anarchy is independent of the network topology. *Journal of Computer and System Sciences*, 67:341–364, 2003.
- [106] T. Roughgarden. Stackelberg scheduling strategies. *SIAM Journal on Computing*, 33(2):332–350, 2004.
- [107] T. Roughgarden. *Selfish Routing and the Price of Anarchy*. MIT Press, 2005.
- [108] T. Roughgarden. On the severity of Braess’s paradox: Designing networks for selfish users is hard. *Journal of Computer and System Sciences*, 72(5):922–953, 2006. Special Issue on FOCS 2001.
- [109] T. Roughgarden and É. Tardos. How bad is selfish routing? *Journal of the ACM*, 49(2):236–259, Mar 2002.
- [110] T. Roughgarden and E. Tardos. Bounding the inefficiency of equilibria in nonatomic congestion games. *Games and Economic Behavior*, 47(2):389–403, 2004.
- [111] W. H. Sandholm. Potential games with continuous player sets. *Journal of Economic Theory*, 97(1):81–108, Mar 2001.
- [112] W. H. Sandholm. Large population potential games. *Journal of Economic Theory*, 144(4):1710–1725, Jul 2009.
- [113] H. Saunders. IP interconnection: trends and emerging issues. http://berec.europa.eu/files/news/ofcom_ipic.pdf, 2012.

- [114] D. Schmeidler. Equilibrium points of nonatomic games. *Journal of Statistical Physics*, 7(4):295–300, 1973.
- [115] A. S. Schulz and N. E. Stier-Moses. Efficiency and fairness of system-optimal routing with user constraints. *Networks*, 48(4):223–234, 2006.
- [116] B. Schwarz. Content delivery networks 3.0. Technical report, CTOiC White paper, 2013.
- [117] F. Sorensen. The Norwegian model for net neutrality. <http://is.gd/HtQuuJ>, 2013.
- [118] Symantec. State of Enterprise Security. http://www.symantec.com/content/en/us/about/presskits/SES_report_Feb2010.pdf, Feb 2010.
- [119] K. T. Talluri and G. J. van Ryzin. *The theory and practice of revenue management*. International series in operations research & management science. Kluwer Academic Publ, Boston, MA, 2004.
- [120] L. Toka and P. Maillé. Managing a peer-to-peer backup system: Does imposed fairness socially outperform a revenue-driven monopoly? In *Proc. of the 4th International Workshop on Grid Economics and Business Models (GECON)*, Rennes, France, 2007.
- [121] H.R. Varian. Position auctions. *International Journal of Industrial Organization*, 25(6):1163–1178, 2007.
- [122] H. von Stackelberg. *Marktform und Gleichgewicht*. Springer, Wein, 1934. English translation: *The Theory of the Market Economy*, published in 1952 by Oxford University Press.
- [123] J. Walrand. Economic models of communication networks. In Z. Liu and C. Xia, editors, *Performance Modeling and Engineering*, chapter 3. Springer, 2008.
- [124] J. G. Wardrop. Some theoretical aspects of road traffic research. *Proceedings of the Institute of Civil Engineers*, 1:325–378, 1952.
- [125] K. Wieland. The customer retention challenge. *Telecommunications*, 40(10):14–17, 2006.
- [126] J. D. Wright. Defining and measuring search bias: Some preliminary evidence. George Mason Law & Economics Research Paper 12-14, George Mason University School of Law, 2012.
- [127] T. Wu. Network neutrality, broadband discrimination. *Journal of Telecommunications and High Technology*, 2003.
- [128] T. Zeller. Black market in stolen credit card data thrives on Internet. *The New York Times*, Jun 2005.

Appendix A

Main publications evoked in Chapter 2

Eliciting Coordination with Rebates

Patrick Mailé

Institut Telecom, Telecom Bretagne, Université Européenne de Bretagne,
Cesson Sévigné, France, patrick.maille@telecom-bretagne.eu

Nicolás E. Stier-Moses

Graduate School of Business, Columbia University, Uris Hall 418,
New York, New York 10027, stier@gsb.columbia.edu

This article considers a mechanism based on rebates that aims at reducing congestion in urban networks. The framework helps select rebate levels so that enough commuters switch to modes that are under used. Indeed, getting a relative small number of drivers to switch to public transportation can significantly improve congestion. This mechanism is modeled by a Stackelberg game in which the transportation authority offers rebates, and participants factor them into the costs of each mode. A new Wardrop equilibrium arises when participants selfishly select one of the modes of transportation with the lowest updated costs. Rebate levels are chosen taking into account not only the potential reduction of the participants' cost, but also the cost of providing those rebates. Part of the budget for rebates may come from the savings that arise from the more efficient use of capacity. We characterize the Stackelberg equilibria of the game, and describe a polynomial-time algorithm to compute the optimal rebates for each mode. In addition, we provide tight results on the worst-case inefficiency of the resulting Wardrop equilibrium, measured by the so-called price of anarchy. Specifically, we describe the tradeoff between the sensitivity of the owner towards rebate costs and the worst-case inefficiency of the system.

Key words: network pricing; subsidies as incentives; Wardrop equilibrium; Stackelberg games; price of anarchy

History: Received: September 2008; revision received: March 2009; accepted: April 2009. Published online in *Articles in Advance* October 16, 2009.

1. Introduction

Congestion in most large cities in the world is prevalent. The Urban Mobility Study, a survey conducted by the Texas Transportation Institute (Schrank and Lomax 2007), estimated that the *congestion bill* related to automobile traffic, in the U.S. alone, amounts to \$78.2 billion in 2005. This cost estimate is based on the following two components: 4.2 billion hours of delay that people lose to highway congestion plus 2.9 billion gallons worth of fuel. Given those figures, even a small improvement in the efficiency of the road traffic system implies that a large sum of money and time could be saved. Furthermore, a recent study by the (Partnership for New York City 2006, p. 3) concluded that “traffic delays add to logistical, inventory, and personnel costs that annually amount to an estimated \$1.9 billion in additional costs of doing business and \$4.6 billion in unrealized business revenue.”

In most urban transportation networks, commuters do not have to pay the cost they impose to others by a particular choice of mode and route. Because of these under-payments, decisions—which are mostly influenced by a desire to get to the destination as fast as possible and as cheap as possible—lead to choices that do not use the available capacity of the network well. Because congestion increases sharply with road use,

having relatively few drivers switch to other modes significantly improves commute times. Starting with the seminal idea of Vickrey (1955, 1969), many transportation economists have advocated the use of congestion pricing to achieve this goal. The scheme forces drivers to pay a toll when entering congested areas as an incentive to switch to other modes of transportation (operational details differ according to the concrete implementation). The underlying idea is to charge drivers the externality they impose to others because when commuters internalize these externalities, the corresponding choices maximize the system welfare.

Singapore introduced congestion pricing in 1975, London in 2003 (Santos 2005; Santos and Fraser 2006), and Stockholm in 2007. Increasingly, many large cities have been debating whether a congestion pricing scheme should be adopted, New York City being the most prominent example in the United States. Nevertheless, it has been very hard to implement congestion pricing because of technical, economical, and political problems (e.g., the proposal in New York was not implemented after the State Assembly blocked it in 2008). Even though proponents claim it will decrease the delay costs generated by congestion, will curb harmful emissions and will reduce the dependence

on oil, the main concern that opponents raise comes from the perspective of social equity. Opponents favor other alternatives such as restricting some cars from driving to congested areas on some days of the week, increasing the taxes for parking, and offering incentives for tele-commuting, among many others.

Introducing congestion pricing is not likely to have a large impact among the better-positioned segment of society. They will continue to drive because they can afford to pay the corresponding charge. In contrast, the not-as-well positioned segment will be relegated to the less desirable options because they cannot afford to pay the tolls. Some articles suggest different measures to alleviate problems of inequity raised by this type of mechanisms (e.g., Starkie 1986; Button and Verhoef 1998).

The most important practical questions are which incentives to offer, at what level, how much they will improve social welfare, and who will be affected. Most cities that do use congestion pricing, use a *second-best* approach because charging tolls on each arc is not feasible or practical, even with electronic toll collection systems.¹ Besides the implementation cost, charging (potentially variable) tolls everywhere makes it more complicated for the driver to select a route. In the future, this may be less of a problem because the market penetration of route guidance devices is likely to be larger, and it is technologically feasible that these devices receive broadcasts with the current values of tolls. Most cities that have adopted congestion pricing decided to charge a flat daily fee that has to be paid on each day a driver wants to access the central business district of the city. Although a flat fee does not elicit the most efficient choices, it is conducive to increase the social welfare. Indeed, the high cost of the charge is enough to provide a detriment to some drivers who will switch to other modes of transportation. Unfortunately, an implementation of a congestion pricing scheme is not likely to allow for much room for experimentation. If not done right initially, expected benefits may not be realized, thereby invalidating the whole effort and potentially jeopardizing the political viability of a second try. Quantitative models can be used upfront to help policymakers make decisions and compare proposals.

This article initiates the study of an approach that complements congestion pricing. Although congestion pricing considers only (positive) tolls, there is no reason not to use negative tolls, which we refer to as

rebates or subsidies. Often, the proceeds of congestion pricing are used to fund improvements in public transportation, but very rarely are they used to reduce operating costs by subsidizing fares. On the contrary, it has been documented that in some cases public transportation fares increased after the introduction of congestion pricing (Wichiensin, Bell, and Yang 2007).

In the context of the debate around the introduction of congestion pricing in New York City, Kheel (2008) recently proposed to completely eliminate the fare for public transportation by paying operating costs with the congestion charges. His own words, “[t]his more balanced plan will result in the equivalent of a \$20 after-tax pay raise for every transit-using worker in the city. Automobile drivers will benefit too, as traffic is vastly reduced” (p. 4), capture why rebates provide a more equitable solution than congestion pricing alone. Because having no fare may or may not be optimal from a system welfare perspective, we focus on finding the optimal level of subsidies. We assume that if congestion pricing is used, toll charges are already fixed. Specifically, we concentrate on mode decisions in the case of linear congestion costs and homogeneous demand. The main assumption of this article is that a city can set apart some funds that it will use to subsidize users of certain modes by offering a rebate on part of the fare. As Kheel said, rebates go to the population segment that selects the least-desirable modes, thus compensating users that switched out from their preferred choices. The fact that most public transportation systems in the world are subsidized provides anecdotal evidence that a mechanism based on subsidies is easier to accept by the constituents than congestion pricing. Well-chosen rebates lead to more efficient choices. Less people will drive, congestion will be reduced, and the total commute time will decrease. Eventually, some of the benefits will be transferred back to the provider of subsidies in the form of additional taxes, reduced CO₂ emissions, reduced health-care costs, etc. For example, operating expenses of companies that do deliveries will be reduced, thereby improving their bottom line. The additional taxes can be used to recover a fraction of the money that was set aside initially.

Cities do not have unlimited resources and, thus, cannot offer large rebates if they do not also implement congestion pricing. For that reason, we look at the problem of finding rebates that maximize user welfare, taking into account a limited budget. This budget relates to the value placed on the reduction of commute times. In the extreme case when commute times are all that matters and the budget is large, rebates will be set to make experienced costs equal to zero. (Compare this to the costs experienced by commuters under Kheel’s proposal, which are not

¹ Electronic toll collection systems—currently in use in cities that implement congestion pricing and in many tolled highways—eliminate the need to stop at a toll plaza. In general, these systems have three components: a toll tag, which is placed inside the vehicle; an overhead antenna, which reads the toll tag and collects the toll; and video cameras to identify toll evaders.

zero because commuters still face the disutility arising from the time invested to complete the trip.) On the other extreme, when the reduction in commute time is not deemed important or when the budget for rebates is small, rebates will not be offered and users will experience the full cost arising from the time and the fare or toll.

Even with optimal rebates, the coordination generated by this approach may not be enough to achieve a significant increase in welfare. Henceforth, we want to quantify the coordinating power of rebates. Koutsoupias and Papadimitriou (1999) defined the *price of anarchy* as the worst-case ratio of the social welfare under a user equilibrium attained without coordination to that with socially-optimal choices. This indicator has been used to estimate the potential increase in welfare provided by a given mechanism, and to gauge whether the opportunity cost is large enough to outweigh the implementation cost and justify its use. To answer this question, we compare the total welfare generated by optimal rebates to that when rebates are set to zero. We show that when the budget is large enough, one can have a transportation pattern that is significantly more efficient than the status quo.

Main Contributions and Structure of the Paper

Although others considered rebates implicitly (as negative taxes), to our knowledge, this is the first article that formally studies the computation of optimal rebates with the goal of coordinating a congestion game. Our main contribution is a mechanism that provides incentives for coordination that does not penalize participants, but instead rewards those that were worse-off without such a mechanism by offering them a rebate. Our social cost function explicitly considers the transfer payments to capture the cost of providing rebates, and the mechanism aims to minimize this more general expression of cost. Instead, most of the earlier articles that studied the coordinating power of tolls and taxes consider a social cost equal to the sum of costs for all participants, thus ignoring the costs and benefits of payments because they are transfers that stay in the system (see Beckmann, McGuire, and Winsten 1956; Bergendorff, Hearn, and Ramana 1997; Labbé, Marcotte, and Savard 1998 for classical references; Cole, Dodis, and Roughgarden 2006 is a notable exception that considers transfer payments as part of the social cost).

We consider a *Stackelberg game* in which the system owner (e.g., the city or the transportation authority) is the *leader* and the participants are *followers* (von Stackelberg 1934). In a first stage, the leader offers rebates in each arc; in a second stage, participants selfishly select arcs that have minimal cost, taking rebates into consideration. Focusing on the modal

choice problem, we characterize the optimal rebates in the case of affine cost functions and networks with multiple arcs that connect two nodes (the alternative modes of transportation are substitutes). Many examples of recent work in this area such as Engel, Fischer, and Galetovic (2004), Xiao, Yang, and Han (2007), Acemoglu and Ozdaglar (2007), and Wichiensin, Bell, and Yang (2007), also consider this type of simple networks. Although Labbé, Marcotte, and Savard (1998) present results for general networks, they do it for a simplified model that ignores congestion effects, which is an important feature of our model.

We first prove that if the system owner values the perceived cost more than rebates, then an optimal strategy for the leader is to refund each participant the perceived cost at each arc under a system optimal solution. When the system owner is more sensitive to the investment in rebates than to the perceived cost, it will offer rebates in the modes that are under-used. We also establish an upper bound on the proportion of participants that receive a positive rebate. Using our characterization of Stackelberg equilibria, we provide a polynomial-time algorithm that selects the arcs where rebates should be offered, and computes the optimal rebates for those arcs. This enables us to derive an explicit formula for the resulting social cost, from where we compute the price of anarchy, expressed as a function of the predisposition of the system owner to offer rebates. The main conclusion is that when the system owner is willing to offer rebates, the resulting solution has low social cost. Conversely, when the system owner cannot afford to provide significant rebates, the resulting outcome is close to a Wardrop equilibrium.

This paper is organized as follows. First, we review the literature in §2. In §3, we introduce the model and the performance measures of interest. Section 4 offers some results for general network topologies, while §5 focuses on instances with parallel arcs (substitutes) and characterizes the optimal rebates. In §6, we compute the price of anarchy for instances with affine cost functions. Finally, we offer some concluding remarks and open questions in §7.

2. Connections to the Literature

We work under the setting first described by Wardrop (1952). The corresponding equilibrium concept has been called a *Wardrop equilibrium*, which under mild conditions coincides with a Nash equilibrium (Haurie and Marcotte 1985). Although in some cases a system may be better off without a coordination mechanism because the overall implementation and operating costs may outweigh the potential benefits, equilibria have been found to be too inefficient in many applications of interest. This makes it necessary to

coordinate participants to mitigate the adverse effects of the misalignment of incentives. As imposing decisions to users is not an option in most real-world situations, equilibria can be improved by system (re)design (Roughgarden 2006), by considering routing part of the flow preemptively (Korilis, Lazar, and Orda 1995), or by using pricing mechanisms to create incentives (Bergendorff, Hearn, and Ramana 1997; Labbé, Marcotte, and Savard 1998). This article considers the third approach.

Even before the work of Vickrey (1955), economists such as Dupuit (1849), Pigou (1920), and Knight (1924) proposed to use pricing so participants internalize the externalities, defined as the additional cost they impose to others. If implemented properly, this results in equilibria that are efficient from a social welfare perspective. For a complete treatment of network pricing and many additional references, see, e.g., the book by Yang and Huang (2005).

We study a mechanism based on rebates. Rebates are used in logistics, supply chain management, and marketing, with the objective of revenue maximization as well as to create incentives for coordination (Gerstner and Hess 1991; Ali, Jolson, and Darmon 1994; Taylor 2002; Chen, Li, and Simchi-Levi 2007). We find the optimal rebates by solving a Stackelberg equilibrium problem, which structurally is a *mathematical program with equilibrium constraints* (MPEC). There are relatively standard optimization techniques to compute solutions to this type of problems. For a background on MPECs and solution methods, we refer the interested reader to the book by Luo, Pang, and Ralph (1996). One could get the optimal rebates and the corresponding modal choices from a Stackelberg equilibrium computed numerically; actually, computational studies are routinely used to analyze congestion-charging systems. In our case, though, finding the optimal rebates numerically is not enough for our purposes because such an analysis does not provide the structure needed to understand how much benefit the mechanism provides.

Recently, many authors have studied the maximum efficiency-loss under an equilibrium, using social welfare to measure the quality of solutions. Koutsoupias and Papadimitriou (1999) defined the price of anarchy as the largest possible ratio of the social cost at an equilibrium to the minimum attainable social cost (the term itself was coined by Papadimitriou 2001). Starting from the work of Roughgarden and Tardos (2002), the price of anarchy in transportation networks (the setting suggested by Wardrop 1952) has been characterized by Roughgarden (2003), Correa, Schulz, and Stier-Moses (2004), Chau and Sim (2003), and Perakis (2007), who successively considered more general assumptions. It turns out that equilibria of these

games are reasonably efficient; for example, when congestion costs increase linearly with flow, the extra total cost of an equilibrium does not exceed 33% more than that of a system optimum. For other typical classes of functions, the inefficiency is somewhat larger but bounded. Nevertheless, for practical purposes these inefficiencies are too high; even smaller improvements translate to big savings for societies and governments (recall the figures provided by the Urban Mobility Study). Hence, some researchers looked for improved measures of inefficiency (Friedman 2004; Qiu et al. 2006; Schulz and Stier-Moses 2006; Correa, Schulz, and Stier-Moses 2008), while others focused on mechanisms to improve the inefficiency itself. Some references that look at pricing mechanisms from the perspective of the price of anarchy are Koutsoupias (2004), Karakostas and Kolliopoulos (2005), Cole, Dodis, and Roughgarden (2006), Wichensin, Bell, and Yang (2007), Xiao, Yang, and Han (2007), and Yang, Xu, and Heydecker (2009).

The study of the inefficiency of equilibria has recently received increased attention from researchers in various communities such as Transportation, Operations Research, Operations Management, Economics, and Computer Science. Consequently, there is a growing amount of interdisciplinary literature on the price of anarchy. For example, some additional references in the application domains of telecommunication and distribution networks are the articles by Johari and Tsitsiklis (2004), Golany and Rothblum (2006), Perakis and Roels (2007), Acemoglu and Ozdaglar (2007), and Johari, Weintraub, and Van Roy (2009).

3. Description of the Model

In this section, we introduce the model and its necessary notation. We consider the framework of *network games*, originally introduced by Wardrop (1952) and first analyzed formally by Beckmann, McGuire, and Winsten (1956). An instance of our problem is given by a network, cost functions, a system owner, and participants. The network encodes the modal and route choices, and the cost functions associated to each arc model congestion and charges. The system owner defines the level of rebates, and participants—who are infinitesimally small—select a route from their origins to their destinations with minimum cost.

The network is represented by a directed graph (V, A) , where V is a set of vertices and A is a set of arcs. In general the graph may be arbitrary, but we will concentrate on the case where A is a set of parallel links that represent each of the modes. When possible, we will present results for general graphs to allow for route choice. For a total flow of x_i in an arc $i \in A$, the cost of traversing it is $c_i(x_i)$. Functions c_i , referred to as *cost functions*, are assumed to be affine on x_i for the

main results of this study. When possible, we will also consider more general cost functions that are nonnegative, nondecreasing, differentiable, and convex. Furthermore, we assume that cost functions are separable, meaning that the only argument of a cost function is the flow along that arc.

As we described in the introduction, the most typical example of this model is given by a urban network in which commuters have to decide between driving their cars, walking or taking one form of public transportation. Cost functions encode commute time, delays, and fares, all of which are assumed to be expressed in monetary units, and indicate the overall equivalent cost perceived by users for traversing a link. Although we do not explicitly model congestion pricing, it can be partially incorporated in our model by adding the corresponding charges to the cost functions.

The system owner offers rebates to elicit coordination. We denote the rebate for arc i by $s_i \geq 0$. As participants will not be reimbursed more than their cost, we restrict the actual reimbursement to not exceed $c_i(x_i)$. Hence, as the rebates are announced before participants make their selections, participants receive a rebate up to the cost of the arc. Indeed, the experienced cost is

$$c_i^s(x_i) := [c_i(x_i) - s_i]^+,$$

where $[y]^+$ denotes the positive part of y . Equivalently, the actual rebate equals $\min(s_i, c_i(x_i))$. Collectively, we denote the vector of all rebates with $s \in \mathbb{R}_+^A$.

Each participant selects the arc in A that corresponds to the mode of choice. For the results in which we also consider route selection, participants are associated with a pair of nodes, called an origin-destination pair (OD-pair), and have to select a path from their origins to their destinations. Let us denote the set of OD-pairs by K , the demand corresponding to OD-pair $k \in K$ by r_k , and the total demand $\sum_{k \in K} r_k$ by r . In addition, we refer to all the possible paths connecting an OD-pair $k \in K$ by \mathcal{P}_k and we let $\mathcal{P} := \bigcup_{k \in K} \mathcal{P}_k$. For the mode-choice model, there is a single OD pair that consists of the only two nodes.

We use flows to encode all participants' decisions, as specific identities are irrelevant. A flow x is feasible if it is nonnegative and it satisfies all demand constraints. Mathematically, this is represented by the set $\{x \in \mathbb{R}_+^{\mathcal{P}} : \sum_{P \in \mathcal{P}_k} x_P = r_k \text{ for all } k \in K\}$. The flow on an arc x_i is given by the sum over the paths $\sum_{P \in \mathcal{P} : P \ni i} x_P$.

Competition leads participants from the same OD-pair to select paths of cheapest equal cost because otherwise they would have an incentive to change their selection. This is the basis of the traditional solution concept called *Wardrop equilibrium* (Wardrop 1952).

DEFINITION 3.1. A flow x^{WE} is a Wardrop equilibrium of a network game (without rebates) if it is feasible, and for all k and all $P, Q \in \mathcal{P}_k$ such that $x_P^{\text{WE}} > 0$, $c_P(x^{\text{WE}}) \leq c_Q(x^{\text{WE}})$, where $c_P(x) := \sum_{i \in P} c_i(x_i)$.

The previous definition provides us with a solution concept that models the behavior of the second stage players:

DEFINITION 3.2. If the system owner selects the rebate vector s , participants select a solution x^s , which is a Wardrop equilibrium with respect to cost functions $[c_i(\cdot) - s_i]^+$.

For a given rebate vector s , the corresponding Wardrop equilibrium x^s always exists because the modified cost functions $[c_i(\cdot) - s_i]^+$ are continuous (Beckmann, McGuire, and Winsten 1956). In general, the equilibrium x^s need not be unique but if there are more than one, the prevailing costs under different equilibria are equal. Because any equilibrium can arise in practice, we consider an arbitrary one.

We now focus on the best strategy for the system owner. Because it is the leader of the Stackelberg game and it fixes the rebates knowing that participants are going to select a Wardrop equilibrium, its optimal strategy is to select the vector s that minimizes the *social cost*, defined as the sum of the costs of all parties in the game (Mas-Colell, Whinston, and Green 1995). This objective function includes the perceived cost experienced by each participant and the amount the system owner invests in rebates. As the system owner may be more sensitive to one of the terms than to the other, we consider a parameter $\rho \geq 0$ that transforms the rebate investment into social cost units. Section 6.1 provides further justification for this choice of social cost functions. (Note that we can alternatively define the social cost as the sum of the real costs that participants face by using a modified coefficient as shown in (1b).)

DEFINITION 3.3. The strategy (s, x^s) is a *Stackelberg equilibrium* if the vector of rebates s minimizes the *social cost*, defined as

$$C_\rho(s) := \underbrace{\sum_{i \in A} x_i^s [c_i(x_i^s) - s_i]^+}_{\text{participants' perceived cost}} + \rho \underbrace{\sum_{i \in A} x_i^s \min(c_i(x_i^s), s_i)}_{\text{cost of rebates}}, \quad (1a)$$

which can also be expressed as

$$\underbrace{\sum_{i \in A} x_i^s c_i(x_i^s)}_{\text{participants' real cost}} + (\rho - 1) \underbrace{\sum_{i \in A} x_i^s \min(c_i(x_i^s), s_i)}_{\text{cost of rebates}}. \quad (1b)$$

In this case, we refer to s as an *optimal rebate vector*.

The parameter ρ allows the system owner to control the tradeoff between the social cost of the solution and its investment. Alternatively, it can be viewed as the Lagrangian multiplier of the system owner's budget constraint. In fact, $1/\rho$ represents the investment

the system owner is willing to commit to make the participants' perceived cost decrease by one unit:

- $\rho = 1$ corresponds to the situation in which the system owner is only interested in minimizing the participants' real cost $\sum_{i \in A} x_i c_i(x_i)$, regardless of the rebate cost (see (1b)).
- $\rho = +\infty$ corresponds to the situation in which the system owner does not want to spend any money on rebates. Here, the outcome will be a Wardrop equilibrium, as without rebates.
- Values of $\rho < 1$ correspond to the case where the network planner values the participants' perceived cost more than its own investments.

As we said in §2, the Stackelberg equilibrium can be found by solving an MPEC. If the leader wants to compute optimal rebates for a particular instance, there are relatively standard optimization techniques to solve this problem, even if more constraints are added to the problem (e.g., restrict rebates to a subset of arcs, or impose that rebates may not exceed monetary charges such as tolls or fares). Instead, we will work with the optimality conditions of this problem to explicitly characterize the Stackelberg equilibrium. This will allow us to design an efficient algorithm and to find the worst-case inefficiency of the corresponding equilibrium.

Not only do we want to compare the social cost of different solutions with rebates, but we also want to compare using rebates to not using them. Therefore, another measure of interest is the participants' real cost, represented by the objective function $C(x) := \sum_{i \in A} x_i c_i(x_i)$. The following definition captures the situation when the system owner controls the whole system.

DEFINITION 3.4. A flow x^{SO} is a system optimum if it is feasible and minimizes $C(\cdot)$.

The following proposition draws on the first-order optimality conditions to the mathematical program that defines a system optimum.

PROPOSITION 3.5 (BECKMANN, MCGUIRE, AND WINSTEN 1956). *For instances with differentiable and convex cost functions, a flow x^{SO} is a system optimum if and only if it is a Wardrop equilibrium with respect to the modified cost functions $c_i^*(x_i) := c_i(x_i) + x_i c'_i(x_i)$, where $c'_i(x)$ is the derivative of $c_i(x)$ with respect to x .*

Note that if $\rho \geq 1$, the social cost of a Stackelberg equilibrium (s, x^s) satisfies

$$C(x^{SO}) \leq C_\rho(s) \leq C(x^{WE}). \tag{2}$$

The lower bound follows from (1b) because its second term is nonnegative, and the upper bound comes from the feasibility of $s = 0$ because $C(x^{WE}) = C_\rho(0)$.

3.1. Examples

In this section we introduce two concrete instances that will be the running examples for the rest of the article. These instances will be used to illustrate the different concepts and calculations along the way.

INSTANCE 1 (ROUGHGARDEN AND TARDOS 2002). The first instance represents a competitive situation first described by Pigou (1920). As illustrated in Figure 1, participants must select one of two available modes: the first is expensive but its cost is not influenced by demand, while the second one is cheap under low demand but becomes expensive if it attracts many participants. This instance models a decision that commuters make daily in many cities. A person can use mass transit and experience an almost constant but large commute time, or can drive to (hopefully) experience a short commute while being exposed to the possibility of congestion.

The total demand in this instance is equal to 1, composed of an infinite number of price-taking users. The Wardrop equilibrium routes all flow in the lower arc because all participants take lowest-cost routes. Under this solution, $C(x^{WE}) = 1$. To exploit the effects of congestion, the system optimum assigns half of the participants to each mode, implying that $C(x^{SO}) = 3/4$.

If $\rho \leq 1$, the system owner will propose rebates equal to $(1, 1/2)$, which is the vector of prevailing costs under the system optimum. This results in an equilibrium that matches the system optimum. Actually, §4.1 shows that, for arbitrary instances, rebates lead to the system optimum when $\rho \leq 1$ because experienced costs are zero. Let us now consider the case $\rho > 1$. It does not make sense to offer a rebate in both arcs because subtracting a constant everywhere will not change the equilibrium. Therefore, the system owner should only consider giving a rebate in the upper arc (the lower one is always cheaper so it should not be subsidized). Denoting this rebate by $s \in [0, 1]$, the perceived cost on this arc equals $1 - s$. Therefore, the corresponding Wardrop equilibrium x^s is the flow that routes s units in the upper arc. After some algebra, $C_\rho(s) = 1 - s + \rho s^2$. The minimum, which provides the Stackelberg equilibrium, is $s = 1/(2\rho)$ and achieves a social cost of $1 - 1/(4\rho)$.

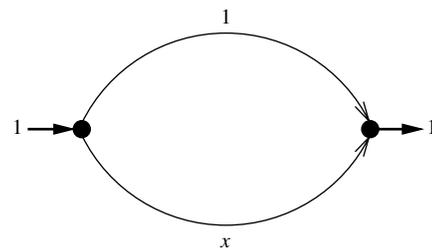


Figure 1 Pigou's Example
 Note. Arcs are labeled with their cost functions.

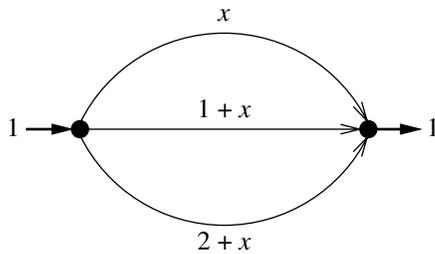


Figure 2 Running Example

Note. Arcs are labeled with their cost functions.

INSTANCE 2. The second network is similar to Pigou’s but contains an extra mode. As depicted in Figure 2, the three modes, numbered from 1 to 3 for simplicity, have cost functions equal to $c_i(x_i) := (i - 1) + x_i$. At the Wardrop equilibrium, all participants select the first mode, and therefore $C(x^{\text{WE}}) = 1$. The system optimum is given by the flow $(3/4, 1/4, 0)$, with total cost $C(x^{\text{SO}}) = 7/8$. Finally, an optimal rebate vector for $\rho > 1$ is $s = (0, 1/(2\rho), y)$, with $0 \leq y \leq 1 + 1/(4\rho)$. The corresponding Wardrop equilibrium x^s is $(1 - 1/(4\rho), 1/(4\rho), 0)$, and its social cost equals $C_\rho(s) = 1 - 1/(8\rho)$.

3.2. An Application to Logistics

The framework that we consider can readily be used to model competition in other settings such as telecommunication and distribution networks. This section briefly comments on an application in the area of logistics.

We consider a freight company that sends goods across a network. The system owner models the corporate headquarters while participants model business units that manage different markets. The system is not controlled centrally; units make their own decisions about how goods are transshipped across the network, considering their individual costs. This network is composed of resources, which may represent different carriers that transport freight or facilities that process it. Resources include sorting facilities, warehouses, flight legs, airports, ship routes, ports, canals, etc. Some of those resources belong to the unit, others belong to the company and are shared between units, and some are controlled by third parties. Resources that are not controlled directly by the unit will be priced according to the laws of offer and demand. Hence, competition for a resource will drive its price up, which can be represented by cost functions (in this case cost-demand curves). We assume that units are not big enough to influence prices independently (i.e., they are price-taking).

Units select a set of resources to transship their goods at minimum cost, and are rewarded by the

profits they generate.² Cost-demand curves create externalities between units, which is what causes competition among them. If nothing is done, the stable situation would be an equilibrium among the business units that is generally inefficient in terms of the company’s total profit. Realizing the problem, the company can compute the system optimum ignoring the goals of the individual business units, and find the rebates that it should offer for each resource. In this way, the headquarters will be offering incentives that help align business units into maximizing the company’s profits.

4. General Network Topologies

We start our study of the structural characteristics of Stackelberg equilibria. In this section, we consider general network topologies, with possibly several OD-pairs. We start by considering the case of the system owner assigning more value to the participants’ perceived cost than to its own rebate investment, and characterize the optimal strategy when setting the rebates. Later, we turn into the opposite case and provide some properties that will be used to characterize optimal rebates.

4.1. Small ρ

This section focuses on achieving a fully efficient coordinated solution for the case of $\rho \leq 1$ and networks with arbitrary topology. As suggested in §3.1, let us consider the rebate vector given by $s_i = c_i(x_i^{\text{SO}})$ for all $i \in A$. With those rebates, the system optimum x^{SO} is an equilibrium for the participants’ game because participants experience a cost equal to zero (which is the absolute minimum because of the nonnegativity of modified cost functions).

Beckmann, McGuire, and Winsten (1956) proved that payments equal to marginal costs at the system optimum also lead to a system optimum (see Proposition 3.5); recalling that $c_i^*(x_i) = c_i(x_i) + x_i c'_i(x_i)$, this corresponds to negative rebates $s_i = -x_i^{\text{SO}} c'_i(x_i^{\text{SO}})$. Moreover, any convex combination of optimal transfers payments (tolls or rebates) is also optimal (Bergendorff, Hearn, and Ramana 1997), which implies that the set of transfers payments that lead to system optimality is a polyhedron. We summarize these claims in the following remark.

² All of the results valid for arbitrary networks are also valid in the more general setting of *nonatomic congestion games* (Rosenthal 1973). In this case, business units will select one set of resources from a list of feasible sets, without insisting that these sets have to be paths. This more general competitive situation is called nonatomic because participants are price taking, and a congestion game because participants are anonymous and costs of resources depend only on the number of participants selecting them.

REMARK 4.1. When rebates equal

$$s = [(1 - \kappa)c_i(x^{\text{SO}}) - \kappa(x_i^{\text{SO}}c'_i(x_i^{\text{SO}}))]_{i \in A}$$

with $0 \leq \kappa \leq 1$, a system optimal solution x^{SO} is at equilibrium. Here, positive values of s_i represent rebates and negative values represent payments. Moreover, if cost functions are strictly increasing, the corresponding equilibrium x^s is unique.

The next proposition shows that the previously-mentioned rebates are optimal when cost functions are strictly increasing. It turns out that this is the only optimal vector and leads to a unique second-stage equilibrium, which matches the system optimum. If we only consider weakly increasing functions, then a system optimum is always at equilibrium for that rebate vector but there may be other equilibria. In that case, though, an optimal rebate vector may not exist.

PROPOSITION 4.2. *Assume that $\rho \leq 1$ and that cost functions are strictly increasing. For arbitrary networks, a Stackelberg equilibrium (s, x^s) satisfies that $s = (c_i(x_i^{\text{SO}}))_{i \in A}$ and $x^s = x^{\text{SO}}$. This equilibrium achieves a social cost of $C_\rho(s) = \rho C(x^{\text{SO}})$.*

PROOF. Considering s as described in the proposition, let us prove that x^s has to be equal to x^{SO} . The Stackelberg flow x^s is a Wardrop equilibrium under the modified cost functions. The variational inequality characterization of Wardrop equilibria (Smith 1979) and the choice of s implies that for a feasible flow x , $\sum_{i \in A} (x_i - x_i^s)[c_i(x_i^s) - c_i(x_i^{\text{SO}})]^+ \geq 0$. Because the optimal flow x^{SO} is feasible, we have $\sum_{i \in A} (x_i^{\text{SO}} - x_i^s) \cdot [c_i(x_i^s) - c_i(x_i^{\text{SO}})]^+ \geq 0$. The summands vanish on arcs i such that $x_i^s \leq x_i^{\text{SO}}$, and are strictly negative on arcs i for which $x_i^s > x_i^{\text{SO}}$. Consequently, $x_i^s \leq x_i^{\text{SO}}$ for all $i \in A$, resulting in $x^s = x^{\text{SO}}$ because x^{SO} is a feasible flow without cycles (because it minimizes the participants' real cost and link cost functions are strictly increasing). Evaluating the social cost, we compute that $C_\rho(s) = \rho C(x^{\text{SO}})$.

Let us now show that this choice of s provides the same social cost as an optimal rebate vector s^* . Using the nonnegativity of the first term of (1a) and the feasibility of s , respectively, $\rho \sum_{i \in A} x_i^{s^*} \min(c_i(x_i^{s^*}), s_i^*) \leq C_\rho(s^*) \leq C_\rho(s)$, from where $\sum_{i \in A} x_i^{s^*} \min(c_i(x_i^{s^*}), s_i^*) \leq C(x^{\text{SO}})$. Bounding each of the terms in (1b) separately, $C_\rho(s^*) \geq C(x^{\text{SO}}) + (\rho - 1)C(x^{\text{SO}}) = \rho C(x^{\text{SO}})$, where we used that x^{SO} minimizes $C(\cdot)$ and that $\rho \leq 1$. Hence, $\rho C(x^{\text{SO}})$ is a lower bound for the optimal social objective that is attained at s , which establishes the proposition. \square

4.2. Large ρ

In this section we consider that $\rho > 1$. For constant cost functions, it is optimal to offer no rebates. Indeed, when $s = 0$, the participants' real cost under a Nash equilibrium equals that of a system optimum and the

cost of rebates is zero. Because both terms of (1b) equal a lower bound, this choice of s is optimal for the leader. Rebates are useful only in the presence of congestion. (Note that we get to a similar conclusion in the model of Labbé, Marcotte, and Savard 1998, who assumed that there is no congestion and that the leader is a revenue-maximizer.)

We will characterize the benefits of offering rebates by studying the structure of Stackelberg equilibria. We start by proving that under an optimal rebate vector there is always at least one used arc with positive experienced cost, and one used arc in which no rebate is offered. We let $\mathcal{F} \subseteq A$ be the set of arcs with positive flow under the equilibrium, which we partition into sets \mathcal{F}_s , containing arcs with positive rebates, and \mathcal{F}_0 , containing arcs with no rebates.

DEFINITION 4.3. For a given rebate vector s , define $\mathcal{F} := \{i \in A \mid x_i^s > 0\} = \mathcal{F}_s \cup \mathcal{F}_0$, where $\mathcal{F}_s := \{i \in \mathcal{F} \mid s_i > 0\}$ and $\mathcal{F}_0 := \{i \in \mathcal{F} \mid s_i = 0\}$.

Without loss of generality, we will sometimes assume that rebates for arcs in $A \setminus \mathcal{F}$ are zero. Indeed, if an unused arc has a positive rebate, it will still be unused without the rebate. Consequently, the corresponding Wardrop equilibrium and all the aggregate measures we considered do not change when the rebate is removed. For example, for the Stackelberg equilibrium of Instance 2, we have that $\mathcal{F}_0 = \{1\}$ and $\mathcal{F}_s = \{2\}$. The third arc does not belong to \mathcal{F} because its flow is zero.

LEMMA 4.4. *Assume that $\rho > 1$ and that all cost functions are strictly increasing. For an arbitrary network, if (s, x^s) is a Stackelberg equilibrium, then there exists an arc $i \in \mathcal{F}$ such that $s_i < c_i(x_i^s)$.*

PROOF. Assume that all perceived costs are zero, i.e., $s_i \geq c_i(x_i^s)$ for all $i \in \mathcal{F}$. Without loss of generality, it is enough that $s_i = c_i(x_i^s)$ for all those arcs. Then the social cost equals $C_\rho(s) = \rho C(x^s) \geq \rho C(x^{\text{SO}})$. As stated in §4.1, the social cost $\rho C(x^{\text{SO}})$ can be attained with rebates $(c_i(x_i^{\text{SO}}))_{i \in A}$. Since s was assumed to be an optimal rebate vector, we must have that $C(x^s) = C(x^{\text{SO}})$, from where we see that x^s is a system optimum. Because of Proposition 3.5, x^s is at equilibrium with respect to modified costs $c_i(x_i^s) + x_i^s c'_i(x_i^s)$.

As perceived costs are zero and cost functions are strictly increasing, $s_i > 0$ for all $i \in \mathcal{F}$, or equivalently $\mathcal{F}_0 = \emptyset$. Hence, there exists a small enough $\epsilon > 0$ such that $\tilde{s} \geq 0$, where

$$\tilde{s}_i := \begin{cases} s_i - \epsilon(c_i(x_i^s) + x_i^s c'_i(x_i^s)) & i \in \mathcal{F}, \\ 0 & i \in A \setminus \mathcal{F}. \end{cases}$$

Under rebates \tilde{s} and flow x^s , the perceived cost on each arc is $[c_i(x_i^s) - \tilde{s}_i]^+ = \epsilon(c_i(x_i^s) + x_i^s c'_i(x_i^s))$ for $i \in \mathcal{F}$. Similarly, $[c_i(x_i^s) - \tilde{s}_i]^+ = c_i(x_i^s) \geq \epsilon c_i(x_i^s) = \epsilon(c_i(x_i^s) + x_i^s c'_i(x_i^s))$ for $i \in A \setminus \mathcal{F}$. The last two equations

imply that x^s is at equilibrium under rebates \tilde{s} , and the perceived cost on each used arc is strictly positive. Finally, x^s is the unique equilibrium under \tilde{s} since the potential function $F(x) := \sum_{i \in A} \int_0^{x_i} [c_i(z_i) - \tilde{s}_i]^+ dz_i$ is convex in general, strictly convex in a vicinity of x^s as the cost functions are strictly increasing, and achieves a minimum at x^s . (We refer the reader to Beckmann, McGuire, and Winsten 1956 for details on the characterization of Wardrop equilibria with this type of potential function.) Consequently,

$$\begin{aligned} C_\rho(\tilde{s}) &= \sum_{i \in A} x_i^s c_i(x_i^s) + (\rho - 1) x_i^s \tilde{s}_i \\ &= C_\rho(s) - \epsilon(\rho - 1) \sum_{i \in \mathcal{J}} x_i^s \underbrace{(c_i(x_i^s) + x_i^s c'_i(x_i^s))}_{>0} < C_\rho(s), \end{aligned}$$

which is a contradiction to the optimality of s . \square

When we presented the examples in §3.1, we mentioned that it cannot be optimal to offer rebates in all arcs. The next lemma generalizes this observation to any network topology. It shows that, if all arcs are used, then \mathcal{J}_0 is necessarily nonempty. In §5, we will further generalize this lemma to instances in which not all arcs are used, but under the restriction that the network has parallel links. Notice that in the case of a general network without the restriction that all arcs are used, we do not know if \mathcal{J}_0 could be empty. If such generalization were valid, Lemma 4.4 would not be necessary because it would be implied by this result. Indeed, because cost functions are nonnegative and strictly increasing, any arc in \mathcal{J}_0 would experience a positive cost because it is used and has no rebate.

LEMMA 4.5. *Assume that $\rho > 1$ and that all cost functions are strictly increasing. For an arbitrary network, if (s, x^s) is a Stackelberg equilibrium and all arcs are used, then there exists an arc $i \in \mathcal{J}$ such that $s_i = 0$.*

PROOF. With the purpose of deriving a contradiction, let us assume that s is an optimal vector of rebates such that $s_i > 0$ for all $i \in \mathcal{J}$. We will show that we can decrease the rebates while maintaining the same user equilibrium. Note that, unless the network only consists of parallel links, subtracting a constant from all rebates may change the user equilibrium because it would make longer paths more attractive to users. Instead, the proposed rebates are such that the resulting perceived cost on all links is a multiple of the original perceived costs. Let us therefore consider new rebates $\tilde{s} = [c_i(x_i^s) - \eta[c_i(x_i^s) - s_i]^+]_{i \in A}$, where

$$\eta := \min_{i \in \mathcal{J}} \frac{c_i(x_i^s)}{[c_i(x_i^s) - s_i]^+}.$$

The definition implies that $\tilde{s} \geq 0$ and Lemma 4.4 implies that $\eta < \infty$, so the new rebates are well-defined. The perceived cost for arc i under the new

rebates equals $[c_i(x_i^s) - \tilde{s}_i]^+ = \eta[c_i(x_i^s) - s_i]^+$, meaning that x^s is also at equilibrium under \tilde{s} . Furthermore, as $s > 0$, we have that $\eta > 1$ and $\tilde{s} \leq s$. Hence, looking at (1b), the participants' real cost is unchanged, whereas the cost of rebates strictly decreases because $\tilde{s}_i = 0$ for the argument i achieving the minimum. This contradicts the optimality of s . \square

5. Networks with Parallel Links

Equipped with the structural results of the previous section, we now embark in the design of an efficient algorithm for computing Stackelberg equilibria. The outline of the procedure described in this section is as follows. First, we will partition arcs into those in which rebates must be offered, those in which no rebates must be offered and those that are not used in an equilibrium. With this partition, we will compute the actual rebates for the corresponding arcs.

We focus on networks in which participants have to select exactly one out of many possible arcs. This primarily models the mode choice problem but one can also use it for other applications in which users choose among substitutes. The network topology that corresponds to this situation comprises two nodes joined by several parallel arcs (see Figure 3). Networks with parallel arcs extend the classic two-route network introduced by Pigou (1920). They have been widely used because of their relevance to practical applications—such as transportation, telecommunication, scheduling, and resource allocation problems—and because of their tractability (see, e.g., Korilis, Lazar, and Orda 1995; Koutsoupias and Papadimitriou 1999; Roughgarden 2004; Engel, Fischer, and Galetovic 2004; de Palma and Picard 2006; de Palma, Kilani, and Lindsey 2007; Wichienstein, Bell, and Yang 2007; Xiao, Yang, and Han 2007; Acemoglu and Ozdaglar 2007; Johari, Weintraub, and Van Roy 2009). Note that the restriction to simple topologies seems necessary if we hope to find the optimal rebates in polynomial time because Cole, Dodis, and Roughgarden (2006) proved that finding

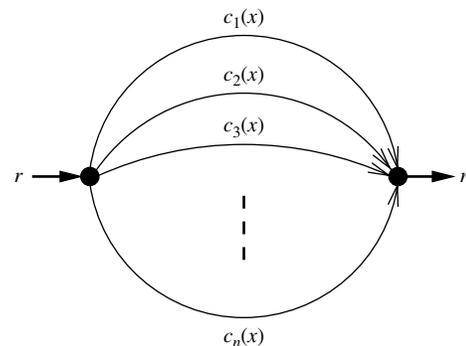


Figure 3 A Network with Parallel Arcs

optimal taxes in general networks with affine cost functions is hard.³ Finally, we only consider the case of $\rho > 1$, because the optimal rebates for $\rho \leq 1$ were already found in §4.1. We can assume without loss of generality that $s_i \leq c_i(x_i^s)$, as it is never beneficial to offer more.

Consider a Stackelberg equilibrium (s, x^s) of an instance in which cost functions are strictly increasing. The equilibrium conditions imply that there is a constant $L_\rho \geq 0$ such that

$$L_\rho = c_i(x_i^s) - s_i \quad \forall i \in \mathcal{F}, \quad (3a)$$

$$L_\rho \leq c_i(0) - s_i \quad \forall i \in A \setminus \mathcal{F}. \quad (3b)$$

Moreover, Lemma 4.4 implies that L_ρ has to be strictly positive. Hence, $c_i(x_i^s) > s_i$ for all $i \in A$. For networks with parallel arcs, then, we do not need to enforce the constraint that the system owner cannot offer rebates that are larger than the cost of arcs. In this case (1) simplifies to $C_\rho(s) = \sum_{i \in A} x_i^s [c_i(x_i^s) + (\rho - 1)s_i]$.

REMARK 5.1. The positivity of L_ρ also implies that when cost functions are strictly increasing there is a unique Wardrop equilibrium corresponding to the optimal s because the potential function $F(x) = \sum_{i \in A} \int_0^{x_i} [c_i(z) - s_i]^+ dz$ is strictly convex in a vicinity of x^s . Later, we shall prove that in this case the optimal s is also unique.

Going back to the examples of §3.1, it is not hard to check that L_ρ for Instances 1 and 2 equals $1 - 1/(2\rho)$ and $1 - 1/(4\rho)$, respectively.

5.1. General Cost Functions

We start with general cost functions and then, in the next section, switch to the particular case of affine cost functions. This section proves a result that will allow us to decide for which arcs we must offer positive rebates. To get there, we first have to present a series of lemmas. The first one establishes that a rebate vector that is optimal for a given network is also optimal when some unused arcs are taken out. In other words, removing $i \in A \setminus \mathcal{F}$ does not affect the optimality of s . Missing proofs are given in the appendix.

LEMMA 5.2. *Consider a network with parallel arcs and an optimal rebate vector s . If l is an arc in $A \setminus \mathcal{F}$, then the vector s with the entry corresponding to l removed is optimal for a similar instance with arc l removed.*

³ Cole, Dodis, and Roughgarden (2006, Theorem 6.2) prove that an approximation algorithm with guarantee better than $4/3 - \epsilon$ cannot exist unless $P = NP$. Although their reduction does not work for our problem, we conjecture that finding the optimal rebates in a general network with affine cost functions is also NP-hard because of the similarity between their social cost function and (1b) (see also §6.1). Another evidence in this direction is given by Labbé, Marcotte, and Savard (1998), who prove that computing taxes and rebates that maximize the leader's profit is an NP-hard problem, even when the network is not subject to congestion effects.

Notice that the previous lemma generalizes Lemma 4.5 to an arbitrary instance with parallel arcs. Indeed, Lemma 5.2 implies that an optimal rebate vector s is still optimal for the network consisting only of arcs in \mathcal{F} . Because that instance makes use of all arcs, it must contain at least one arc without rebate.

In the following propositions, we derive necessary conditions for a rebate vector s to be optimal. The next proposition shows that the optimal rebates satisfy the following equilibrium conditions: rebates are offered only in arcs for which the expression $c_i^*(\cdot)$ is minimal. This is implied by the first-order optimality conditions of the MPEC that characterizes the optimal rebates. Contrast this to Proposition 3.5 that states that in a system optimum, *participants* are assigned only to arcs for which the expression $c_i^*(\cdot)$ is minimal.

PROPOSITION 5.3. *Consider a network with parallel arcs and strictly increasing and differentiable cost functions, and let (s, x^s) be a Stackelberg equilibrium. There exists $V_\rho > 0$ such that*

$$V_\rho = c_i(x_i^s) + x_i^s c'_i(x_i^s) \quad \forall i \in \mathcal{F}_s, \quad (4a)$$

$$V_\rho \leq c_i(x_i^s) + x_i^s c'_i(x_i^s) \quad \forall i \in A \setminus \mathcal{F}_s. \quad (4b)$$

From (3a) and (4), we get that there exists a constant $D_\rho := 2L_\rho - V_\rho$ such that

$$D_\rho = c_i(x_i^s) - x_i^s c'_i(x_i^s) - 2s_i \quad \forall i \in \mathcal{F}_s, \quad (5a)$$

$$D_\rho \geq c_i(x_i^s) - x_i^s c'_i(x_i^s) - 2s_i \quad \forall i \in \mathcal{F}_0. \quad (5b)$$

The common perceived cost at equilibrium therefore equals $L_\rho = (V_\rho + D_\rho)/2$. Comparing the expressions, it is clear that $D_\rho < L_\rho < V_\rho$. For example, looking at a Stackelberg equilibrium of Instance 2, the constants are $V_\rho = 1 + 1/(2\rho)$ and $D_\rho = 1 - 1/\rho$.

In the sequel, we will make extensive use of the following definition to characterize and to compute optimal rebates:

DEFINITION 5.4. For $X \subseteq A$, let $K(X) := \sum_{i \in X} c'_i(x_i^s)^{-1}$. For the special case of an empty set, it is assumed that $K(\emptyset) := 0$.

The following technical lemma provides a formula that will be useful later. Its proof considers another feasible direction from the optimal rebate vector.

LEMMA 5.5. *Consider a network with parallel arcs and strictly increasing and differentiable cost functions. Letting (s, x^s) be a Stackelberg equilibrium, then*

$$\sum_{i \in \mathcal{F}_s} \left(x_i^s K(\mathcal{F}) + \frac{s_i}{c'_i(x_i^s)} K(\mathcal{F}_0) \right) = \frac{r}{\rho} K(\mathcal{F}_s). \quad (6)$$

Using the previous results, we can characterize the sets \mathcal{F}_0 and \mathcal{F}_s , which will allow us to compute the optimal rebates.

PROPOSITION 5.6. Consider a network with parallel arcs and strictly increasing and differentiable cost functions. Letting (s, x^s) be a Stackelberg equilibrium, for all $i \in A$ we have that

$$i \in \mathcal{F}_0 \Leftrightarrow D_\rho \geq c_i(x_i^s) - x_i^s c'_i(x_i^s), \quad (7a)$$

$$i \in A \setminus \mathcal{F} \Leftrightarrow V_\rho \leq c_i(0). \quad (7b)$$

PROOF. We start with (7a). The forward implication is (5b). Conversely, consider $i \in A$, and assume that $D_\rho \geq c_i(x_i^s) - x_i^s c'_i(x_i^s)$. If $i \in A \setminus \mathcal{F}$, then $x_i^s = 0$ and $c_i(0) \leq D_\rho < L_\rho$, contradicting the Wardrop equilibrium condition. If $i \in \mathcal{F}_s$, then (5a) implies that $c_i(x_i^s) - x_i^s c'_i(x_i^s) = D_\rho + 2s_i > D_\rho$, yielding a contradiction again.

The forward implication of (7b) follows from (4b). Conversely, consider an $i \in A$, and assume that $V_\rho \leq c_i(0)$. If $i \in \mathcal{F}_0$, then $c_i(0) < L_\rho < V_\rho$, which yields a contradiction. If $i \in \mathcal{F}_s$ then (4a) implies that $c_i(0) < c_i(x_i^s) + x_i^s c'_i(x_i^s) = V_\rho$, which is again a contradiction. \square

In other words, we have the following partition of the arcs according to the expression $c_i(x_i^s) - x_i^s c'_i(x_i^s)$: considering $i_0 \in \mathcal{F}_0$, $i_s \in \mathcal{F}_s$, and $j \in A \setminus \mathcal{F}$, we have

$$\begin{aligned} c_{i_0}(x_{i_0}^s) - x_{i_0}^s c'_{i_0}(x_{i_0}^s) &\leq D_\rho < c_{i_s}(x_{i_s}^s) - x_{i_s}^s c'_{i_s}(x_{i_s}^s) \\ &< V_\rho \leq c_j(0). \end{aligned} \quad (8)$$

This characterizes which arcs are used naturally because they are cheap, which arcs are used because of the rebates offered, and which arcs are not used, even having the possibility of offering rebates, because they are too expensive. Of course, to use this result constructively one would first need to know the Stackelberg equilibrium. In the next section, we will see how to work around that problem for affine cost functions. Going back to Instance 2, one can see that $c_1(x_1^s) - x_1^s c'_1(x_1^s) = 0 \leq D_\rho = 1 - 1/\rho < c_2(x_2^s) - x_2^s c'_2(x_2^s) = 1 < V_\rho = 1 + 1/(2\rho) \leq c_3(0) = 2$.

We can now use the, so far partial, characterization of Stackelberg equilibria to determine how many participants extract a benefit from the availability of rebates in the network.

PROPOSITION 5.7. Consider a network with parallel arcs and strictly increasing and differentiable cost functions, and let (s, x^s) be a Stackelberg equilibrium. The proportion of participants that receive a rebate is strictly lower than $1/\rho$.

PROOF. Assume that $\mathcal{F}_s \neq \emptyset$ because otherwise the claim is obvious. Dividing (6) by $K(\mathcal{F}_s)$,

$$\begin{aligned} \frac{r}{\rho} &= \frac{K(\mathcal{F}_0)}{K(\mathcal{F}_s)} \sum_{i \in \mathcal{F}_s} \left(x_i^s + \frac{s_i}{c'_i(x_i^s)} \right) + \sum_{i \in \mathcal{F}_s} x_i^s \\ &= K(\mathcal{F}_0)(V_\rho - L_\rho) + \sum_{i \in \mathcal{F}_s} x_i^s. \end{aligned}$$

Therefore, $\sum_{i \in \mathcal{F}_s} x_i^s / r = 1/\rho - K(\mathcal{F}_0)(V_\rho - L_\rho) / r < 1/\rho$, as we wanted to show. \square

As expected, there is a strong correlation between how many participants respond to the incentive and the gains in the social cost that arise from it. The previous bound turns out to be tight as demonstrated by the following instance.

INSTANCE 3. Consider a network similar to that depicted in Figure 1 but with cost functions $c_1(x) = 1 - (1 - \epsilon)/\rho + \alpha x$ and $c_2(x) = x$, where $0 < \epsilon < 1$ and $\alpha > 0$. Using results we will develop in §5.2, we must have that $\mathcal{F}_0 = \{2\}$ and $\mathcal{F}_s = \{1\}$ (because $b_2 < L_\infty(1 - 1/\rho) = 1 - 1/\rho < b_1 < L_\infty(1 + 1/\rho)$; see the next section for the notation). Hence, the rebate $s = (\epsilon/(2\rho), 0)$ is optimal and the corresponding equilibrium is given by

$$x^s = \left(\frac{2 - \epsilon}{2(1 + \alpha)\rho}, 1 - \frac{2 - \epsilon}{2(1 + \alpha)\rho} \right).$$

The proportion of participants that receive positive rebates is x_1^s , which tends to $1/\rho$ as ϵ and α tend to 0.

5.2. Affine Cost Functions

Having derived properties for general cost functions, this section considers instances with affine cost functions and explicitly provides expressions for the optimal rebates. Instances with this type of cost functions are rich enough for many congestion phenomena to appear. For example, the well-known Braess paradox was initially formulated with affine cost functions (Braess 1968). Even for applications in which cost functions are more complex, an affine approximation can already show evidence of first-order effects (Acemoglu and Ozdaglar 2007; Johari, Weintraub, and Van Roy 2009). We denote the cost function on arc $i \in A$ by $c_i(x) = a_i x + b_i$, with $a_i > 0$ and $b_i \geq 0$. Without loss of generality, we consider that arcs are sorted according to b_i , so we have that $b_1 \leq b_2 \leq \dots \leq b_{|A|}$. For ease of notation, we let $[i] := \{1, \dots, i\}$, and $b_{|A|+1} = +\infty$.

In the case of affine functions, we can simplify some of the formulas we provided in previous sections. For example, Definition 5.4 becomes $K(X) = \sum_{i \in X} 1/a_i$ for $X \subset A$. Notice also that a consequence of (5) is that $D_\rho \geq 0$ and $s_i \leq b_i/2$ for all $i \in \mathcal{F}_s$. Furthermore, (8) allows us to partition the arcs into the sets \mathcal{F}_0 , \mathcal{F}_s , and $A \setminus \mathcal{F}$ as follows.

PROPOSITION 5.8. Consider a network with parallel arcs and affine cost functions. If we consider $i_0 \in \mathcal{F}_0$, $i_s \in \mathcal{F}_s$, and $j \in A \setminus \mathcal{F}$, then $b_{i_0} \leq D_\rho < b_{i_s} < V_\rho \leq b_j$.

The following lemma and theorem show that if we know how the arcs are partitioned, we can compute the optimal rebate values for all arcs.

LEMMA 5.9. Consider a network with parallel arcs and affine cost functions, and let (s, x^s) be a Stackelberg equilibrium. If rebates are beneficial (i.e., if $\mathcal{F}_s \neq \emptyset$), then

$$D_\rho = \frac{1}{K(\mathcal{F}_0)} \left(r \frac{\rho - 1}{\rho} + \sum_{i \in \mathcal{F}_0} \frac{b_i}{a_i} \right),$$

$$V_\rho = \frac{1}{K(\mathcal{F})} \left(r \frac{\rho + 1}{\rho} + \sum_{i \in \mathcal{F}} \frac{b_i}{a_i} \right).$$

If \mathcal{F}_0 is known, making use of the previous lemma, we can compute the optimal rebates using the relations that we developed in the previous section. This result implies that, essentially, there is a unique optimal rebate vector.

THEOREM 5.10. Consider a network with parallel arcs and affine cost functions. Then, the optimal rebates must satisfy that

$$s_i = \left[\frac{b_i - D_\rho}{2} \right]^+$$

for all $i \in \mathcal{F}$. Moreover, if this formula is used for all arcs, the corresponding solution (s, x^s) is a Stackelberg equilibrium.

PROOF. Consider an arc $i \in \mathcal{F}$. If $i \in \mathcal{F}_0$, then $s_i = 0$ by definition and this agrees with the proposed formula because of Proposition 5.8. If $i \in \mathcal{F}_s$, then solving for s_i in (5a) also gives the proposed formula.

Now consider using the proposed formula for all $i \in A$. We must prove that each arc $j \in A \setminus \mathcal{F}$ is not used under the corresponding Wardrop equilibrium. Proposition 5.8 implies that $b_j > V_\rho$. Therefore, the rebate computed by the theorem is positive and $b_j - s_j = (b_j + D_\rho)/2$. We conclude that the experienced cost when the flow is zero equals $b_j - s_j \geq L_\rho$, which means that $x_j^s = 0$. \square

Evidently, plugging the values into the expression of the previous theorem for Examples 1 and 2 gives us the rebates that we indicated in §3.1. What remains to be done to finish the characterization of optimal rebates is to find \mathcal{F}_0 , which will allow us to determine the value of D_ρ . The following result provides a characterization of the common cost experienced by participants under a Stackelberg equilibrium. We will use it to compute the values of D_ρ and V_ρ .

PROPOSITION 5.11. Consider a network with parallel arcs, affine cost functions, and total demand $r > 0$. For $j \in A$, define $\gamma(j, r) := (r + \sum_{i=1}^j (b_i/a_i))/K([j])$. There exist unique arcs $i_0, i_1 \in A$ such that

$$b_{i_0} \leq \gamma(i_0, r) < b_{i_0+1}, \quad (9a)$$

$$b_{i_1} < \gamma(i_1, r) \leq b_{i_1+1}. \quad (9b)$$

Moreover, $\gamma(i_0, r) = \gamma(i_1, r) = L_\infty$, where L_∞ is the common cost experienced by participants under a Wardrop equilibrium (without rebates).

PROOF. Let us define $i_0 := \max\{i \in A: b_i \leq L_\infty\}$, and let x be the Wardrop equilibrium. From the definition, i_0 satisfies that $b_{i_0} \leq L_\infty < b_{i_0+1}$. The equilibrium condition implies that $x_i = (L_\infty - b_i)/a_i$ for all $i \leq i_0$. Summing over that range we get that $L_\infty = \gamma(i_0, r)$. What is left to prove is that there is no other i_0 that satisfies (9a). Hence, assume that there is another index \tilde{i}_0 , and define \tilde{x} equal to $(\gamma(\tilde{i}_0, r) - b_i)/a_i$ for $i \leq \tilde{i}_0$ and 0 otherwise. This flow is feasible because it is nonnegative and its total demand equals r . Furthermore, it satisfies the Wardrop equilibrium conditions with cost equal to $\gamma(\tilde{i}_0, r)$ for all participants. Recall that because cost functions are strictly increasing, there exists a unique Wardrop equilibrium. Because x and \tilde{x} are both at equilibrium, they must be equal. This implies that $\gamma(i_0, r) = \gamma(\tilde{i}_0, r)$, from where $\tilde{i}_0 = i_0$ because of (9a).

A similar argument proves the existence of a unique index $i_1 := \max\{i \in A: b_i < L_\infty\}$ that satisfies (9b). We highlight that i_0 and i_1 differ only when there is a link i with $b_i = L_\infty$, in which case $i_0 > i_1$. \square

Computing $\gamma(i, r)$ for the different arcs in Instance 2, we get that $\gamma(1, r) = r$, $\gamma(2, r) = (r + 1)/2$, and $\gamma(3, r) = (r + 3)/3$. Then $i_0 = 1$ when $0 \leq r < 1$, $i_0 = 2$ when $1 \leq r < 3$, and $i_0 = 3$ when $r \geq 3$. Similarly, $i_1 = 1$ when $0 \leq r \leq 1$, $i_1 = 2$ when $1 < r \leq 3$, and $i_1 = 3$ when $r > 3$.

In the sequel, we consider the function $L_\infty(z)$ which represents the perceived cost under a Wardrop equilibrium (without rebates) when the total demand is z . When we do not denote a demand explicitly, we assume that the regular demand of r is used. It is well known that the function $L_\infty(z)$ is nondecreasing and continuous (Hall 1978). In addition, Proposition 5.11 implies that it is piecewise linear with slope $1/K([i])$ when its value is between b_i and b_{i+1} . Therefore, it is a concave function. For an illustration, see Figure 4 in the following section. Under our assumptions, $L_\infty(\cdot)$ is easy to compute using an incremental loading algorithm.

Using the previous result, we can now express the perceived cost of participants at the Stackelberg equilibrium. In addition, the next proposition will clearly identify the sets \mathcal{F}_0 and \mathcal{F}_s . First, $\mathcal{F}_0 = [i_0]$, where i_0 corresponds to the index introduced in Proposition 5.11 for a demand of $r(1 - 1/\rho)$. The arcs without rebates that are used in a Stackelberg equilibrium coincide with those that are used under a Wardrop equilibrium (without rebates) with a total demand of $r(1 - 1/\rho) + \epsilon$, for a sufficiently small $\epsilon > 0$. Likewise, $\mathcal{F}_s = [i_1]$, where i_1 is the index introduced in Proposition 5.11 for a demand of $r(1 + 1/\rho)$. The arcs used under a Stackelberg equilibrium coincide with those that are used under a Wardrop equilibrium with a total demand of $r(1 + 1/\rho)$.

PROPOSITION 5.12. Consider a network with parallel arcs and affine cost functions, and a Stackelberg equilibrium (s, x^s) . If rebates are beneficial (i.e., if $\mathcal{F}_s \neq \emptyset$), then

$$D_\rho = L_\infty\left(r \frac{\rho-1}{\rho}\right) \quad \text{and} \quad V_\rho = L_\infty\left(r \frac{\rho+1}{\rho}\right),$$

and the perceived cost of each participant under x^s is

$$L_\rho = \frac{1}{2} \left(L_\infty\left(r \frac{\rho-1}{\rho}\right) + L_\infty\left(r \frac{\rho+1}{\rho}\right) \right).$$

PROOF. From Proposition 5.8 and Lemma 5.9, we know that there exist $i_0, i_1 \in A$ such that $\mathcal{F}_0 = [i_0]$, $\mathcal{F} = [i_1]$,

$$b_{i_0} \leq \gamma\left(i_0, r \frac{\rho-1}{\rho}\right) < b_{i_0+1}, \quad \text{and}$$

$$b_{i_1} < \gamma\left(i_1, r \frac{\rho+1}{\rho}\right) \leq b_{i_1+1}.$$

Hence, Proposition 5.11 implies the first two claims. The third follows simply from the relation displayed right after (5). \square

Using the values of i_0 that we previously computed for Instance 2, it is easy to see that $L_\infty(r) = r$ when $0 \leq r < 1$, $L_\infty(r) = (r+1)/2$ when $1 \leq r < 3$, and $L_\infty(r) = (r+3)/3$ when $r \geq 3$. Using this, $D_\rho = 1 - 1/\rho$, $V_\rho = 1 + 1/(2\rho)$, and $L_\rho = 1 - 1/(4\rho)$ as expected.

Notice that Proposition 5.12 provides an explicit way to compute D_ρ . Hence, this value is unique and, relying on Proposition 5.11, the vector of optimal rebates is unique as well (disregarding that a rebate for an arc $l \in A \setminus \mathcal{F}$ can take any value between 0 and $c_l(0) - L_\rho$, which does not count as multiple equilibria because l is unused). Because there is a unique Wardrop equilibrium for any given rebate vector such that $L_\rho > 0$, the Stackelberg game has an essentially unique solution. This means that any two different Stackelberg equilibria will be undistinguishable from a practical point of view because flows and costs under both solutions will be equal.

The following proposition provides an easily verifiable condition to check whether rebates can help lower the social cost in a specific instance. Note that when the inequality does not hold, the formula must hold with equality because of the concavity of $L_\infty(r)$.

PROPOSITION 5.13. Consider a network with parallel arcs and affine cost functions. Rebates are beneficial (i.e., $\mathcal{F}_s \neq \emptyset$) if and only if

$$\frac{1}{2} \left(L_\infty\left(r \frac{\rho-1}{\rho}\right) + L_\infty\left(r \frac{\rho+1}{\rho}\right) \right) < L_\infty.$$

5.3. A Polynomial-Time Algorithm for Computing Optimal Rebates

The results we have presented in the previous section lead to a polynomial-time algorithm for finding the optimal rebates. The following algorithm receives an instance described by a network with parallel arcs, affine cost functions, and a fixed demand as input, and computes a Stackelberg equilibrium.

Algorithm

- (1) Sort the arcs with respect to b_i to cast the instance into the form we considered.
- (2) Compute the function $L_\infty(z)$ for the instance.
- (3) Use Proposition 5.13 to decide whether rebates need to be used.
- (4) If rebates are not beneficial, we are done.
- (5) Compute D_ρ using Proposition 5.12.
- (6) Finally, compute the rebate to offer in each arch using Theorem 5.10.

Each of these steps requires a computation that can be done in polynomial time. The bottleneck is computing $L_\infty(z)$, which requires solving at most $|A|$ systems of linear equations to load the network incrementally and compute the breakpoints of the piecewise linear function.

At this point, it is convenient to discuss how to estimate the information needed to create an instance in practice. This estimation has been discussed at length in the literature of transportation engineering (see, e.g., Sheffi 1985). We provide a short overview. First, one needs to list the modes and their costs as a function on the flows. Cost functions are calibrated from historical information, taking into account how different modes operate. Overall, one needs to sum the travel time and the fare or toll for the mode, which can be converted to the same units by using the average value of time for the population. The latter can usually be estimated from socioeconomic information coming from census data. The demand can be measured directly or may come from historical OD matrices that can be calibrated using up-to-date traffic counts. The most difficult parameter to estimate in our model is ρ because it is hard to attach a dollar figure to a reduction in the total cost experienced by travelers. This estimation has been attempted by the Partnership for New York City (2006), who measure the economic impact of reducing traffic congestion. Alternatively, one can compute the optimal rebates, social costs, and total cost experienced by commuters, as a function of ρ . This can be done easily because the algorithm above runs fast enough to solve the problem for many different values of ρ . With this curve in hand, one can look at the tradeoff between the budget invested in rebates and the overall social benefit. This can guide policymakers in selecting the optimal rebates to be used in a concrete situation.

6. The Benefits of Using Rebates

6.1. Coordination Mechanisms Based on Transfer Payments

First, we introduce some measures derived from the price of anarchy that will be useful to quantify the quality of equilibria resulting from a coordination mechanism. As we said in §2, Roughgarden and Tardos (2002) were the first to measure the price of anarchy in the network competition model introduced by Wardrop (1952). They defined the coordination ratio of an instance as

$$\frac{C(x^{\text{WE}})}{C(x^{\text{SO}})}, \quad (10)$$

and the price of anarchy as the supremum of (10) among all Wardrop equilibria and all possible instances (meaning all possible networks, demands, and allowed cost functions). Note that this value is at least 1 and it can be interpreted as follows: if it is low, then there is not much improvement to be expected from the introduction of a coordination mechanism in the game that was considered. On the other hand, a large price of anarchy suggests that there is a potentially large benefit to be made. For example, the coordination ratio of Pigou's instance (Instance 1) is 4/3. The following result establishes that this ratio is the largest possible.

PROPOSITION 6.1 (ROUGHGARDEN AND TARDOS 2002). *The price of anarchy for instances with affine cost functions is 4/3.*

For quadratic, cubic, and quartic cost functions, the price of anarchy is 1.626, 1.896, and 2.151, respectively (Roughgarden 2003; Correa, Schulz, and Stier-Moses 2004). For a simple proof of these results we direct the reader to Correa, Schulz, and Stier-Moses (2008).

Traditionally, the efficiency of a solution involving congestion pricing has been defined in terms of the total cost $C(\cdot)$ because charges are transfer payments that stay inside the system, or alternatively by assuming that these payments can be redistributed back to the users. For that social cost function, as Proposition 3.5 shows, charging users the externalities they introduce produces a socially efficient outcome. Some more recent articles look at social cost functions that include a term corresponding to taxation, similar to what we do in (1a). Under these more general social cost functions, a system owner may take a more holistic view, and not only care about outcomes, but also about investments. Cole, Dodis, and Roughgarden (2006) considered the problem of finding the taxes τ that minimize $\sum_{i \in A} x_i(c_i(x_i) + \tau_i)$, where x is a Wardrop equilibrium with respect to modified cost functions $c(\cdot) + \tau$. Unfortunately, finding an optimal mechanism for this

social cost function is NP-hard for arbitrary instances. Although they did not explicitly specialize their results to networks with parallel arcs, a generalization of the results of §5 can be used to compute optimal payments in polynomial time (still considering a general conversion factor ρ like in (1a)). Karakostas and Kolliopoulos (2005) extended the previous analysis and found bounds for the social cost achieved by an extension of the marginal taxation mechanism to heterogeneous values-of-time. Under this setting, the ratio of the social cost of an equilibrium to the solution of minimum social cost with respect to the optimal taxes is bounded with a smaller constant than that of Proposition 6.1 and its generalizations. In addition, the social cost is not too large compared to the minimum possible total cost (without taxes).

One can use different variations of the concept of the price of anarchy to quantify the power of a coordination mechanism. We consider the two definitions that are most interesting in our opinion. Both consist of ratios of the same cost function under two different solutions, thereby not falling into the situation of comparing apples and oranges. In addition, both compare the outcome provided by the coordination mechanism to an upper or lower bound, depending on the circumstances.

The first measure we consider is a straightforward extension of (10). Indeed, to quantify the loss of efficiency because of the limited coordinating power of the system owner, we consider the ratio

$$\frac{C(x^s)}{C(x^{\text{SO}})}. \quad (11)$$

For example, looking at Instance 2, this ratio equals 1 for $\rho \leq 1$ and $(8 - 2/\rho + 1/\rho^2)/7$ for $\rho > 1$. Note that although the previous ratio measures the quality of a given solution (s, x^s) for a fixed instance, our main interest is on the supremum of the coordination ratio of an arbitrary Stackelberg equilibrium over all possible instances, as it is done for the price of anarchy. Another option would have been to define the price of anarchy as in (11) but using perceived costs, as Cole, Dodis, and Roughgarden (2006) proposed for their study of taxes in networks. Remark 6.5 shows that the bound that can be obtained is the same as that for (11).

Previous research has determined that the price of anarchy is sometimes a pessimistic measure, as can be expected from a general worst-case bound. For example, Correa, Schulz, and Stier-Moses (2008) proposed to restrict the analysis to instances with fixed congestion loads to get more realistic estimates. Another aspect of the previous definitions is that they do not consider that in certain settings a system optimum is unrealistic and cannot be implemented. For example, Schulz and Stier-Moses (2006) proposed

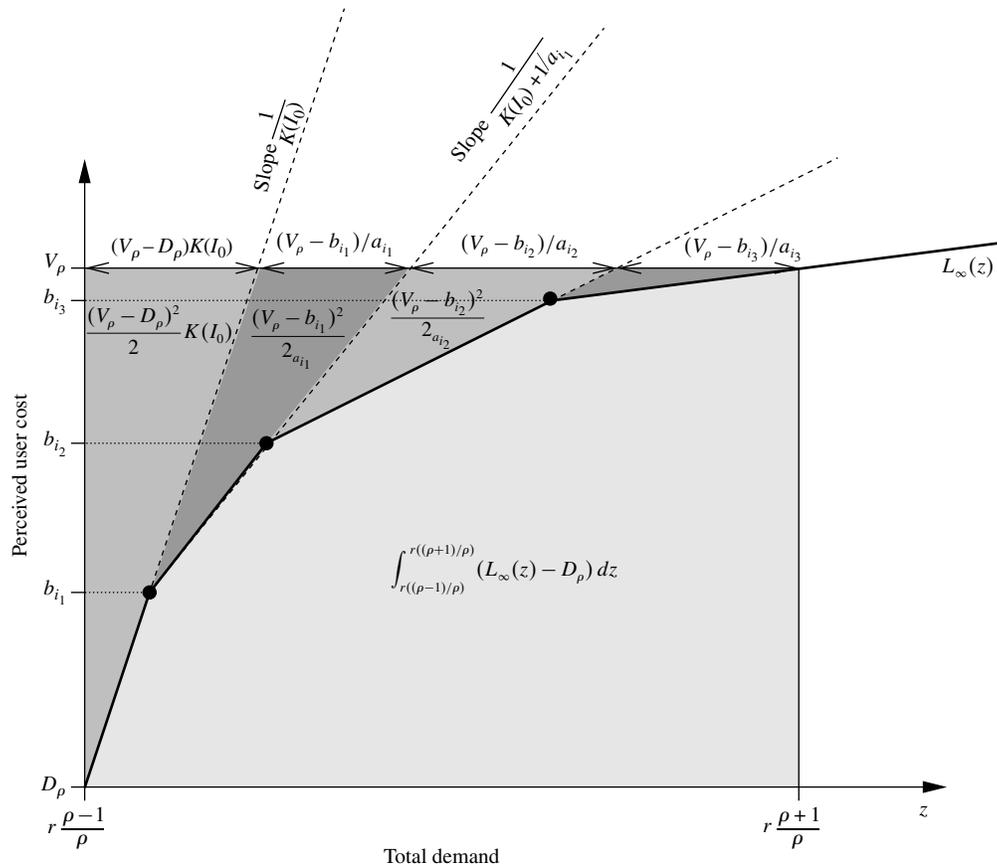


Figure 4 Illustration of the Decomposition of $2r(V_\rho - D_\rho)/\rho$

to quantify the performance of a route guidance system for vehicular traffic by comparing the solutions with and without guidance instead of using a social optimum.

To get a measure that is both less pessimistic and more realistic, we consider that the best possible outcome is what the system owner can enforce by setting rebates correctly. Hence, we consider the ratio of the social cost of a Wardrop equilibrium to that of a Stackelberg equilibrium. Letting s be the optimal rebate vector, this ratio is expressed as

$$\frac{C_\rho(0)}{C_\rho(s)} = \frac{C(x^{\text{WE}})}{C_\rho(s)}. \quad (12)$$

When $\rho < 1$, this quantity may be large because the denominator of (12) can be arbitrary small. Instead, when $\rho \geq 1$, the lower bound in (2) implies that this ratio is less pessimistic (smaller) than (10). For the examples provided before, we get that this ratio equals $4\rho/(4\rho - 1)$ for Instance 1, while the coordination ratio displayed in (10) equals $4/3$. The corresponding values for Instance 2 are $8\rho/(8\rho - 1)$ and $8/7$, respectively.

6.2. Computing the Price of Anarchy

Now that we have already characterized the optimal rebates for a particular instance of the problem, we are ready to analyze the performance of this coordination mechanism. We continue to work with networks consisting of parallel arcs and affine cost functions.

We start by providing a bound between the uncoordinated solution (no rebates) and the Stackelberg equilibrium. The case of $\rho \leq 1$ follows from Proposition 4.2. Indeed, using Proposition 6.1, we have that $C_\rho(0)/C_\rho(s) = C(x^{\text{WE}})/(\rho C(x^{\text{SO}})) \leq 4/(3\rho)$. This means that the price of anarchy arising from (12) is $4/(3\rho)$ for an arbitrary network with affine cost functions. The case of $\rho > 1$ is more involved. We start by computing the social cost of the Stackelberg equilibrium making use of the relations developed in the previous section.

LEMMA 6.2. Consider a network with parallel arcs and affine cost functions. For $\rho > 1$, the optimal social cost equals $(\rho/2) \int_{r(1-1/\rho)}^{r(1+1/\rho)} L_\infty(z) dz$.

PROOF. We rewrite the expression $2r(V_\rho - D_\rho)/\rho$ using the graphical decomposition shown in Figure 4.

Indeed, the area of the rectangle equals

$$\begin{aligned} & K(\mathcal{J}_0) \frac{(V_\rho - D_\rho)^2}{2} + \sum_{i \in \mathcal{J}_s} \frac{(V_\rho - b_i)^2}{2a_i} + \int_{r(1-1/\rho)}^{r(1+1/\rho)} (L_\infty(z) - D_\rho) dz \\ &= \frac{V_\rho - D_\rho}{2} \underbrace{\left((V_\rho - D_\rho)K(\mathcal{J}_0) + \sum_{i \in \mathcal{J}_s} \frac{V_\rho - b_i}{a_i} \right)}_{=2r/\rho} \\ & \quad - 2 \sum_{i \in \mathcal{J}_s} x_i^s s_i + \int_{r(1-1/\rho)}^{r(1+1/\rho)} (L_\infty(z) - D_\rho) dz, \end{aligned}$$

where we used the expression for x_i^s in the proof of Lemma 5.9, the expression for s_i in Theorem 5.10, and that $(V_\rho - b_i)^2 = (V_\rho - b_i)(V_\rho - D_\rho + D_\rho - b_i)$. The term with the brace equals $2r/\rho$ because of (17). After some algebra,

$$\begin{aligned} \sum_{i \in \mathcal{J}_s} x_i^s s_i &= \frac{1}{2} \int_{r(1-1/\rho)}^{r(1+1/\rho)} (L_\infty(z) - D_\rho) dz - \frac{r}{2\rho} (V_\rho - D_\rho) \\ &= \frac{1}{2} \int_{r(1-1/\rho)}^{r(1+1/\rho)} L_\infty(z) dz - \frac{r}{\rho} L_\rho. \end{aligned} \tag{13}$$

Consequently, the optimal social cost is

$$\begin{aligned} C_\rho(s) &= rL_\rho + \rho \sum_{i \in \mathcal{J}_s} x_i^s s_i \\ &= (\rho/2) \int_{r(1-1/\rho)}^{r(1+1/\rho)} L_\infty(z) dz. \quad \square \end{aligned} \tag{14}$$

THEOREM 6.3. Consider a network with parallel arcs and affine cost functions. For $\rho > 1$, the unique Stackelberg equilibrium (s, x^s) satisfies that

$$\frac{C_\rho(0)}{C_\rho(s)} \leq \frac{4\rho}{4\rho - 1}.$$

PROOF. Let us assume that $\mathcal{J}_s \neq \emptyset$ because otherwise the result is trivial. We need to compare the cost $C_\rho(s)$ computed in the previous lemma to $rL_\infty(r)$. Since $L_\infty(z)$ is a positive and concave function, $L_\infty(z)/z$ is a nonincreasing function. Bounding the integral from below as Figure 5 illustrates, we get that

$$C_\rho(s) \geq \frac{\rho r L_\infty(r)}{2\rho} \left(2 - \frac{1}{2\rho} \right) = C_\rho(0) \left(1 - \frac{1}{4\rho} \right),$$

as claimed. \square

The previous result characterizes the tradeoff between willingness to offer rebates and coordination power of the mechanism. The corresponding bound is tight, as Instance 1 demonstrates. (Note that the topmost arc has a constant cost, but one can take that cost equal to $ax + 1$ for an arbitrarily small a and nothing changes.) When the system owner’s willingness to offer rebates is high (ρ is not much larger than 1),

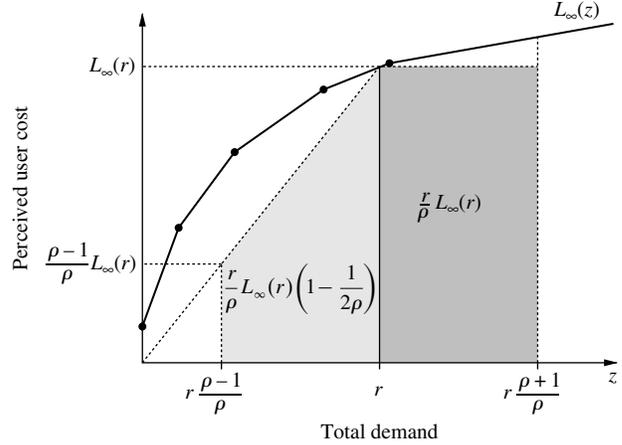


Figure 5 Illustration of the Bound for $\int_{r(1-1/\rho)}^{r(1+1/\rho)} L_\infty(z) dz$

the optimal social cost is approximately equal to the total cost under a system optimum; hence, the previous theorem provides a bound that is close to $4/3$. Here, recall that $4/3$ is the price of anarchy when the coordination mechanism can achieve a socially optimal solution (Proposition 6.1). Not surprisingly, when the willingness to offer rebates decreases (big ρ), the previous theorem gives a bound that is close to 1 because the system owner cannot do much better than in a Wardrop equilibrium.

Finally, we compute the worst-case ratio between the participants’ real cost under a Stackelberg equilibrium and under a system optimum, as we proposed in (11). In the case of $\rho \leq 1$, the flow x^{SO} is at equilibrium (and it is the unique one for strictly increasing cost functions, see §4.1), which implies that for an arbitrary network with affine cost functions the mechanism coordinates the network. The following results provide the bound corresponding to the case of $\rho > 1$.

THEOREM 6.4. Consider a network with parallel arcs and affine cost functions. The Stackelberg equilibrium (s, x^s) described in the previous section satisfies that

$$\frac{C(x^s)}{C(x^{SO})} \leq \frac{4\rho}{3\rho + 1}. \tag{15}$$

This bound is close to 1 for $\rho \approx 1$ because in that case a Stackelberg equilibrium is similar to a system optimum, and close to $4/3$ when ρ is large because in that case it is similar to a Wardrop equilibrium. As for the previous bound, Theorem 6.4 provides the curve that characterizes the tradeoff between willingness to offer rebates and coordinating power. We highlight that this bound is tight, which can be observed by taking $\epsilon = 0$ and letting α tend to 0 in Instance 3.

REMARK 6.5. The bound provided by Theorem 6.4 is also valid if one takes the ratio of the participants’ perceived cost in the Stackelberg equilibrium to that in the system optimum. This holds

because $\sum_i x_i^s [c_i(x_i^s) - s_i]^+ \leq C(x^s)$. Moreover, the same instance as before shows that this bound is tight.

7. Conclusions

We have studied the possible improvement that can stem from the use of rebates to coordinate an urban transportation network. If a system owner can afford to offer rebates and the system is highly congestible, rebates can significantly lower the social cost, which includes commute times and costs, as well as the cost of providing the rebates themselves. The algorithm we have presented can be used to determine optimal subsidies for each mode of transportation. Subsidies only affect a limited proportion of the demand, implying that the cost of providing them will not be exceedingly large. We have also estimated how much improvement this coordination mechanism brings to the system, as a function of the city's sensitivity to the cost of offering rebates. The coordinating power of a rebate scheme increases as the owner's sensitivity to the rebate cost decreases.

Several questions related to this study remain open. First and foremost, we have worked under the assumption that an instance has parallel arcs and affine cost functions. It would be interesting to generalize our results to more general instances. Another interesting problem is to determine the computational complexity of finding optimal rebates. Proving its hardness would shed light into this problem and would motivate the need to look for good heuristics. For quadratic cost functions for example, optimal rebates can be irrational numbers.⁴ Hence, an optimal rebate vector cannot be computed exactly in polynomial time. Nevertheless, it would be interesting to find a way to approximate it. Finally, another interesting open question is whether optimal rebates are unique in general. We have shown that this is true for networks with parallel arcs and affine cost functions.

Our model has some limitations that we would like to address in future research. On the one hand, we plan to incorporate the possibility that the system owner considers congestion pricing and rebates at the same time. Such extension will be useful to model systems in which both incentive mechanisms co-exist to create a larger differential between the total cost of driving and that of public transportation. On the other hand, we also want to look at an heterogeneous population because the valuation of time is user-dependent. This extension would allow a modeler to look at more precise measures of equity among commuters. Furthermore, it is important to consider

elastic demands because in practice some trips are optional and will not happen if the price of transportation is too high. The last element that would be interesting to consider is a situation in which multiple agencies in the government have to coordinate their efforts and budgets to offer incentives to the population. Because each agency has its own goal and agenda, they may not agree in the policy that should be chosen.

Acknowledgments

The authors thank two anonymous referees for their constructive remarks that improved the presentation of this paper, Tom Liebling for stimulating discussions about the motivations of this paper, and participants of the Mittagsemnar in the Operations Research Center at MIT for their insightful comments. This work has been supported by the Center for International Business Education and Research (CIBER), and by the Center for Excellence in E-Business (CEBIZ), both at Columbia University. The research was done while the first author was visiting Columbia Business School.

Appendix A. Proofs

A.1. Proof of Lemma 5.2

PROOF. Let $\tilde{A} := A \setminus \{l\}$ and \tilde{s} be the restriction of s to \tilde{A} . Assume that \tilde{s} is not optimal for \tilde{A} , and let \tilde{s}^* be an optimal rebate vector for that network. Then, $C_{\rho}^{\tilde{A}}(\tilde{s}^*) < C_{\rho}^{\tilde{A}}(\tilde{s}) = C_{\rho}^A(s)$, where the superscript represents the instance and the equality holds because if no participant selects the arc, it makes no difference whether the arc exists. Now, we take the optimal rebate vector \tilde{s}^* and extend it to the original network by setting $s_l^* := 0$ and $s_i^* := \tilde{s}_i^*$ for $i \in \tilde{A}$. Because a situation like Braess' paradox (1968) cannot occur in networks with parallel arcs, the participants' real cost at a Wardrop equilibrium decreases when link l is (re)introduced. Together with the fact that arc l is not subsidized, we have that $C_{\rho}^A(s^*) \leq C_{\rho}^A(\tilde{s}^*)$, which contradicts the optimality of s in the original instance. \square

A.2. Proof of Proposition 5.3

PROOF. Without loss of generality assume that $s_i = c_i(0) - L_{\rho}$ for all $i \in A \setminus \mathcal{F}$. Consider two fixed arcs $i \in \mathcal{F}_s$ and $j \in A$. Since s_i is strictly positive, it is possible to simultaneously reduce s_i by a positive infinitesimal ds_i and increase s_j so that the only effect is that some participants switch from arc i to j . In other words, we have that $dx_j = -dx_i$, where we denote an infinitesimal variation of a quantity w by dw . By design, the perceived cost L_{ρ} at equilibrium remains the same. The local effect at the arcs in question is $d(c_i(x_i^s) - s_i) = 0$ and $d(c_j(x_j^s) - s_j) = 0$. Because s was optimal, this modification cannot decrease the total rebate cost $\sum_{i \in A} x_i^s s_i$, as it does not modify the total participants' perceived cost. This implies that $d(x_i^s s_i + x_j^s s_j) \geq 0$. Putting all together,

$$dx_i^s (x_i^s c_i'(x_i^s) + s_i - x_j^s c_j'(x_j^s) - s_j) \geq 0.$$

As $dx_i^s < 0$, we must have $x_i^s c_i'(x_i^s) + s_i \leq x_j^s c_j'(x_j^s) + s_j$, and adding $c_i(x_i^s) - s_i = c_j(x_j^s) - s_j$, we finally obtain that $c_i(x_i^s) + x_i^s c_i'(x_i^s) \leq c_j(x_j^s) + x_j^s c_j'(x_j^s)$. We get the claim by letting i and j vary. \square

⁴ For example, considering the instance shown in Figure 1 with costs functions 1 and x^2 , and $\rho = 2$, it is optimal to offer a rebate of $(11 - \sqrt{13})/18$ for the arc with constant cost.

A.3. Proof of Lemma 5.5

PROOF. To ensure that the modification to the rebates we are going to make does not change the sets \mathcal{F}_s and \mathcal{F}_0 , we first remove all unused arcs. Indeed, Lemma 5.2 proves that if s is optimal for the original network, it is also optimal for the instance containing the arcs in \mathcal{F} only. The proposition is obvious for $\mathcal{F}_s = \emptyset$, so let us assume the opposite. We consider adding or subtracting a common infinitesimal ds to all rebates that are strictly positive. After modifying s the outcome is still at equilibrium and all arcs are still used; hence, differentials of perceived costs are equal for all arcs in \mathcal{F} . For a fixed $i_s \in \mathcal{F}_s$ and a fixed $i_0 \in \mathcal{F}_0 \neq \emptyset$, we have that $c'_i(x_i^s)dx_{i_s} - ds = c'_{i_0}(x_{i_0}^s)dx_{i_0}^s$ and

$$dx_i^s = \begin{cases} c'_i(x_i^s)dx_{i_s}^s / c'_i(x_i^s) & i \in \mathcal{F}_s, \\ c'_{i_0}(x_{i_0}^s)dx_{i_0}^s / c'_i(x_i^s) & i \in \mathcal{F}_0. \end{cases}$$

As the total demand does not change, we must have that $0 = \sum_{i \in \mathcal{F}} dx_i^s = K(\mathcal{F}_s)c'_i(x_i^s)dx_{i_s}^s + K(\mathcal{F}_0)c'_{i_0}(x_{i_0}^s)dx_{i_0}^s$. After some algebra, $c'_i(x_i^s)dx_{i_s}^s = dsK(\mathcal{F}_0)/K(\mathcal{F})$. Finally, let us consider how the social cost changes.

$$\begin{aligned} dC_\rho(s) &= d\left(r(c_{i_s}(x_{i_s}) - s_{i_s}) + \rho \sum_{i \in \mathcal{F}} x_i s_i\right) \\ &= r(c'_{i_s}(x_{i_s}^s)dx_{i_s}^s - ds) + \rho \sum_{i \in \mathcal{F}_s} \left(c'_{i_s}(x_{i_s}^s)dx_{i_s}^s \frac{s_i}{c'_i(x_i^s)} + x_i^s ds\right) \\ &= c'_{i_s}(x_{i_s}^s)dx_{i_s}^s \left(r + \rho \sum_{i \in \mathcal{F}_s} \frac{s_i}{c'_i(x_i^s)}\right) + ds \left(\rho \sum_{i \in \mathcal{F}_s} x_i^s - r\right) \\ &= ds \left(\left(\frac{K(\mathcal{F}_0)}{K(\mathcal{F})} - 1\right)r + \rho \frac{K(\mathcal{F}_0)}{K(\mathcal{F})} \sum_{i \in \mathcal{F}_s} \frac{s_i}{c'_i(x_i^s)} + \rho \sum_{i \in \mathcal{F}_s} x_i^s\right) \\ &= ds \frac{\rho}{K(\mathcal{F})} \left(-\frac{r}{\rho}K(\mathcal{F}_s) + K(\mathcal{F}_0) \sum_{i \in \mathcal{F}_s} \frac{s_i}{c'_i(x_i^s)} + K(\mathcal{F}) \sum_{i \in \mathcal{F}_s} x_i^s\right). \end{aligned}$$

The claim follows because the optimality of s implies that $dC_\rho(s) \geq 0$ for feasible directions $ds > 0$ and $ds < 0$. \square

A.4. Proof of Lemma 5.9

PROOF. On the one hand, (3a) and (4a), respectively, imply that

$$x_i^s = \begin{cases} \frac{V_\rho + D_\rho - 2b_i}{2a_i} & i \in \mathcal{F}_0, \\ \frac{V_\rho - b_i}{2a_i} & i \in \mathcal{F}_s. \end{cases}$$

Since $\sum_{i \in \mathcal{F}} x_i^s = r$, we have

$$\frac{V_\rho}{2}K(\mathcal{F}) = r - \frac{D_\rho}{2}K(\mathcal{F}_0) + \sum_{i \in \mathcal{F}_s} \frac{b_i}{2a_i} + \sum_{i \in \mathcal{F}_0} \frac{b_i}{a_i}. \quad (16)$$

On the other hand, (4a), (5a), and Lemma 5.5 imply that

$$\begin{aligned} \frac{r}{\rho}K(\mathcal{F}_s) &= \frac{V_\rho}{2}K(\mathcal{F}_s)K(\mathcal{F}) + (K(\mathcal{F}_0) - K(\mathcal{F})) \sum_{i \in \mathcal{F}_s} \frac{b_i}{2a_i} \\ &\quad - \frac{D_\rho}{2}K(\mathcal{F}_0)K(\mathcal{F}_s), \end{aligned}$$

and since $\mathcal{F}_s \neq \emptyset$,

$$\frac{V_\rho}{2}K(\mathcal{F}) = \frac{r}{\rho} + \sum_{i \in \mathcal{F}_s} \frac{b_i}{2a_i} + \frac{D_\rho}{2}K(\mathcal{F}_0). \quad (17)$$

Adding and subtracting (16) and (17) yield the claim. \square

A.5. Proof of Proposition 5.13

PROOF. Rebates are beneficial only if the social cost of a Stackelberg equilibrium is lower than that of the Wardrop equilibrium. In that case, $L_\rho < L_\infty$ and, hence, the strict inequality of the claim holds because of Proposition 5.12.

We now focus on the reverse implication. Assuming that the inequality in the hypothesis holds, there exists an $i \in A$ such that

$$L_\infty \left(r \frac{\rho - 1}{\rho}\right) < b_i < L_\infty \left(r \frac{\rho + 1}{\rho}\right).$$

Proposition 5.8 implies that if $i \in A \setminus \mathcal{F}$ then $b_i \geq V_\rho$, and if $i \in \mathcal{F}_0$ then $b_i \leq D_\rho$. Therefore, $i \in \mathcal{F}_s$, which is consequently nonempty. \square

A.6. Proof of Theorem 6.4

First, we express the cost of the system optimum as a function of L_∞ to be able to relate it to a Stackelberg equilibrium.

LEMMA A.1. For networks with parallel arcs and affine cost functions, the minimal value of the participants' real cost is

$$C(x^{\text{SO}}) = \frac{1}{2} \int_{z=0}^{2r} L_\infty(z) dz.$$

PROOF. Proposition 3.5 implies that there exists a constant $L^{\text{SO}} > 0$ such that

$$\begin{aligned} L^{\text{SO}} &= 2a_i x_i^{\text{SO}} + b_i \quad \forall i \text{ s.t. } x_i^{\text{SO}} > 0, \\ L^{\text{SO}} &\leq b_i \quad \forall i \text{ s.t. } x_i^{\text{SO}} = 0. \end{aligned}$$

Proceeding as in Stackelberg equilibrium case, $x_i^{\text{SO}} = [L^{\text{SO}} - b_i]^+ / (2a_i)$. If we let $i^{\text{SO}} := \max\{i \in A : b_i \leq L^{\text{SO}}\}$, we have that $r = \sum_{j=1}^{i^{\text{SO}}} (L^{\text{SO}} - b_j) / (2a_j)$. Hence,

$$L^{\text{SO}} = \frac{1}{K([i^{\text{SO}}])} \left(2r + \sum_{j=1}^{i^{\text{SO}}} \frac{b_j}{a_j}\right).$$

Since $b_{i^{\text{SO}}} \leq L^{\text{SO}} < b_{i^{\text{SO}}+1}$, Proposition 5.11 implies that $L^{\text{SO}} = L_\infty(2r)$. Then

$$\begin{aligned} C(x^{\text{SO}}) &= \sum_{j=1}^{i^{\text{SO}}} x_j^{\text{SO}} \left(L_\infty(2r) - \frac{L_\infty(2r) - b_j}{2}\right) \\ &= rL_\infty(2r) - \frac{1}{2} \sum_{j=1}^{i^{\text{SO}}} \frac{(L_\infty(2r) - b_j)^2}{2a_j}. \end{aligned}$$

Finally, using a similar decomposition as in Figure 4, it can be shown that

$$2rL_\infty(2r) = \int_{z=0}^{2r} L_\infty(z) dz + \sum_{i \in A : b_i \leq L_\infty(2r)} \frac{(L_\infty(2r) - b_i)^2}{2a_i}.$$

The claim follows from the last two equations. \square

Now we are ready to offer the proof of Theorem 6.4.

PROOF. Equations (1b), (13), and (14) imply that

$$\sum_{i \in A} x_i^s c_i(x_i^s) = r \frac{\rho - 1}{\rho} L_\rho + \frac{1}{2} \int_{r(1-1/\rho)}^{r(1+1/\rho)} L_\infty(z) dz. \quad (18)$$

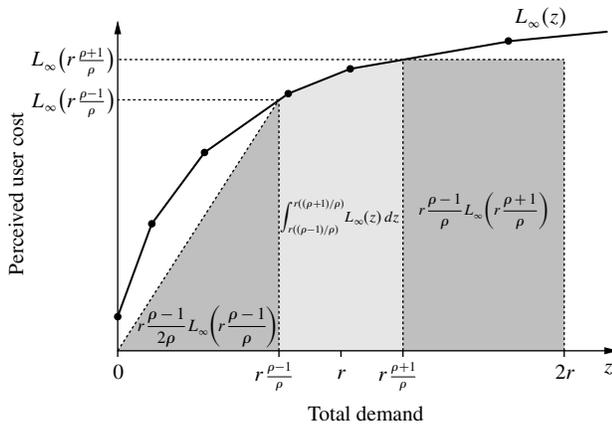


Figure A.1 Illustration of the Bound for $C(x^{SO})$

From Lemma A.1, the concavity of L_{∞} , and decomposing the area under the curve as Figure A.1 illustrates, we have that

$$\begin{aligned} 2C(x^{SO}) &\geq r \frac{\rho-1}{\rho} \frac{D_{\rho}}{2} + r \frac{\rho-1}{\rho} V_{\rho} + \int_{r^{(1-1/\rho)}}^{r^{(1+1/\rho)}} L_{\infty}(z) dz \\ &= 2 \sum_{i \in A} x_i^s c_i(x_i^s) - r \frac{\rho-1}{\rho} \frac{D_{\rho}}{2} \\ &\geq 2 \sum_{i \in A} x_i^s c_i(x_i^s) - r \frac{\rho-1}{\rho} \frac{L_{\rho}}{2} \\ &\geq C(x^s) \left(2 - \frac{\rho-1}{2\rho} \right), \end{aligned}$$

where the second, third, and fourth lines hold because of (18), $D_{\rho} \leq L_{\rho}$, and $rL_{\rho} \leq \sum_i x_i^s c_i(x_i^s)$, respectively. \square

References

Acemoglu, D., A. Ozdaglar. 2007. Competition and efficiency in congested markets. *Math. Oper. Res.* **32**(1) 1–31.

Ali, A., M. A. Jolson, R. Y. Darmon. 1994. A model for optimizing the refund value in rebate promotions. *J. Bus. Res.* **29**(3) 239–245.

Beckmann, M. J., C. B. McGuire, C. B. Winsten. 1956. *Studies in the Economics of Transportation*. Yale University Press, New Haven, CT.

Bergendorff, P., D. W. Hearn, M. V. Ramana. 1997. Congestion toll pricing of traffic networks. P. M. Pardalos, D. W. Hearn, W. W. Hager, eds. *Network Optimization. Lecture Notes in Economics and Mathematical Systems*, Vol. 450. Springer, Berlin, 51–71.

Braess, D. 1968. Über ein Paradoxon aus der Verkehrsplanung. *Unternehmensforschung* **12** 258–268. [Translated by: D. Braess, A. Nagurney, T. Wakolbinger. 2005. On a paradox of traffic planning. *Transportation Sci.* **39**(4) 446–450.]

Button, K. J., E. T. Verhoef. 1998. *Road Pricing, Traffic Congestion and the Environment: Issues of Efficiency and Social Feasibility*. Edward Elgar, Cheltenham, UK.

Chau, C. K., K. M. Sim. 2003. The price of anarchy for nonatomic congestion games with symmetric cost maps and elastic demands. *Oper. Res. Lett.* **31**(5) 327–334.

Chen, X., C.-L. Li, D. Simchi-Levi. 2007. The impact of manufacturer rebates on supply chain profits. *Naval Res. Logist.* **54**(6) 667–680.

Cole, R., Y. Dodis, T. Roughgarden. 2006. How much can taxes help selfish routing? *J. Comput. System Sci.* **72**(3) 444–467.

Correa, J. R., A. S. Schulz, N. E. Stier-Moses. 2004. Selfish routing in capacitated networks. *Math. Oper. Res.* **29**(4) 961–976.

Correa, J. R., A. S. Schulz, N. E. Stier-Moses. 2008. A geometric approach to the price of anarchy in nonatomic congestion games. *Games Econ. Behav.* **64** 457–469.

de Palma, A., N. Picard. 2006. Equilibria and information provision in risky networks with risk-averse drivers. *Transportation Sci.* **40**(4) 393–408.

de Palma, A., M. Kilani, R. Lindsey. 2007. Maintenance, service quality and congestion pricing with competing roads. *Transportation Res.* **41B**(5) 573–591.

Dupuit, J. 1849. De l’Influence des Péages sur l’Utilities des Voies de Communication. *Annales des Ponts et Chaussées* **17**. [Translated by: E. Henderson, 1962, as: *On Tolls and Transport Charges. International Economic Papers*, Vol. 11. Macmillan, London, 7–31.]

Engel, E., R. Fischer, A. Galetovic. 2004. Toll competition among congested roads. *Topics Econ. Anal. Policy* **4**(1).

Friedman, E. J. 2004. Genericity and congestion control in selfish routing. *Proc. 43rd Conf. Decision and Control*, Institute of Electrical and Electronics Engineers, Paradise Island, Bahamas, 4667–4672.

Gerstner, E., J. D. Hess. 1991. Who benefits from large rebates: Manufacturer, retailer or consumer? *Econom. Lett.* **36** 5–8.

Golany, B., U. G. Rothblum. 2006. Inducing coordination in supply chains through linear reward schemes. *Naval Res. Logist.* **53**(1) 1–15.

Hall, M. A. 1978. Properties of the equilibrium state in transportation networks. *Transportation Sci.* **12**(3) 208–216.

Haurie, A., P. Marcotte. 1985. On the relationship between Nash-Cournot and Wardrop equilibria. *Networks* **15**(3) 295–308.

Johari, R., J. N. Tsitsiklis. 2004. Network resource allocation and a congestion game. *Math. Oper. Res.* **29**(3) 407–435.

Johari, R., G. Y. Weintraub, B. Van Roy. 2009. Investment and market structure in industries with congestion. *Oper. Res.* Forthcoming.

Karakostas, G., S. G. Kolliopoulos. 2005. The efficiency of optimal taxes. A. López-Ortiz, A. Hamel, eds. *Combinatorial and Algorithmic Aspects of Networking (CAAN’04)*, Banff, Canada. *Lecture Notes in Computer Science*, Vol. 3405. Springer, Heidelberg, 3–12.

Kheel, T. W. 2008. Balancing free transit and congestion pricing in New York City. Report, Nurture New York’s Nature, New York.

Knight, F. H. 1924. Some fallacies in the interpretation of social cost. *Quart. J. Econom.* **38** 582–606.

Korilis, Y., A. Lazar, A. Orda. 1995. Architecting noncompetitive networks. *IEEE J. Selected Areas Comm.* **13**(7) 1241–1251.

Koutsoupias, E. 2004. Coordination mechanisms for congestion games. *ACM SIGACT News* **35**(4) 58–71.

Koutsoupias, E., C. H. Papadimitriou. 1999. Worst-case equilibria. C. Meinle, S. Tison, eds. *Proc. 16th Annual Sympos. Theoret. Aspects Comput. Sci. (STACS)*, Trier, Germany. *Lecture Notes in Computer Science*, Vol. 1563. Springer, Heidelberg, 404–413.

Labbé, M., P. Marcotte, G. Savard. 1998. A bilevel model of taxation and its application to optimal highway pricing. *Management Sci.* **44**(12) 1608–1622.

Luo, Z.-Q., J.-S. Pang, D. Ralph. 1996. *Mathematical Programs with Equilibrium Constraints*. Cambridge University Press, Cambridge, UK.

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

Papadimitriou, C. H. 2001. Algorithms, games, and the Internet. *Proc. 33rd Annual ACM Sympos. Theory Comput. (STOC)*, Hersonissos, Greece, ACM Press, New York, 749–753.

Partnership for New York City. 2006. Growth or gridlock? The economic case for traffic relief and transit improvement for a greater New York, New York. Report, Partnership for New York City, New York.

Perakis, G. 2007. The “price of anarchy” under nonlinear and asymmetric costs. *Math. Oper. Res.* **32**(3) 614–628.

Perakis, G., G. Roels. 2007. The price of anarchy in supply chains: Quantifying the efficiency of price-only contracts. *Management Sci.* **53**(8) 1249–1268.

Pigou, A. C. 1920. *The Economics of Welfare*. Macmillan, London.

Qiu, L., Y. R. Yang, Y. Zhang, S. Shenker. 2006. On selfish routing in Internet-like environments. *Trans. Networking* **14**(4) 725–738.

- Rosenthal, R. W. 1973. A class of games possessing pure-strategy Nash equilibria. *Internat. J. Game Theory* 2(1) 65–67.
- Roughgarden, T. 2003. The price of anarchy is independent of the network topology. *J. Comput. System Sci.* 67(2) 341–364.
- Roughgarden, T. 2004. Stackelberg scheduling strategies. *SIAM J. Comput.* 33(2) 332–350.
- Roughgarden, T. 2006. On the severity of Braess's paradox: Designing networks for selfish users is hard. *J. Comput. System Sci.* 72(5) 922–953.
- Roughgarden, T., É. Tardos. 2002. How bad is selfish routing? *J. ACM* 49(2) 236–259.
- Santos, G. 2005. Urban congestion charging: A comparison between London and Singapore. *Transport Rev.* 25(5) 511–534.
- Santos, G., G. Fraser. 2006. Road pricing: Lessons from London. *Econom. Policy* 21(46) 263–310.
- Schrank, D., T. Lomax. 2007. The 2007 urban mobility study. Texas Transportation Institute. Accessed September 13, 2009, <http://mobility.tamu.edu/ums>.
- Schulz, A. S., N. E. Stier-Moses. 2006. Efficiency and fairness of system-optimal routing with user constraints. *Networks* 48(4) 223–234.
- Sheffi, Y. 1985. *Urban Transportation Networks*. Prentice-Hall, Englewood, NJ.
- Smith, M. J. 1979. The existence, uniqueness and stability of traffic equilibria. *Transportation Res.* 13B(4) 295–304.
- Starkie, D. 1986. Efficient and politic congestion tolls. *Transportation Res.* 20A(2) 169–173.
- Taylor, T. 2002. Supply chain coordination under channel rebates with sales effort effects. *Management Sci.* 48(8) 992–1007.
- Vickrey, W. S. 1955. A proposal for revising New York's subway fare structure. *J. Oper. Res. Soc. America* 3(1) 38–68.
- Vickrey, W. S. 1969. Congestion theory and transport investment. *Amer. Econom. Rev.* 59 251–261.
- von Stackelberg, H. 1934. *Marktform und Gleichgewicht*. Springer, Wien. [Translated by: A. T. Peacock. 1952. *The Theory of the Market Economy*. Oxford University Press, New York].
- Wardrop, J. G. 1952. Some theoretical aspects of road traffic research. *Proc. Institution of Civil Engineers, Part II*, Vol. 1, 325–378.
- Wichiensin, M., M. G. H. Bell, H. Yang. 2007. Impact of congestion charging on the transit market: An inter-modal equilibrium model. *Transportation Res.* 41A(7) 703–713.
- Xiao, F., H. Yang, D. Han. 2007. Competition and efficiency of private toll roads. *Transportation Res.* 41B(3) 292–308.
- Yang, H., H.-J. Huang. 2005. *Mathematical and Economic Theory of Road Pricing*. Elsevier, Oxford, UK.
- Yang, H., W. Xu, B. Heydecker. 2009. Bounding the efficiency of road pricing. *Transport. Res. Part E*. Forthcoming.

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

Vladimir Fux

Patrick Maillé

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

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

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

I. INTRODUCTION

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

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

from the diversity of technologies, operators have to improve resource management.

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

II. RELATED WORK

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

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

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

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

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

III. MODEL AND PROBLEM FORMULATION

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

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

A. Mathematical formulation

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

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

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

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

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

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

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

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

with f_i the total flow on network i .

With this type of latency function we also have to assume that $D < \sum_{i=1}^n c_i$, in other words the aggregated capacity is enough to treat all demand. We assume that the provider

owning all considered networks is interested in minimizing the social cost (or total cost) expressed as:

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

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

B. Routing game interpretation

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

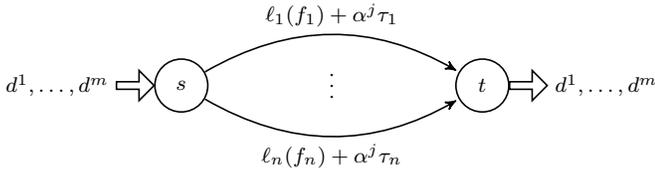


Fig. 1. Logic representation of the network selection problem as a routing problem: the perceived cost on each route i depends on the load f_i and the tax τ_i , but also on the user type j through the sensitivity α^j .

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

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

C. The case of several providers

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

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

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

IV. USER EQUILIBRIUM AND OPTIMAL SITUATIONS

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

Assumption B:

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

A. User equilibrium

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

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

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

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

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

B. User equilibrium without taxes

Consider the case when the provider does not charge taxes for using his networks (or equivalently all taxes are the same), and thus users make their choices without any intervention

from the provider. Then the flows at a Wardrop equilibrium have the form stated in the following proposition.

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

$$f_i^{\text{WE}} = \begin{cases} \frac{D - \sum_{q=1}^t c_q + t c_t}{t} & \text{if } i \leq t, \\ 0 & \text{otherwise,} \end{cases} \quad (5)$$

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

$$D - \sum_{i=1}^t c_i + t c_t > 0, \quad (6)$$

and represents the number of used networks.

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

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

C. Optimal situation

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

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

$$\text{s.t.} \begin{cases} \sum_{i=1}^n f_i = D \\ f_i \geq 0, \text{ for } i = 1, \dots, n \end{cases} \quad (8)$$

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

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

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

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

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

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

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

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

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

$$\begin{cases} f_i^{\text{opt}} > 0 \Rightarrow \ell_i(f_i^{\text{opt}}) + f_i^{\text{opt}} \ell'_i(f_i^{\text{opt}}) = H, \\ f_i^{\text{opt}} = 0 \Rightarrow \ell_i(f_i^{\text{opt}}) + f_i^{\text{opt}} \ell'_i(f_i^{\text{opt}}) = \ell_i(0) \geq H. \end{cases} \quad (12)$$

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

$$f_i^{\text{opt}} > 0 \Leftrightarrow \frac{1}{c_i} < H, \quad (13)$$

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

$$f_i^{\text{opt}} = c_i - \frac{\sqrt{c_i}}{\sqrt{H}}, i = 1, \dots, k, \quad (14)$$

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

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

V. ELICITING OPTIMAL USER-NETWORK ASSOCIATIONS WITH TAXES

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

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

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

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

for some constant K .

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

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

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

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

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

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

thus

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

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

Now we provide a method to calculate the optimal taxes:

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

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

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

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

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

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

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

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

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

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

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

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

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

These two inequalities imply that

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

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

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

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

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

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

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

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

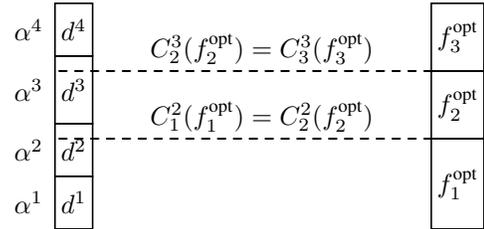


Fig. 2. Example of user distribution among networks with optimal taxes for the case $m = 4, n = 3$: class-1 (resp. class-4) users all attach to network 1 (resp. 3), while class-2 (resp. class-3) users are split among networks 1 and 2 (resp. 2 and 3).

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

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

VI. EFFICIENCY ANALYSIS

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

A. Influence of heterogeneity on the PoA

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

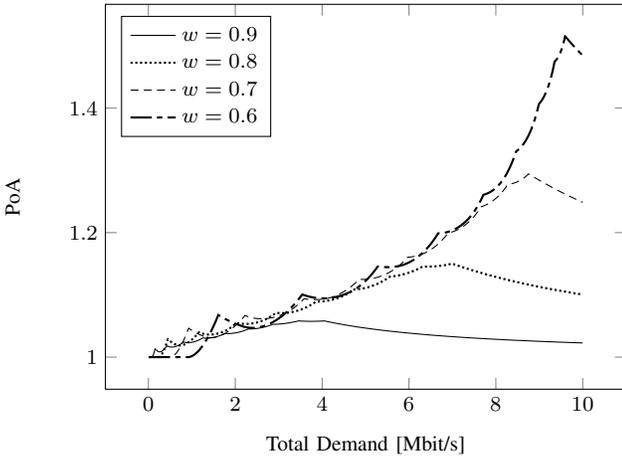


Fig. 3. PoA versus total demand D with $n = 10$ and total capacity equal to 10 [Mbit/s].

B. The PoA interpretation

Finally, we present two counterparts for the Price of Anarchy in our model. For simplicity, we consider only a case with two networks in which $c_1 = 4$ [Mbit/s] and $c_2 = 11$

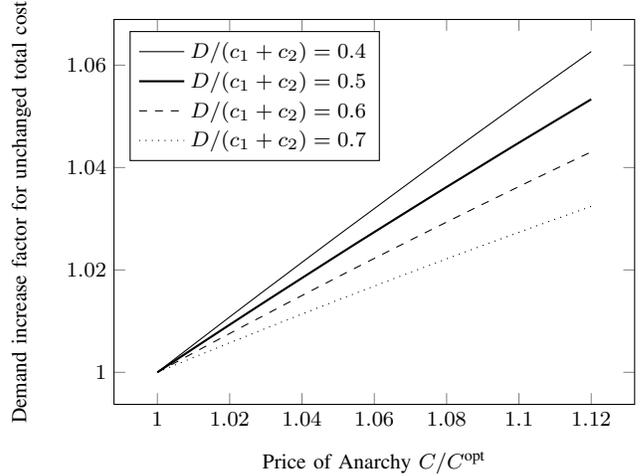


Fig. 4. Demand gain versus PoA, for different demand levels in the case of two networks.

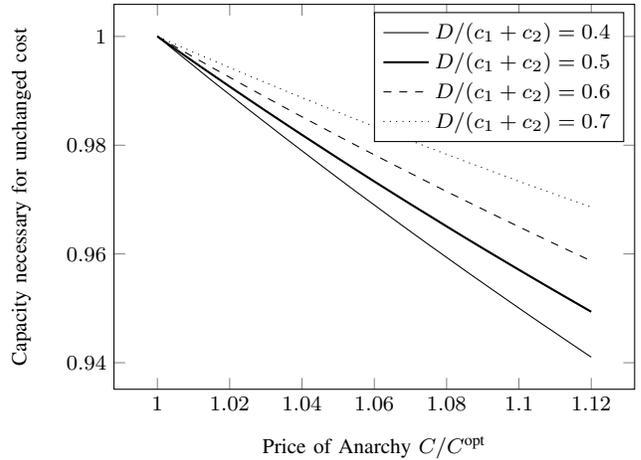


Fig. 5. Capacity gain versus PoA, for different demand levels in the case of two networks.

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

VII. CONCLUSIONS AND PERSPECTIVES

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

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

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

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

ACKNOWLEDGEMENTS

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

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

REFERENCES

- [1] R. J. Aumann and L. S. Shapley, *Values of Non-Atomic Games*. Princeton University Press, 1974.
- [2] M. Beckmann, C. B. McGuire, and C. B. Winsten, *Studies in the Economics of Transportation*. Yale University Press, 1956.
- [3] M. Cesana, N. Gatti, and I. Malanchini, “Game theoretic analysis of wireless access network selection: Models, inefficiency bounds, and algorithms,” in *Proceedings of 2nd Workshop on Game Theory in Communication Networks (GameComm)*, October 2008.
- [4] R. Cole, Y. Dodis, and T. Roughgarden, “Pricing network edges for heterogeneous selfish users,” in *Proceedings of 35th Annual ACM Symposium on Theory of Computing*, 2003.

- [5] P. Coucheny, E. Hyon, C. Touati, and B. Gaujal, “Myopic versus clairvoyant admission policies in wireless networks,” in *Proceedings of the Fourth International ICST Conference on Performance Evaluation Methodologies and Tools (VALUETOOLS)*, 2009.
- [6] M. Coupechoux, J.-M. Kelif, and P. Godlewski, “Network controlled joint radio resource management for heterogeneous networks,” in *Proceedings of IEEE Vehicular Technology Conference (VTC)*, 2008.
- [7] A. De La Oliva, A. Banchs, I. Soto, T. Melia, and A. Vidal, “An overview of IEEE 802.21: media-independent handover services,” *IEEE Wireless Communications*, vol. 15, no. 4, pp. 96–103, August 2008.
- [8] L. Fleischer, K. Jain, and M. Mahdian, “Tolls for heterogeneous selfish users in multicommodity networks and generalized congestion games,” in *Proceedings of IEEE Symposium on Foundations of Computer Science*, 2004.
- [9] V. Fux, P. Maillé, J.-M. Bonnin, and N. Kaci, “Efficiency or fairness: managing applications with different delay sensitivities in heterogeneous wireless networks,” in *Proceedings of WoWMoM*, 2013.
- [10] E. Gustafsson and A. Jonsson, “Always best connected,” *IEEE Wireless Communications*, February 2003.
- [11] N. Kaci, P. Maillé, and J.-M. Bonnin, “Performance of wireless heterogeneous networks with always-best connected users,” in *Proceedings of 5th Conference on Next Generation Internet Networks (NGI)*, 2009.
- [12] G. Karakostas and S. G. Kolliopoulos, “Edge pricing of multicommodity networks for heterogeneous selfish users,” in *Proceedings of 45th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2004.
- [13] G. Karakostas, S. G. Kolliopoulos, and G. Stavros, “Edge pricing of multicommodity networks for heterogeneous selfish users with elastic demands,” in *Proceedings of 12th annual international conference on Computing and Combinatorics*, 2006.
- [14] D. Kumar, E. Altman, and J.-M. Kelif, “User-network association in a WLAN-UMTS hybrid cell: Global individual optimality,” in *Proceedings of International Teletraffic Congress (ITC)*, 2006.
- [15] P. Maillé, B. Tuffin, and J.-M. Vigne, “Technological investment games among wireless telecommunications service providers,” *International Journal on Network Management - Special Issue: Economic Traffic Management*, vol. 21, no. 1, pp. 65–82, 2011.
- [16] —, “Competition between wireless service providers sharing a radio resource,” in *Proceedings of 11th IFIP Networking Conference*, Prague, Czech Republic, May 2012.
- [17] K. Mittal, E. M. Belding, and S. Suri, “A game-theoretic analysis of wireless access point selection by mobile users,” *Computer Communications*, vol. 31, no. 10, 2008.
- [18] A. C. Pigou, *The economics of welfare*. Macmillan, 1920.
- [19] J. Wardrop, “Some theoretical aspects of road traffic research,” *Proceedings of the Institute of Civil Engineers*, 1952.

Interplay between security providers, consumers, and attackers: a weighted congestion game approach

Patrick Maillé¹, Peter Reichl², and Bruno Tuffin³

¹ Institut Telecom; Telecom Bretagne
2 rue de la Châtaigneraie CS 17607
35576 Cesson Sévigné Cedex, France
`patrick.maille@telecom-bretagne.eu`

² FTW

Donau-City-Str. 1
A-1220 Wien, Austria
`reichl@ftw.at`

Université européenne de Bretagne
³ INRIA Rennes - Bretagne Atlantique
Campus universitaire de Beaulieu
35042 Rennes Cedex, France
`bruno.tuffin@inria.fr`

Abstract. Network users can choose among different security solutions to protect their data. Those solutions are offered by competing providers, with possibly different performance and price levels. In this paper, we model the interactions among users as a noncooperative game, with a negative externality coming from the fact that attackers target popular systems to maximize their expected gain. Using a nonatomic weighted congestion game model for user interactions, we prove the existence and uniqueness of a user equilibrium, and exhibit the tractability of its computation, as a solution of a convex problem. We also compute the corresponding Price of Anarchy, that is the loss of efficiency due to user selfishness, and investigate some consequences for the (higher-level) pricing game played by security providers.

Keywords: Security, Game theory, Competition

1 Introduction

Within the current evolution towards the Future Internet, the provision of appropriate network security is considered to be one of the most difficult as well as most challenging tasks. Among the broad range of related research approaches, the attempt to better understand the mindset of attackers serves for sure as one of the key sources for developing advanced protection mechanisms. Cybercrime concerns colossal amounts of money, and is highly organized so that attacker

efforts are rationalized to maximize the associated gains. This is why we model here an interesting negative externality effect of security architectures and systems, through the attractiveness for potential attackers: majority products are likely to be larger targets for hackers, and therefore become less attractive for consumers. Then, the choice of a particular system and security protection -that we will call a security provider from now on- by the whole online population can now be considered as a congestion game, where congestion is not considered in the common sense of an excessive demand for a finite resource amount, but more generally as a degradation of the performance on a given choice when it gets too popular. Here the performance degradation is indirect, since it stems from the behavior of attackers.

In the specific context of security, the link between the audience of a system and its attractiveness to attackers can be further described when attacks are intended to steal or damage data: an attacker would be attracted by the potential gain (or damage) of the attack, which depends on the value of the users' data, but that value affects (and is therefore, to some extent, revealed by) the security option users choose. For example, the "safest" solutions may attract users with high-value data to protect, making those solutions an interesting target for an attacker even if their market share is small.

In this paper, we propose a model that encompasses that effect, by considering users with heterogeneous data values making a choice among several security possibilities. The criteria considered in that choice are the security protection level -measured by the likeliness of having one's data stolen or damaged, that is subject to negative externalities- and the price set by the security provider.

The literature on network security involving game-theoretic models and tools is recent and still not very abundant. Some very interesting works have been published regarding the interactions between attacking and defending entities, where the available strategies can consist in spreading effort over the links of a network [6,15] or over specific targets [8], or in selecting some particular attack or defense measures [5,11]. In those references, the security game is a zero-sum game between two players only, and therefore no externalities among several potential defenders are considered.

Another stream of work considers security protection investments, through models that encompass positive externalities among users: indeed, when considering epidemic attacks (like, e.g., worms), the likeliness of being infected decreases with the proportion of neighbors that are protected. Since protection has a cost and users selfishly decide to protect or not without considering the externality they generate, the equilibrium outcome is such that investment is suboptimal [12] and needs to be incentivized through specific measures [17]. For more references on game theory applied to network security contexts, see [1,18].

In contrast, the work presented here considers negative externalities in the choices of security software/procedures. As highlighted above, the negative externality comes from the attractiveness of security solutions for attackers. Such situations can arise when attacks are not epidemic but rather direct, as are attacks targeting randomly chosen IP addresses. The interaction among users can

then be modeled as a population game, that is a game where the user payoffs for a given strategy (here, a security solution) change as more users choose that same strategy [10]. Such games are particular cases of so-called *congestion games* where user strategies are subsets of a given set of resources, and the total cost experienced by users is the sum of the costs on each resource [2,22]. Here, users select only one resource, and congestion corresponds to the fact that the more customers, the more likely an attack.

In this paper, we consider a very large population, where the extra congestion created by any individual user is negligible. The set of players can therefore be considered as a continuum; note that such games are called *nonatomic* [29]. The study of nonatomic congestion games has seen recent advances for the case when all users are identical or belong to a finite set of populations [7,14,24,25,26], but we want here to encompass the larger attractiveness to attackers of “rich” users, compared to the ones with no valuable data online. More precisely, we intend to model the heterogeneity in users congestion effects, by introducing a distribution among users valuation for the data to protect. The congestion game is therefore *weighted* in the sense that not all users contribute to congestion in an identical manner. Fewer results exist for those games [4,21], even when user strategies only consist in choosing one resource among a common strategy set.

Moreover, in our model users undergo the congestion cost of the security solution they select - which depends on the congestion as well as on their particular data valuation -, but also the monetary cost associated to that solution - which is the same for all users -. As a result, following [20,21] the game would be called a *weighted congestion game with separable preferences*, and can be transformed into an equivalent *weighted congestion game with player-specific constants* [19] (i.e., the payoffs of users selecting the same strategy only differ through a user-specific additive constant). In general, the existence of an equilibrium is not ensured for such games when the number of users is finite [19,20,21]. In the nonatomic case, the existence of a mixed equilibrium is ensured by [29] and the loss of efficiency due to user selfishness is bounded [4], but the existence of a pure equilibrium in the general case is not guaranteed.

In this paper, we establish the *existence* and essential *uniqueness* of a pure equilibrium for our model, as well as its *tractability* by proving that an equilibrium solves a strictly convex optimization problem. To the best of our knowledge, such proofs for nonatomic games had only been given for unweighted games [27,28], with a finite number of different user populations; here we consider a weighted game with possibly an infinity of different weight values, with the specificity that the differences in user congestion weights are directly linked to their user-specific valuations.

The remainder of the paper is organized as follows. The model is formally introduced in Section 2. We focus on the user equilibrium existence, uniqueness and tractability in Section 3, and give an upper bound on the loss of efficiency due to user selfishness. The results are then applied in Section 4 to give some insights about the prices that profit-oriented security providers should set. We conclude and suggest directions for future work in Section 5.

2 Model

We consider a set \mathcal{I} of security providers (each one on a given architecture), and define $I := |\mathcal{I}|$.

2.1 User data valuation

Users differ with the valuation for their data. When an attack is successful over a target user u , that user is assumed to experience a financial loss $v_u \geq 0$, which we call her data valuation. The distribution of valuations over the population is given by a cumulative distribution function F on \mathbb{R}^+ , where $F(v)$ represents the proportion of users with valuation lower than or equal to v . Since users who do not value their data (i.e., for whom $v_u = 0$) will not play any role in our model, we can ignore them; the distribution function F is therefore such that $F(0) = 0$. The overall total “mass” of users is finite, and through a unit change we can assume it to be 1 without loss of generality.

Equivalently, the repartition F of user preferences among the population can be represented by its corresponding *quantile function* $q : [0, 1) \rightarrow \mathbb{R}^+$. For $x \in [0, 1)$, the quantity $q(x)$ represents the valuation⁴ of the (infinitesimal) user at (continuous) position x on a valuation-related increasing ranking. Formally, we have

$$\forall x \in [0, 1), \quad q(x) = \inf\{v \in \mathbb{R}^+ : F(v) \geq x\}, \quad (1)$$

$$\forall v \in \mathbb{R}^+, \quad F(v) = \inf\{x \in [0, 1) : q(x) > v\}, \quad (2)$$

with the convention $\inf \emptyset := 1$ in the latter equation. Note that F is right-continuous, while the quantile function q is left-continuous. Both functions are nonnegative and nondecreasing.

We may not suppose that the support of F , that we denote by S_v , is bounded, but we assume that the overall value of the data in the population is finite, i.e.,

$$V_{\text{tot}} := \int_{S_v} v \, dF(v) < +\infty.$$

Finally, we define $\mathcal{N}(V)$ as the user mass⁵ such that the total data valuation for the $\mathcal{N}(V)$ users with smallest valuation exactly equals V :

$$\forall V \in [0, V_{\text{tot}}), \quad \mathcal{N}(V) := \min \left\{ x : \int_{y=0}^x q(y) dy = V \right\}.$$

$\mathcal{N}(V)$ is obtained by inverting the bijective function

$$\begin{aligned} \mathcal{V} : [0, 1] &\mapsto [0, V_{\text{tot}}] \\ x &\mapsto \mathcal{V}(x) = \int_{y=0}^x q(y) dy. \end{aligned} \quad (3)$$

⁴ Except, possibly, on a zero-measure set of users.

⁵ i.e., proportion since we normalized the total user mass to 1.

Notice that \mathcal{V} is continuous and differentiable on $[0, 1]$, with left-derivative $q(x)$ and right-derivative $q(x^+)$, where $q(x^+) = \lim_{y \rightarrow x, y > x} q(y)$. Since q is nondecreasing and strictly positive for $x > 0$, then \mathcal{V} is convex and strictly increasing on $[0, 1]$. As a result, its inverse function \mathcal{N} is concave on $(0, V_{\text{tot}})$, and has left-derivative

$$\mathcal{N}'_l(V) = \frac{1}{q(\mathcal{N}(V))} \quad (4)$$

and right-derivative

$$\mathcal{N}'_r(V) = \frac{1}{q(\mathcal{N}(V)^+)}. \quad (5)$$

The distribution F , the quantity V_{tot} as well as the functions q and \mathcal{N} are illustrated in Figure 1.

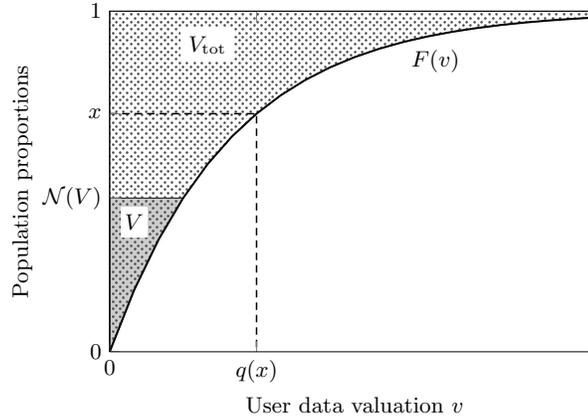


Fig. 1. Values and functions of interest regarding the user valuation distribution F .

2.2 Security systems performance

In this paper, we focus on direct attacks targeting some specific machines, which may for instance come from an attack-generating robot that randomly chooses IP addresses and launches attacks to those hosts.

The attacks generated by such a scheme have to target a specific vulnerability of a given security system. As a result, the attacker has to select which security system $i \in \mathcal{I}$ to focus on. If an attack is launched to a security system i , we consider that all machines protected by a system $j \neq i$ do not run any risk, while the success probability of the attack is supposed to be fixed, denoted by π_i , on machines with protection system i . In other terms, the parameter π_i measures the effectiveness of the security defense.

2.3 The attacker point of view

Successful attacks bring some revenue to the attacker. Be it in terms of damage done to user data, or in terms of stolen data from users, it is reasonable to consider that for a given attack, the gain for the attacker is proportional to the value that the data had to the victim. Indeed, in the case of data steal, more sensitive data (e.g., bank details) are more likely to bring high revenues when used. Likewise, when the objective of the attacker is simply to maximize user damage, then the link between attacker utility and user data valuation is direct.

For a given distribution of the population among providers, let F_i be the (unconditional) distribution of valuations of users associated with provider i , so that $F = \sum_{i \in \mathcal{I}} F_i$. We then define for each provider $i \in \mathcal{I}$ the total value of the protected data, as

$$V_i := \int v dF_i(v). \quad (6)$$

For an attacker, the expected benefit from launching an attack targeted at system i (without knowing which users are with provider i) is thus proportional to $\pi_i V_i$. We therefore assume that the likeliness of attacks occurring on system i is a nondecreasing function of $\pi_i V_i$. We discretize time, and denote by $R_i(\pi_i V_i)$ the probability that a particular user is the target of a system- i attack over a time period. Remark that we consider system-specific functions $(R_i)_{i \in \mathcal{I}}$, so that the model can encompass some heterogeneity in the difficulty of creating system-targeted attacks.

To simplify a bit the writing, let us define $T_i(V_i)$ as the risk, for a user, of having one's data compromised when choosing security provider i . Note that it can be written as a function of the total protected data value V_i :

$$T_i(V_i) := \pi_i R_i(\pi_i V_i) = \pi_i R_i\left(\pi_i \int v dF_i(v)\right). \quad (7)$$

We will often make use of the assumption below.

Assumption A *For all $i \in \mathcal{I}$, T_i is a continuous and strictly increasing function of V_i , and $T_i(0) = 0$.*

For T_i functions of the form given in (7), Assumption A is equivalent to

- $\pi_i > 0$ for all $i \in \mathcal{I}$ (no provider offers a perfect protection against attacks),
- R_i is a continuous and strictly increasing function with $R_i(0) = 0$, for all $i \in \mathcal{I}$ (attackers do not target providers not protecting valuable data).

2.4 User preferences

For a user u with data valuation v_u , the *total expected cost* at provider i depends on the risk of being (successfully) attacked, and on the price p_i charged by the security provider. That total cost is therefore given by

$$v_u T_i(V_i) + p_i.$$

To ensure that all users select one option, we can assume that there exists a provider i with $p_i = 0$, which would correspond to security solutions offered by free software communities (e.g., avast!⁶). Indeed, if $p_i = 0$, the total cost is the valuation times a product of probabilities, and therefore less than the valuation itself, so that this choice of a free service is always a valuable option⁷.

Remark that we consider risk-neutral users here, as may be expected from large entities, while private individuals should rather be considered risk-averse. Nevertheless, one can imagine some extra mechanisms (e.g., insurance [17]) to reach a risk-neutral equivalent formulation.

3 User equilibrium

In this section, we investigate how demand is split among providers, when their prices p_i and security levels π_i are fixed. Recall we assumed that users are infinitely small: their individual choices do not affect the overall user distribution among providers (and therefore the total values $(V_i)_{i \in \mathcal{I}}$).

The outcome from such user interactions should be determined by user selfishness: demand should be distributed in such a way that each user u chooses one of the cheapest providers (in terms of perceived price) with respect to her valuation v_u and the current risk values $(T_i(V_i))_{i \in \mathcal{I}}$. Such a distribution of users among providers, if it exists, will be called a *user equilibrium*. In other words, if provider $i \in \mathcal{I}$ is chosen by some users u , then it is cheaper for those users (in terms of total expected cost) than any other provider $j \in \mathcal{I}$, otherwise they would be better off switching to j . Formally,

$$i \in \arg \min_{j \in \mathcal{I}} v_u T_j(V_j) + p_j.$$

We use here the nonatomicity assumption: each user u considers the values $(V_j)_{j \in \mathcal{I}}$ as fixed when making her individual choice.

3.1 Structure of a user equilibrium

We now investigate the existence and uniqueness of a user equilibrium, for fixed values of prices and attack success probabilities. To do so, we first define the notion of *user repartition*.

Definition 1. Denote by $\mathcal{P}_{\mathcal{I}}$ the set of probability distributions over providers in \mathcal{I} , i.e., $\mathcal{P}_{\mathcal{I}} := \{(y_1, \dots, y_I) \geq 0, \sum_{i \in \mathcal{I}} y_i = 1\}$. For a given price profile $p = (p_1, \dots, p_I)$, a user repartition is a mapping $A : S_v \mapsto \mathcal{P}_{\mathcal{I}}$, that is interpreted as follows:

For all $v \in S_v$, among users with valuation v , a proportion $A_i(v)$ chooses provider i , where $A(v) = (A_1(v), \dots, A_I(v))$.

⁶ <http://www.avast.com>

⁷ We implicitly assume here that each user u is willing to pay at least v_u to benefit from the online service.

Therefore, to a given user repartition A corresponds a unique distribution $\mathbf{V} = (V_i)_{i \in \mathcal{I}}$ of the total data valuation V_{tot} among providers, given by

$$V_i(A) = \int_{v \in S_v} v A_i(v) dF(v) \quad \forall i \in \mathcal{I}. \quad (8)$$

Remark also that $F_i(v) = \int_{w \leq v} A_i(w) dF(w)$.

Reciprocally, we say that a distribution $\mathbf{V} = (V_i)_{i \in \mathcal{I}}$ of the data valuation is *feasible* if $V_i \geq 0$ for all i , and $\sum_{i \in \mathcal{I}} V_i = V_{\text{tot}}$. For a feasible distribution \mathbf{V} , when providers are sorted such that $p_1 \leq \dots \leq p_I$, we define for each $i \in \mathcal{I} \cup \{0\}$ the quantity

$$V_{[i]} := \sum_{j=1}^i V_j,$$

with $V_{[0]} = 0$. $V_{[i]}$ therefore represents the total value of the data protected by the i cheapest providers.

We now formally define the outcome that we should expect from the interaction of users, i.e., an *equilibrium* situation.

Definition 2. *A user equilibrium is a user repartition A^{eq} such that no user has an interest to switch providers. In other words, for any value $v \in S_v$, a user with valuation v cannot do better than following the provider choice given by $A^{eq}(v)$. Formally, A^{eq} is a user equilibrium if and only if*

$$\forall v \in S_v, \quad A_i^{eq}(v) > 0 \quad \Rightarrow \quad i \in \arg \min_{j \in \mathcal{I}} v T_j(V_j(A^{eq})) + p_j, \quad (9)$$

where $V_j(A^{eq})$ is given by (8).

We now establish some monotonicity properties that should be verified by a user equilibrium: if a user y values her data strictly less than another user x , then she selects cheaper (in terms of price) providers than x .

Lemma 1. *Consider a user equilibrium A^{eq} . Then user choices -in terms of price of the chosen provider(s)- are monotone in their valuation: for any two users x and y with respective valuations v_x and v_y , and any providers i and j ,*

$$(v_x - v_y) \cdot A_i^{eq}(v_x) \cdot A_j^{eq}(v_y) > 0 \quad \Rightarrow \quad p_i \geq p_j. \quad (10)$$

Proof. Let us write $V_i := V_i(A^{eq})$ and $V_j := V_j(A^{eq})$. From (9) applied to users x and y , the left-hand inequality of (10) implies

$$\begin{aligned} v_x T_i(V_i) + p_i &\leq v_x T_j(V_j) + p_j \\ \text{and} \quad v_y T_i(V_i) + p_i &\geq v_y T_j(V_j) + p_j. \end{aligned} \quad (11)$$

Subtracting those inequalities gives $T_i(V_i) \leq T_j(V_j)$ since $(v_x - v_y) > 0$. Then (11) yields the right-hand side of (10).

We then use that result to prove that for a given value repartition $(V_i)_{i \in \mathcal{I}}$ over the providers, there can be only one equilibrium repartition if all providers set different prices.

Lemma 2. *Assume that all providers set different prices. If a user equilibrium exists, it is completely characterized (unless for a zero-measure set of users) by the total values $(V_i)_{i \in \mathcal{I}}$ of protected data for each provider $i \in \mathcal{I}$, provided that $\sum_{i \in \mathcal{I}} V_i = V_{\text{tot}}$.*

Proof. Without loss of generality, assume that provider prices are sorted, such that $p_1 < p_2 < \dots < p_I$.

From Definition 1 and (8), to a given equilibrium corresponds a unique set of values $(V_i)_{i \in \mathcal{I}}$.

Reciprocally, consider a feasible data value repartition $\mathbf{V} = (V_i)_{i \in \mathcal{I}}$, and assume it corresponds to a user equilibrium A^{eq} . Since we do not differentiate users with similar valuations, we can sort them -still without loss of generality- in an increasing order of the price of their chosen provider: if $x < y$ and $q(x) = q(y)$ then we can impose that $p_{i_x} \leq p_{i_y}$, where i_x (resp. i_y) would be the (unique) provider chosen by user at position x (resp. y) in the user valuation ranking. Therefore from Lemma 1, at the user equilibrium A^{eq} , provider prices can be considered as sorted in a increasing order of user valuations among *all* users. Thus, user choices are uniquely (unless on a zero-measure user set) determined by their position $x \in [0, 1]$ in the user valuation ranking, and given by

$$\mathcal{V}(x) \in (V_{[i-1]}, V_{[i]}) \Rightarrow \text{user } x \text{ selects provider } i, \quad (12)$$

where \mathcal{V} is defined in (3).

3.2 The case of several providers with the same price

In this subsection, we establish a way to consider several providers with the same price as one single option from the user point of view. Let us consider a common price p , and define $\mathcal{I}_p := \{i \in \mathcal{I} : p_i = p\}$.

First, if one such provider i gets positive demand (i.e., $V_i > 0$), then at a user equilibrium all providers with the same price also get positive demand: indeed, Assumption A implies that $T_i(V_i) > 0$, and thus the total cost of a user u with positive valuation choosing provider $i \in \mathcal{I}_p$ is $v_u T_i(V_i) + p > p$. Therefore each provider $j \in \mathcal{I}_p$ necessarily has a strictly positive T_j , otherwise it would have cost $v_u T_j(0) + p = p$ for user u , who would be better off switching from i to j . Consequently, at a user equilibrium we necessarily have $T_i(V_i) = T_j(V_j)$.

When the set of users choosing one of the providers with price p is fixed, so is the total valuation $V_{\mathcal{I}_p}$ of those users' data. Consequently, the distribution of users among all providers in \mathcal{I}_p should be such that

$$\begin{cases} i, j \in \mathcal{I}_p \Rightarrow T_i(V_i) = T_j(V_j) \\ \sum_{i \in \mathcal{I}_p} V_i = V_{\mathcal{I}_p}. \end{cases} \quad (13)$$

Following [2], we reformulate (13) as a minimization problem:

$$\begin{aligned} (V_i)_{i \in \mathcal{I}_p} \in \arg \min_{(x_i)_{i \in \mathcal{I}_p} \geq 0} \sum_{i \in \mathcal{I}_p} \int_{y=0}^{x_i} T_i(y) dy \\ \text{s.t. } \sum_{i \in \mathcal{I}_p} x_i = V_{\mathcal{I}_p}. \end{aligned} \quad (14)$$

Under Assumption A, there exists a unique vector of values $(V_i)_{i \in \mathcal{I}_p}$ satisfying the above system. In the following, we will denote by $T_{\mathcal{I}_p}(V)$ the corresponding common value of $T_i(V_i)$. Interestingly, remark that the function $T_{\mathcal{I}_p}$ that we have defined also satisfies Assumption A. As a result, in the rest of the analysis of user equilibria, we will associate providers with the same price p and consider them as a single choice \mathcal{I}_p that we assimilate as a single provider k , with corresponding risk function $T_k(V) := T_{\mathcal{I}_p}(V)$ satisfying Assumption A.

3.3 Game equilibrium as a solution of an optimization problem

Based on the reasoning in Subsection 3.2, we assume that all providers submit a different price, and we sort them such that $p_1 < \dots < p_I$. Now let us consider the following measure:

$$\mathcal{L}(\mathbf{V}, \mathbf{p}) := \sum_{i \in \mathcal{I}} \left(\int_{y=0}^{V_i} T_i(y) dy + p_i \underbrace{\left(\mathcal{N}(V_{[i]}) - \mathcal{N}(V_{[i-1]}) \right)}_{\text{Market share of prov. } i} \right) \quad (15)$$

$$= \sum_{i=1}^I \int_{y=0}^{V_i} T_i(y) dy + p_I - \sum_{i=1}^{I-1} (p_{i+1} - p_i) \mathcal{N}(V_{[i]}), \quad (16)$$

with $p_0 := 0$. Remark that the first part of the quantity $\mathcal{L}(\mathbf{V}, \mathbf{p})$ in (15) is the potential function usually associated to unweighted congestion games (see, e.g., [2]), while the second part stands for the total price paid by all users.

The expression (16) highlights the fact that \mathcal{L} is a strictly convex function of \mathbf{V} , since \mathcal{N} is concave and under Assumption A, T_i is strictly increasing. It thus admits a unique minimum \mathbf{V}^* on the (convex) domain of feasible value shares; and \mathbf{V}^* is completely characterized by the first-order conditions. We now prove that this valuation repartition \mathbf{V}^* actually corresponds to a user equilibrium.

Proposition 1. *Let Assumption A hold. For any price profile \mathbf{p} , there exists a user equilibrium, that is completely characterized by the valuation repartition \mathbf{V}^* , unique solution of the convex optimization problem*

$$\min_{\mathbf{V} \text{ feasible}} \mathcal{L}(\mathbf{V}, \mathbf{p}). \quad (17)$$

Proof. We first consider the feasible directions consisting in switching some infinitesimal amount of value from $i > 1$ to $j < i$, when $V_i^* > 0$. The optimality condition in (16) then yields

$$\begin{aligned} 0 &\leq T_j(V_j^*) - T_i(V_i^*) - \sum_{k=j}^{i-1} (p_{k+1} - p_k) \mathcal{N}'_r(V_{[k]}^*) \\ &\leq T_j(V_j^*) - T_i(V_i^*) - (p_i - p_j) \mathcal{N}'_r(V_{[i-1]}^*), \end{aligned} \quad (18)$$

where the second line comes from the concavity of \mathcal{N} .

Notice that since $p_j < p_i$ and \mathcal{N} is nondecreasing, (18) and Assumption A imply that $V_j^* > 0$. Consequently, if we define $i^* := \max\{i \in \mathcal{I} : V_i^* > 0\}$, then

$$V_i^* > 0 \Leftrightarrow i \leq i^*. \quad (19)$$

As a result, since $V_i > 0$ and $i > 1$ in (18), then $0 < V_{[i-1]}^* < V_{\text{tot}}$. Thus, from (5), $\mathcal{N}'_r(V_{[i-1]}^*) = \frac{1}{q(\mathcal{N}(V_{[i-1]}^*))}$ is strictly positive. (18) is then equivalent to

$$\underline{v}_i^* T_i(V_i^*) + p_i \leq \underline{v}_i^* T_j(V_j^*) + p_j, \quad (20)$$

with $\underline{v}_i^* := q(\mathcal{N}(V_{[i-1]}^*))^+ = \inf\{v : \int_{u=0}^v udF(u) > V_{[i-1]}^*\}$. Remark that necessarily from (20), $T_i(V_i^*) < T_j(V_j^*)$ since $p_i > p_j$.

For $i < I$ such that $V_i^* > 0$ (i.e., $i \leq i^*$), we now investigate the possibility of switching some value from i to $j > i$. Still applying the optimality condition for \mathbf{V}^* , we get

$$\begin{aligned} 0 &\leq T_j(V_j^*) - T_i(V_i^*) + \sum_{k=i}^{j-1} (p_{k+1} - p_k) \mathcal{N}'_l(V_{[k]}^*) \\ &\leq T_j(V_j^*) - T_i(V_i^*) + (p_j - p_i) \mathcal{N}'_l(V_{[i]}^*), \end{aligned} \quad (21)$$

where we used again the concavity of \mathcal{N} .

Applying (4), Relation (21) is equivalent to

$$\bar{v}_i^* T_i(V_i^*) + p_i \leq \bar{v}_i^* T_j(V_j^*) + p_j, \quad (22)$$

with $\bar{v}_i^* = q(\mathcal{N}(V_{[i]}^*)) = \inf\{v : \int_{u=0}^v udF(u) \geq V_{[i]}^*\}$.

Relations (20) and (22) can be interpreted as users with valuation $v \in [\underline{v}_i^*, \bar{v}_i^*]$ preferring provider i over any other one, for the repartition value \mathbf{V}^* . Formally,

$$v \in [\underline{v}_i^*, \bar{v}_i^*] \Rightarrow i \in \arg \min_{j \in \mathcal{I}} v T_j(V_j^*) + p_j. \quad (23)$$

Now, consider the provider choices induced by the value repartition \mathbf{V}^* as given in (12). We prove here that this repartition is a user equilibrium: no user has an interest to change providers. Take a provider $i \in \mathcal{I}$. Then for $x \in [0, 1]$,

$$\begin{aligned} \mathcal{V}(x) \in (V_{[i-1]}^*, V_{[i]}^*) &\Leftrightarrow V_{[i-1]}^* < \int_{y=0}^x q(y) dy < V_{[i]}^* \\ &\Leftrightarrow \mathcal{N}(V_{[i-1]}^*) < x < \mathcal{N}(V_{[i]}^*) \\ &\Rightarrow \underline{v}_i^* \leq q(x) \leq \bar{v}_i^*. \end{aligned}$$

The last line and (23) imply that the considered user, that is at position x in the population when it is ranked according to valuations, cannot do better than choosing the provider suggested by (12). In other words, each user is satisfied with her current provider choice, i.e., we have a user equilibrium.

We now establish the uniqueness of the equilibrium value repartition \mathbf{V}^* (and thus, of the user equilibrium due to Lemma 2 when all prices are different).

Proposition 2. *Under Assumption A, the value repartition at a user equilibrium necessarily equals $\mathbf{V}^* = \arg \min_{\mathbf{V} \text{ feasible}} \mathcal{L}(\mathbf{V}, \mathbf{p})$. Consequently, there exists a unique value equilibrium value repartition, and the user equilibrium is unique (unless for a zero-measure set of users) when all providers set different prices.*

Proof. We consider a user equilibrium, and prove that the corresponding value repartition $\tilde{\mathbf{V}}$ satisfies the first-order conditions of the convex optimization problem (17), that has been shown to have a unique solution \mathbf{V}^* .

We actually only need to show the counterpart of Relation (18) (resp., (21)) for $j = i-1$ (resp., $j = i+1$), since the other cases immediately follow. From (12), at a user equilibrium we should have for all $x \in (0, 1)$ and all $i, j \in \mathcal{I}$,

$$x \in \left(\mathcal{N}(\tilde{V}_{[i-1]}), \mathcal{N}(\tilde{V}_{[i]}) \right) \Rightarrow q(x)(T_i(\tilde{V}_i) - T_j(\tilde{V}_j)) + p_i - p_j \leq 0. \quad (24)$$

Consider $i \in \mathcal{I}$ such that $\tilde{V}_i > 0$.

– If $j = i-1$, then $T_i(\tilde{V}_i) < T_j(\tilde{V}_j)$. When x tends to $\mathcal{N}(\tilde{V}_{[i-1]})$, (24) yields

$$\underbrace{q(\mathcal{N}(\tilde{V}_{[i-1]}^+))}_{=\mathcal{N}'_r(\tilde{V}_{[i-1]})}(T_i(\tilde{V}_i) - T_j(\tilde{V}_j)) + p_i - p_j \leq 0,$$

which is exactly the counterpart of (18).

– Likewise for $j = i+1$, from (24) for x tending to $\mathcal{N}(\tilde{V}_{[i]})$ we get the counterpart of (21) (using the fact that q is left-continuous)

$$\underbrace{q(\mathcal{N}(\tilde{V}_{[i]}))}_{=\mathcal{N}'_l(\tilde{V}_{[i]})}(T_i(\tilde{V}_i) - T_j(\tilde{V}_j)) + p_i - p_j \leq 0.$$

The repartition $\tilde{\mathbf{V}}$ satisfies the first-order conditions of the convex optimization problem (17) and is feasible, therefore $\tilde{\mathbf{V}} = \mathbf{V}^*$, the unique solution of (17).

The second claim of the proposition is a direct application of Lemma 2.

Note that the uniqueness of the equilibrium value repartition \mathbf{V}^* implies that even when several user equilibria exist, for all users the cost of each provider at equilibrium is unique; the user equilibrium is then said *essentially unique* [2].

Note also that it was not compulsory to aggregate providers with the same price p : at the minimum of $\mathcal{L}(\cdot, \mathbf{p})$ we notice from (14) that the term $\int_0^{V_{\mathcal{I}_p}} T_{\mathcal{I}_p}$ involving the aggregated function coincides with $\sum_{i \in \mathcal{I}_p} \int_{y=0}^{x_i} T_i(y) dy$. Therefore, the equilibrium value distribution \mathbf{V}^* can directly be found by solving the potential minimization problem (17). Nevertheless, the interpretation of the potential is changed, since the terms $\mathcal{N}(V_{[i]}) - \mathcal{N}(V_{[i-1]})$ of (15) do not necessarily correspond anymore to provider i 's market share.

The next result shows some continuity properties of the user equilibrium.

Proposition 3. *The (unique) equilibrium value repartition \mathbf{V}^* is continuous in the price profile. Moreover, at any price profile such that all prices are different, the provider market shares are continuous in the price profile.*

Proof. Remark that $\mathcal{L}(\mathbf{V}, \mathbf{p})$ is jointly continuous in \mathbf{V} and \mathbf{p} , and that the set of feasible value repartitions is compact. Therefore, from the Theorem of the Maximum (see [3]) applied to the minimization problem (17), the set of equilibrium distributions is upper hemicontinuous in \mathbf{p} . It is actually continuous due to the uniqueness of the equilibrium distribution \mathbf{V}^* .

For a given price profile $\bar{\mathbf{p}}$ where all prices differ, the strict order of prices is maintained within a vicinity of $\bar{\mathbf{p}}$, where the market share of provider i is $\mathcal{N}(V_{[i]}^*) - \mathcal{N}(V_{[i-1]}^*)$, which is jointly continuous in \mathbf{V} and \mathbf{p} since \mathcal{N} is continuous.

Note that while the equilibrium value repartition \mathbf{V}^* is continuous for all price profiles, that is not the case of provider market shares. Indeed, market shares $(\theta_i)_{i \in \mathcal{I}}$ strongly depend on the *order* of prices through the expression $\mathcal{N}(V_{[i]}^*) - \mathcal{N}(V_{[i-1]}^*)$, that holds when prices are sorted in an increasing order. Since \mathcal{N} is a concave function, then the market share of a provider may drastically decrease when a slight price modification changes his position from k to $k + 1$ in the price ranking. This effect is more prominent when \mathcal{N} is more concave, i.e., when user valuations are heterogeneous.

3.4 Price of Anarchy of the user game

In non-cooperative games, the Price of Anarchy measures the loss of efficiency due to user selfishness [16]. This metric is usually defined as the worst-case ratio of the total cost at an equilibrium to the minimal feasible total cost, and has been extensively studied in the last years [7,24,25,26]. The results closest to the one presented in this subsection come from [4]: the authors consider weighted congestion games, where the cost experienced by each user would correspond to the situation where all prices are set to 0 in our model. Then the authors prove that the upper bound for the Price of Anarchy is not greater for the weighted game than for its unweighted counterpart. We actually establish the same kind of result for any value of the provider price profile \mathbf{p} , except that in our case the total user cost (sum of the costs perceived by all users) for any feasible user valuation repartition \mathbf{V} is

$$C_u := \sum_{i \in \mathcal{I}} (V_i T_i(V_i) + p_i (\mathcal{N}(V_{[i]}) - \mathcal{N}(V_{[i-1]}))). \quad (25)$$

Proposition 4. *Assume that the risk functions $(T_i)_{i \in \mathcal{I}}$ belong to a family \mathcal{C} , and define as in [7] the quantity $\beta(\mathcal{C}) := \sup_{T \in \mathcal{C}, (x,y) \in [0, V_{\text{tot}}]^2} \frac{x(T(y) - T(x))}{yT(y)}$. Then for any nonnegative price profile \mathbf{p} ,*

$$\frac{C_u^*}{C_u^{\text{opt}}} \leq \frac{1}{1 - \beta(\mathcal{C})}, \quad (26)$$

where C_u^* (resp. C_u^{opt}) is the total user cost at the user equilibrium (resp. the minimum total user cost) for the price profile \mathbf{p} .

Proof. We apply a variational inequality that is satisfied by the user equilibrium value repartition \mathbf{V}^* , and that directly stems from the fact that users only select their preferred provider: for any feasible value repartition \mathbf{V} , we have

$$\sum_{i \in \mathcal{I}} \left(V_i^* T_i(V_i^*) + p_i (\mathcal{N}(V_{[i]}^*) - \mathcal{N}(V_{[i-1]}^*)) \right) \leq \sum_{i \in \mathcal{I}} \left(V_i T_i(V_i) + p_i (\mathcal{N}(V_{[i]}) - \mathcal{N}(V_{[i-1]})) \right).$$

This yields

$$C_u^* \leq C_u + \sum_{i \in \mathcal{I}} V_i (T_i(V_i^*) - T_i(V_i)) \leq C_u + \beta(\mathcal{C}) \sum_{i \in \mathcal{I}} V_i^* T_i(V_i^*) \leq C_u + \beta(\mathcal{C}) C_u^*,$$

which establishes the proposition.

It is shown in [7] that if \mathcal{C} is the set of affine risk functions the bound $1/(1 - \beta(\mathcal{C}))$ equals $4/3$, resulting in a moderate loss of efficiency due to selfishness. Values 1.626 and 1.896 have also been found respectively for the sets of quadratic and cubic cost risk functions, and $\beta(\mathcal{C}) = d/(d+1)^{1+1/d}$ for the set of polynomials of degree at most d with non-negative coefficients.

As in [4], we find that the introduction of weights among user congestion effects (and here, in addition, among user perceived costs) does not worsen the Price of Anarchy. The bound given in Proposition 4 can indeed be attained, when \mathcal{C} includes the constant functions, with a simple 2-provider instance with prices set to zero, and all users having the same weight.

4 Pricing decisions of security providers

We now focus on the decisions made by security providers when choosing their charging price. We consider that providers are able to anticipate user reactions when fixing their prices. We then have a two-stage game, where at a first step (larger time scale) providers compete on setting their prices so as to maximize revenue, considering that at a second step (smaller time scale) users selfishly select their provider.

The utility of provider i is given by his revenue $r_i := p_i \theta_i$, where θ_i is the market share of provider i . When all providers propose different prices and providers are ranked such that $p_1 < p_2 < \dots < p_I$, from Proposition 2 the user equilibrium exists and is unique, and we simply have $\theta_i = \mathcal{N}(V_{[i]}^*) - \mathcal{N}(V_{[i-1]}^*)$, where \mathbf{V}^* is the equilibrium value repartition. On the other hand, if several providers in a set \mathcal{I}_p propose the same price p , then the equilibrium valuation repartition \mathbf{V}^* is unique, but the user equilibrium choices need not be unique: indeed, any price-monotone user repartition consistent with \mathbf{V}^* is a user equilibrium, and several such repartitions may exist. For those special cases, a reasonable assumption could be that users make their provider choice independently of their valuation when they have several equally preferred providers. As a result, the total market share of providers in \mathcal{I}_p would be split among them proportionally to the data value V_i^* that they attract, yielding

$$\theta_i = \frac{V_i^*}{\sum_{j: p_j = p_i} V_j^*} \left(\mathcal{N} \left(\sum_{j: p_j \leq p_i} V_j^* \right) - \mathcal{N} \left(\sum_{j: p_j < p_i} V_j^* \right) \right).$$

We now establish that, when there exists a bounded price alternative, the revenue of any provider tends to zero if he increases his price to infinity. In practice, such a bounded-price option always exists, even if it has bad performance: one just needs to consider any free security possibility. Therefore, prices will not be arbitrarily high when providers want to maximize revenue.

Proposition 5. *Assume that there exists a provider i_0 with price $p_{i_0} \leq \bar{p}_{i_0} < \infty$. Then for any provider $j \neq i_0$, the revenue $r_j = p_j \theta_j$ tends to 0 when $p_j \rightarrow \infty$.*

Proof. Let us consider a user with valuation v , for whom provider j is among the favorite providers. In particular, that user prefers j over i_0 , thus at a user equilibrium we have

$$v(T_{i_0}(V_{i_0}) - T_j(V_j)) \geq p_j - p_{i_0} \geq p_j - \bar{p}_{i_0}. \quad (27)$$

Therefore if $p_j > \bar{p}_{i_0}$ then $T_j(V_j) < T_{i_0}(V_{i_0})$ and

$$v \geq \frac{p_j - \bar{p}_{i_0}}{T_{i_0}(V_{i_0}) - T_j(V_j)} \geq \frac{p_j - \bar{p}_{i_0}}{T_{i_0}(V_{\text{tot}})} := v_{\min}.$$

The revenue $r_j = p_j \theta_j$ of provider j can then be upper bounded:

$$r_j \leq p_j \int_{v=v_{\min}}^{+\infty} dF(v) = T_{i_0}(V_{\text{tot}}) \underbrace{\frac{p_j - \bar{p}_{i_0}}{T_{i_0}(V_{\text{tot}})} \int_{v=\frac{p_j - \bar{p}_{i_0}}{T_{i_0}(V_{\text{tot}})}}^{+\infty} dF(v)}_{\xrightarrow[p_j \rightarrow \infty]{} 0} + \bar{p}_{i_0} \underbrace{\int_{v=\frac{p_j - \bar{p}_{i_0}}{T_{i_0}(V_{\text{tot}})}}^{+\infty} dF(v)}_{\xrightarrow[p_j \rightarrow \infty]{} 0},$$

where the two terms tend to zero since $\int_0^\infty v dF(v) = V_{\text{tot}} < \infty$.

4.1 Licensed versus free security provider

We consider here a simple situation with two providers, but only one trying to maximize his profit through subscription benefits. The other provider (or, more likely, a community of developers) offers the security service for free.

Denote by 0 and 1 the freeware provider and the licensed provider, respectively. From Proposition 1, there exists a unique value repartition $(V_0(p), V_{\text{tot}} - V_0(p))$ at the user equilibrium, for any price p set by provider 1. Likewise, for any $p > 0$ the equilibrium market share of provider 1 is unique and given by $\theta_1 = 1 - \mathcal{N}(V_0(p))$; the profit maximization problem of provider 1 can therefore be written as

$$\max_{p \geq 0} p \cdot (1 - \mathcal{N}(V_0(p))). \quad (28)$$

Note that provider 1 gets demand as soon as his price is strictly below $\sup(S_v) \times T_0(V_{\text{tot}})$, therefore by choosing $p \in (0, \sup(S_v) T_0(V_{\text{tot}}))$ he can ensure a positive revenue. Therefore from Propositions 3 and 5, the provider revenue optimization problem (28) has a solution, that is finite.

Corollary 1. *When a profit-oriented provider faces only a competitor with null price, then under Assumption A there exists a finite price $\bar{p} > 0$ that maximizes his revenue, whose maximum value is strictly positive.*

4.2 Competition among providers: the risk of price war

Competitive contexts where providers play on price to attract customers often lead to *price war* situations, i.e., situations where each provider has an interest in decreasing his price below the price of his competitor. The outcome then corresponds to providers making no profit, and possibly not surviving.

With the model presented in this paper, not all demand goes to the cheapest provider because of the congestion effect due to attackers' behavior. However, some threshold effect still exist, as illustrated by the non-continuity of provider market shares when provider prices cross each other.

Let us for example consider two identical profit-oriented providers and a free alternative. Due to the symmetry of the game, one would expect a situation where both providers set their price to the same level, say $p > 0$. As a result, again from symmetry arguments both providers would be chosen by users to protect, at equilibrium, the same value $V_1^* = V_2^* := V^*$ of data each, while the free provider covers a total data value V_0 . Then, if provider 1 sets his price to $p - \varepsilon$ for a small $\varepsilon > 0$, the market share repartition is such that when $\varepsilon \rightarrow 0$,

$$\begin{aligned}\theta_0 &= \mathcal{N}(V_0^*), \\ \theta_1 &= \mathcal{N}(V_0^* + V^*) - \mathcal{N}(V_0^*), \\ \theta_2 &= \mathcal{N}(V_0^* + 2V^*) - \mathcal{N}(V_0^* + V^*).\end{aligned}$$

When users choosing provider 1 or 2 are not all homogeneous in their data valuations (which is for example the case if the valuation distribution F admits a density), then $\theta_1 > \theta_2$. In other words, provider 1 strictly improves his market share (and thus his revenue) by setting his price just below the price of his competitor. But provider 2 can make the exact same reasoning, resulting in a price war situation.

Consequently, there can be no symmetric Nash equilibrium (i.e., a price profile such that no provider can improve his revenue by a unilateral change) where $p_1 = p_2 > 0$, despite the symmetry of the pricing game. Furthermore, the price profile where all prices are set to 0 is not an equilibrium either: both providers would get no revenue, which each one could strictly improve by a small price increase as stated in Corollary 1.

Remark that this reasoning does not rule out the possibility of the pricing game having a (non-symmetric) Nash equilibrium, however we cannot always guarantee that such an equilibrium exists. An explanation to the existence of stable price profiles can nevertheless still be found from game-theoretic arguments, since the pricing game among providers is not played only once but repeatedly over time. When considering *repeated games* (i.e., where players take into account not only their current payoff but also a discounted sum of the future ones), the set of Nash equilibria is indeed much larger than for their one-shot counterpart, as evidenced by the *Folk theorem* [23]. The stability of prices can then stem from the threat of being sanctioned by competitors for an (immediate-profit) price change.

We illustrate those results when user valuations are distributed according to an exponential law with average value $1/\lambda = 10$ monetary units. Such a distribution models an unbounded continuum of valuations among the population, where a large majority of users have limited valuations, but there exist few people with extremely high value data to protect. The risk function considered in our numerical computations is $R_i(x) = 1 - e^{-x}$ for each provider i , which models the fact that systems with no valuable data are not targeted while successful systems are very likely to attract attacks.

In our numerical illustration, we consider here three providers: a provider 0 with performance parameter $\pi_0 = 0.05$, that is always free: $p_0 = 0$; and two profit-oriented providers, namely 1 and 2, with respective performance values $\pi_1 = 0.01$ and $\pi_2 = 0,005$. Providers protected data values and market shares are shown in Figures 2 and 3, and the revenue of provider 2 is displayed in Figure 4. The curves illustrate the continuity results of Proposition 3. Interestingly, we

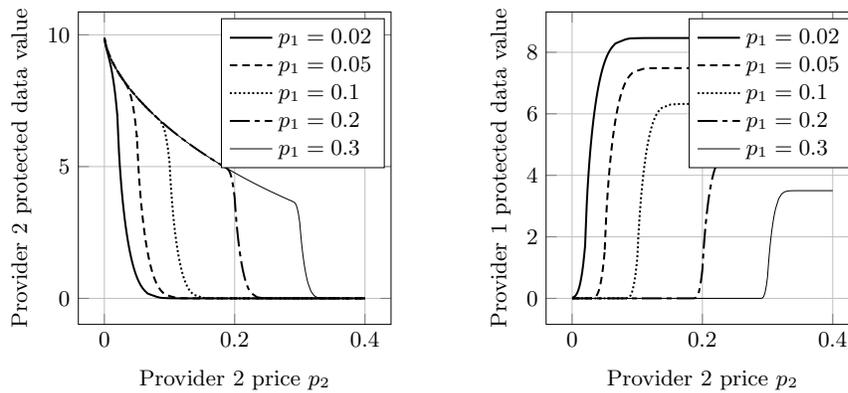


Fig. 2. Protected data values when provider 2 varies his price.

remark in Figure 4 that despite the discontinuity in revenue when prices cross each other, provider 2 actually has a revenue-maximizing price $p_2^{\text{BR}}(p_1)$ strictly below the price of his competitor. That last figure shows the price war situation: if providers engage in successive best-reply price adaptations to the competition, then prices tend to very low values, which jeopardizes the viability of security providers. However, a situation with strictly positive prices from both providers could be stable in a repeated game context. Consider a price profile (p_1, p_2) such that each provider obtains at least what he could obtain with an aggressive competitor (i.e., a competitor that tries to minimize the provider revenue); when providers value the future almost as much as the present (i.e., when the discount factor that relates current prices to future prices is close to 1), that price profile can be maintained as a subgame-perfect equilibrium of the repeated game [9].

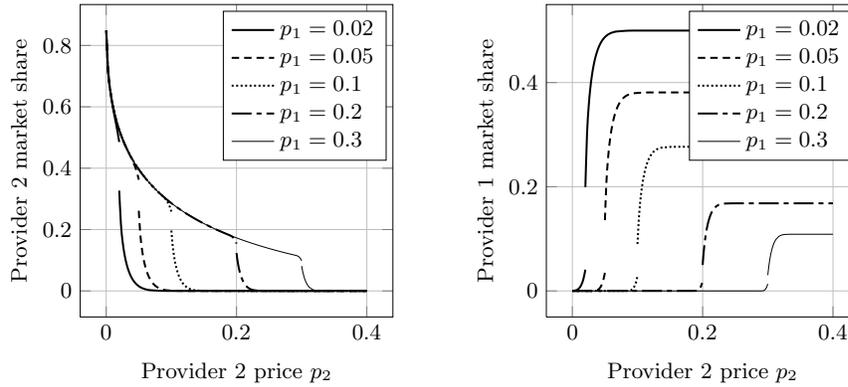


Fig. 3. Market shares when provider 2 varies his price.

5 Conclusions

The model introduced in this paper takes into account the attractiveness that successful security systems represent to profit-minded attackers. This constitutes a negative externality among users: their (selfish) security choices then form a noncooperative congestion game. We have considered heterogeneity among user valuations for data protection, which affects both the externality level and the user cost functions. The corresponding game is therefore a weighted congestion game with user-specific payoffs. We have studied that game for the case of a continuum of infinitesimal users, and have proved that it admits a potential and therefore an equilibrium, that is unique when providers submit different prices.

The study of the user selection game has helped us understand the interaction among security providers, who have to attract customers but are then subject to quality degradation due to more attacks, hence a trade-off. Our analysis shows that providers will keep their prices low, and that competition may lead to price war situations, unless providers consider long-term repeated interactions.

Future work can focus on the information asymmetry and uncertainty among actors: we have studied the interactions in a complete information context, whereas users may not have a perfect knowledge of the performance level of the different providers, or of their total protected data value. Likewise, attackers can only estimate the potential gain from targeting a given system.

Another interesting direction for future research concerns the investment strategies that security providers should implement: indeed, improving the protection performance has a cost, that has to be compensated by the extra revenue due to user subscription decisions. While there exist references for this kind of problem when users are homogeneous [13], the case when users have different weights deserves further attention.

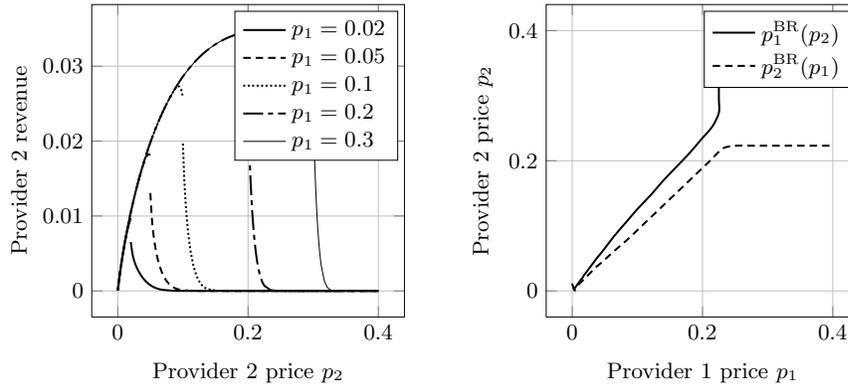


Fig. 4. Revenue of provider 2 ($\pi_2 = 0.005$) facing provider 1 ($\pi_1 = 0.01$) and free provider 0 ($\pi_0 = 0.05$) (left), and best-reply functions of providers 1 and 2 (right).

Acknowledgements

The authors acknowledge the support of European initiative COST IS0605, Econ@tel. Part of this work has been supported by the Austrian government and the city of Vienna in the framework of the COMET competence center program, and by the French research agency through the CAPTURES project.

References

1. Alpcan, T., Başar, T.: Network Security: A Decision and Game Theoretic Approach. Cambridge University Press (2011)
2. Beckmann, M., McGuire, C.B., Winsten, C.B.: Studies in the economics of transportation. Yale University Press, New Haven, Connecticut (1956)
3. Berge, C.: Espaces topologiques. Fonctions multivoques, Collection Universitaire de Mathématiques, vol. III. Dunod, Paris (1959)
4. Bhawalkar, K., Gairing, M., Roughgarden, T.: Weighted congestion games: Price of anarchy, universal worst-case examples, and tightness. In: Proc. of 18th European Symposium on Algorithms (ESA). Liverpool, UK (2010)
5. Bistarelli, S., Dall'Aglio, M., Peretti, P.: Strategic games on defense trees. In: Proc. of 4th Intl Workshop on Formal Aspects in Security and Trust (FAST'06), LNCS 4691. pp. 1–15. Hamilton, Ontario, Canada (2006)
6. Bohacek, N., Hespanha, J.P., Lee, J., Lim, C., Obraczka, K.: Game theoretic stochastic routing for fault tolerance and security in computer networks. IEEE Transactions on Parallel and Distributed Systems 18(9), 1227–1240 (2007)
7. Correa, J.R., Schulz, A.S., Stier-Moses, N.: A geometric approach to the price of anarchy in nonatomic congestion games. Games and Economic Behavior 64(2), 457–469 (2008)

8. Cremonini, M., Nizovtsev, D.: Understanding and influencing attackers' decisions: Implications for security investment strategies. In: Proc. of 5th Workshop on the Economics of Information Security (WEIS). Cambridge, UK (2006)
9. Fudenberg, D., Maskin, E.: The folk theorem in repeated games with discounting or with incomplete information. *Econometrica* 54(3), 533–554 (1986)
10. Fudenberg, D., Tirole, J.: *Game Theory*. MIT Press (1991)
11. Ganesh, A., Gunawardena, D., Jey, P., Massoulié, L., Scott, J.: Efficient quarantining of scanning worms: Optimal detection and co-ordination. In: Proc. of IEEE INFOCOM 2006. Barcelona, Spain (2006)
12. Jiang, L., Anantharam, V., Walrand, J.: Efficiency of selfish investments in network security. In: Proc. of 3rd Workshop on the Economics of Networks, Systems, and Computation. Seattle, WA, USA (2008)
13. Johari, R., Weintraub, G.Y., Van Roy, B.: Investment and market structure in industries with congestion. *Operations Research* 58(5), 1303–1317 (2010)
14. Karakostas, G., Kolliopoulos, S.G.: Edge pricing of multicommodity networks for heterogeneous selfish users. In: Proc. of FOCS, pp. 268–276 (2004)
15. Kodialam, M., Lakshman, T.V.: Detecting network intrusions via sampling: A game theoretic approach. In: Proc. of IEEE INFOCOM. San Francisco, CA, USA (2003)
16. Koutsoupias, E., Papadimitriou, C.: Worst-case equilibria. In: Proc. of 16th Annual Symposium on Theoretical Aspects of Computer Science (STACS 1999). LNCS 1563, pp. 404–413 (1999)
17. Lelarge, M., Bolot, J.: Economic incentives to increase security in the internet: The case for insurance. In: Proc. of IEEE INFOCOM. Rio de Janeiro, Brazil (2009)
18. Maillé, P., Reichl, P., Tuffin, B.: Of threats and costs: A game-theoretic approach to security risk management. In: Gülpınar, N., Harrison, P., Rüstem, B. (eds.) *Performance Models and Risk Management in Communication Systems*. Springer (2010)
19. Mavronicolas, M., Milchtaich, I., Monien, B., Tiemann, K.: Congestion games with player-specific constants. In: Proc. of 32nd International Symposium on Mathematical Foundations of Computer Science (MFCS). LNCS 4708, pp. 633–644. Springer-Verlag, Český Krumlov, Czech Republic (Aug 2007)
20. Milchtaich, I.: Congestion games with player-specific payoff functions. *Games and Economic Behavior* 13(1), 111–124 (1996)
21. Milchtaich, I.: Weighted congestion games with separable preferences. *Games and Economic Behavior* 67(2), 750–757 (Nov 2009), <http://www.sciencedirect.com/science/article/B6WFW-4W0SK2D-1/2/63c8ccc38f57d26ba8db4ed12c1596d3>
22. Monderer, D., Shapley, L.S.: Potential games. *Games and Economic Behaviour* 14, 124–143 (1996)
23. Osborne, M.J., Rubinstein, A.: *A Course in Game Theory*. MIT Press (1994)
24. Perakis, G.: The “Price of Anarchy” under nonlinear and asymmetric costs. *Mathematics of Operations Research* 32(3), 614–628 (2007)
25. Roughgarden, T.: *Selfish Routing and the Price of Anarchy*. MIT Press (2005)
26. Roughgarden, T., Tardos, E.: Bounding the inefficiency of equilibria in nonatomic congestion games. *Games and Economic Behavior* 47(2), 389–403 (2004)
27. Sandholm, W.H.: Potential games with continuous player sets. *Journal of Economic Theory* 97(1), 81–108 (2001)
28. Sandholm, W.H.: Large population potential games. *Journal of Economic Theory* 144(4), 1710–1725 (2009)
29. Schmeidler, D.: Equilibrium points of nonatomic games. *Journal of Statistical Physics* 7(4), 295–300 (1973)

Managing a Peer-to-Peer Data Storage System in a Selfish Society

Patrick Maillé, László Toka

Abstract—We compare two possible mechanisms to manage a peer-to-peer storage system, where participants can store data online on the disks of peers in order to increase data availability and accessibility. Due to the lack of incentives for peers to contribute to the service, we suggest that either each peer's use of the service be limited to her contribution level (symmetric schemes), or that storage space be bought from and sold to peers by a system operator that seeks to maximize profit. Using a noncooperative game model to take into account user selfishness, we study those mechanisms with respect to the social welfare performance measure, and give necessary and sufficient conditions for one scheme to socially outperform the other.

Index Terms—Peer-to-peer networks, game theory, incentives, pricing.

I. INTRODUCTION

THE “digital society” that has been soaring since the creation of the Internet implies that all kinds of digital documents are now likely to be created, accessed, and modified from several types of devices. Therefore, an appropriate system for storing the data of a user should offer various services, such as versioning, ease of access, protection against device failures, and short transfer time to a given device.

In that context, the possibility of storing data online appears as a promising solution. Indeed, having access to the Internet becomes easier and easier, with the multiplication of WiFi hotspots, the development of WiMAX and third generation wireless networks, and the appearance of other access modes, such as multi-hop networks that work in an ad-hoc fashion to reach an access point. Let us also highlight the high rise of available transmission rates in access networks, which renders transfer times reasonable, even for large files. Finally, online storage systems are able to cope with document versioning, and to protect data not only against user device failures but also against disk failures, through the use of data replicates stored on different disks.

For those reasons, many companies now propose online data storage services, most of them offering a given storage capacity (between 2 and 25 gigabytes) for free, with the possibility of extending that quota to a higher value for a fixed price per year (the price per year per gigabyte being of the

order of 1\$). However, while creating such a storage service implies owning huge memory capacities and affording the associated energy and warehouse costs, one can imagine using the smaller but numerous storage spaces of the service users themselves, as is done in peer-to-peer file sharing systems.

In a peer-to-peer storage system, the participants are at the same time the providers and the users of the service: each participant offers some memory capacity (possibly from multiple locations in the network: part of her disk space at home, storing device devoted to the service, ...) to provide the service to the others, and benefits from storing her own data onto the system. The added value of the service then comes from the protection against failures provided by the system, from the ease of data access, from the versioning management that may be included, and from the difference in the amount of data stored into the system versus offered to the service.

An online storage service is valuable only if data are available: therefore to cope with disk failures and with participants disconnecting their disk from the system, data replicates must be spread over several (sufficiently reliable) peers to guarantee that data are not lost and are almost always available; the data replication rate then depends on the reliability of the participants. To work properly, a peer-to-peer storage network therefore needs that participants offer a sufficient part of their disk space to the system, and remain online often enough. However, both of those requirements imply costs (or at least constraints) for participants, who may be reluctant to devote some of their storage capacity to the system instead of using it for their own needs.

In this paper, we consider that users behave selfishly, i.e. are only sensitive to the quality of service they experience, regardless of the effects of their actions on the other users. The framework of *noncooperative Game Theory* [1] is therefore particularly well-suited to study the interactions among peers. For a peer-to-peer storage system, it is clear that without any reward for contributing participants, selfishly behaving users will only benefit from the service without providing any part of it¹. In other words, the only Nash equilibrium of the noncooperative game is the situation where the system actually does not exist due to the lack of offering peers.

The work presented in this paper focuses on the incentives that can be used to make participants contribute to the system, i.e. the changes that can be brought to the game to modify its Nash equilibria. While the economic aspects of peer-to-peer file sharing networks have already been extensively studied (see [3], [4], [5], [6] and references therein), there are to our

P. Maillé is with the Department of Network, Security and Multimedia, TELECOM Bretagne, 2, rue de la Châtaigneraie, 35576 Cesson-Sévigné Cedex, FRANCE. E-mail: patrick.maille@telecom-bretagne.eu

L. Toka is with the Department of Network and Security, EURECOM, 2229 route des Crêtes, 06560 Sophia-Antipolis Cedex, FRANCE and with the Department of Telecommunications and Media Informatics, Budapest University of Technology and Economics, 2 Magyar Tudósok krt., 1117 Budapest, HUNGARY. E-mail: toka@eurecom.fr

This work has been partially supported by the TELECOM Institute project DisPairSe.

Manuscript received July 31, 2007; revised March 5, 2008.

¹Such a behavior, called *free-riding* [2], also appears in peer-to-peer file sharing networks, and is problematic for the survival of those altruism-based networks.

knowledge no references on the economics of peer-to-peer storage networks. Now, the economic models developed for peer-to-peer file sharing systems do not apply to peer-to-peer storage services: in file sharing systems, when a peer provides some files to the community, she adds value to the system for all users since they all can access the data she proposes; in that sense the resource offered to the system is a public good. On the contrary, in a peer-to-peer storage system the memory space offered by a peer is a private good: it can be shared among different users but each part is then devoted to only one user. Therefore the economic implications of those systems are necessarily different.

The existing literature on peer-to-peer storage systems mainly focuses on security, reliability and technical feasibility issues [7], [8], [9], whereas the incentive aspect received little attention. Only solutions that do not imply financial transactions are considered in current works, therefore to create some incentives to participate, the counter payment for providing service is usually the service in question as well. This approach finally leads to a scheme where every peer should contribute to the system in terms of service at least as much as she benefits from others [10], [11]. We call such a mechanism imposing the contribution of each peer to equal her use of the system a *symmetric* scheme.

In this paper, we also investigate solutions based on monetary exchanges: users can “buy” storage space for a fixed unit price, and “sell” their own memory space to the system at another unit price. It is known from economic theory that when those unit prices are fixed by the supply and demand curves (as in a perfect market [12]), then user selfish choices lead to a socially efficient situation. However, it is more likely here that the system be managed by a profit-maximizing entity that fixes prices so as to maximize revenue. That entity then acts as the leader of a Stackelberg game [1].

The main question addressed in this paper is whether it is socially better to impose a symmetric scheme or to let a profit-maximizing monopoly set prices. The performance measure we consider is social welfare, i.e. the total value that the system has for all participants. Under some assumptions on the peers utility functions, we derive a necessary and sufficient condition for symmetry-based systems to outperform revenue-oriented management. We obtain that user heterogeneity tends to favor pricing-based schemes that are more flexible, and above a given user heterogeneity threshold even a monopoly-managed system will be socially better than a system imposing symmetry. The results presented here are a generalization of our preliminary work [13] that did not consider incentives to stay online and where the only source of heterogeneity came from the price sensitivities.

This paper is organized as follows. Section II introduces the model we consider for user preferences, and for the two incentive mechanisms studied in this paper, namely symmetry-based and profit oriented price-based schemes. In Section III we define the social welfare performance measure and compute its value for those two types of schemes. We compare them in Section IV to determine the management scheme that is best suited to the society, and present our conclusions in Section V.

II. MODEL

A. Content availability management and associated costs

In a peer-to-peer storage system the availability of the stored data is considered as the most important factor in user’s appreciation. As the storage disks are users’ property, there are no direct means to guarantee that a given user disk storing a specific file will be online 100% of the time. To ensure data availability, the system can introduce several tools, such as data replication and coding [14]. We suppose here that the system detecting that a peer has gone offline triggers a recovery of the data stored in that peer from the replicas in the system, and a new storage of those data into other peers. Likewise, when a peer comes back online, then new data will be transferred into her offered storage space, independently what and whose data she was storing before. Such a scheme is purely reactive (actions are taken when a user departure is detected). One could also imagine using proactive approaches, or a combination of both, to smoothe the incurred traffic [15].

This data protection mechanism implies data transfers, and therefore nonmonetary costs due to resource consumption (CPU, bandwidth utilization, etc.). A peer i is concerned by those data transfers in two situations: when she comes back online after an offline period (new data load), and when other peers enter and leave the system (upload traffic if user i stores replicates of the leaving user’s data, download traffic when user i has to store more data). The mean data transfer associated to the first situation is thus proportional to the amount of capacity C_i she offers to the system, and to the mean number of online-offline cycles per unit of time: denoting by t_i^{on} (resp. t_i^{off}) the mean duration of online (resp. offline) periods of user i , the corresponding mean amount of data transferred is then proportional to $C_i/(t_i^{\text{on}} + t_i^{\text{off}})$. The mean amount of data transferred to and from user i per unit of time in the second situation is proportional to the weighted (by the offered capacity) mean $\bar{\mu}$ of peer status changes per unit of time². This term appears only at those peers who offer storage space (proportionally to their offered capacity since the probability that user i be concerned by a peer’s departure is proportional to C_i), and only during the time they are online (it is therefore also proportional to the mean availability of user i , $\pi_i := t_i^{\text{on}}/(t_i^{\text{on}} + t_i^{\text{off}})$).

Consequently, the transfer cost perceived by user i for offering capacity C_i with the mean availability π_i expresses $C_i\pi_i(\delta_i/t_i^{\text{on}} + \gamma_i\bar{\mu})$, where δ_i and γ_i are parameters that reflect the user characteristics such as sensitivity, access bandwidth, or hardware profile.

B. User preferences

We describe the preferences of a user i in the user set denoted by \mathcal{I} by a utility function, that reflects the benefit of using the service by storing an amount C_i^s of data in the system, the cost of offering storage space $C_i^o := \pi_i C_i$ for

²Actually peer i should only be sensitive to the status change rate of all other peers but hers. However we consider here a system with a very large number of users, so that taking the mean of the change rates over all participants but one is equivalent to considering all participants.

other users, and the monetary transactions, if any. We suggest to use a separable additive function.

Definition 1: The utility U_i of a user $i \in \mathcal{I}$ is of the form

$$U_i(C_i^s, C_i, t_i^{\text{on}}, t_i^{\text{off}}, \epsilon_i) = V_i(C_i^s) - \underbrace{O_i(C_i \pi_i) - C_i \pi_i (\delta_i / t_i^{\text{on}} + \gamma_i \bar{\mu})}_{:= P_i(C_i, t_i^{\text{on}}, t_i^{\text{off}})} - \epsilon_i, \quad (1)$$

where

- $V_i(C_i^s)$ is user i 's valuation of the storage service, i.e. the price she is willing to pay to store an amount C_i^s of data in the system³. We assume that $V_i(\cdot)$ is positive, continuously differentiable, increasing and concave in its argument, and that $V_i(0) = 0$ (no service yields no value).

- $P_i(C_i, t_i^{\text{on}}, t_i^{\text{off}})$ is the overall non-monetary cost of user i for offering capacity C_i to the system with mean online and offline durations respectively equal to t_i^{on} and t_i^{off} , i.e. with availability $\pi_i = \frac{t_i^{\text{on}}}{t_i^{\text{on}} + t_i^{\text{off}}}$. It consists of two distinct costs:

- an opportunity cost $O_i(C_i \pi_i)$ of offering storage capacity for other users (during online periods) instead of using it for her own needs⁴, where $O_i(\cdot)$ is assumed positive, continuously differentiable, increasing and strictly convex, and such that $O_i(0) = 0$ (no contribution brings no cost);
- data transfer costs $C_i \pi_i (\delta_i / t_i^{\text{on}} + \gamma_i \bar{\mu})$ due to the data protection mechanism implemented by the system as described in the previous subsection.
- ϵ_i is the monetary price paid by user i . This term is 0 in case of a symmetric scheme, and otherwise equals the price difference between the charge for storing her data into the system and the remuneration for offering her disk space.

Remark that we implicitly say that the storage space necessary to safely store some data in the system equals the size of those data. This is done without loss of generality, taking into account the redundancy factor r added by the system in users' cost function: a user considered to offer space to store an amount C_i of data actually devotes more of her disk space (rC_i) to the service. Likewise, prices are then per unit of "protected data".

C. Incentive schemes for cooperation

Users selfishly choose strategies that maximize their utility. We assume here that apart from C_i^s and C_i^o , each user i can also decide about her behavior related to availability. In this subsection, we describe the two types of incentive mechanisms that we intend to compare in this paper. Both schemes may imply the existence of a central authority or clearance service to supervise the peers behavior and/or manage payments: as the model aims to give hints for commercial applications, we do not try to avoid such a centralized system control.

³We assume here that data replication ensures a given availability, so that this availability does not appear in the utility function.

⁴We implicitly assume here that the opportunity cost depends only on the mean capacity offered over time, since during offline periods the user can use the disk space for other purposes than the service.

1) *Symmetric schemes:* We follow here the ideas suggested in the literature for schemes without pricing. As evoked in the introduction, the principle of those schemes is that users are invited to contribute to, at least as much as they take from, the other users. The availability of the peer is therefore checked (e.g. at randomly chosen times) to ensure that $C_i^o = \pi_i C_i$ exceeds the peer's service use C_i^s .

We assume in this paper that this verification is technically feasible. Determining whether and how it can be done remains an active topic of research and is beyond the scope of this paper, since we only focus here on incentives.

2) *Payment-based schemes:* We consider a simple payment-based mechanism where users can "buy" storage space in the system for a unit price p^s (per byte and per unit of time) and "sell" some of their (time-average available) disk capacity for a unit price p^o .

The (possibly negative) amount that user i is charged is then

$$\epsilon_i = p^s C_i^s - p^o C_i^o.$$

In this paper, we assume that prices are set by the system operator so as to maximize her revenue, knowing a priori the reactions of the users. The operator can thus drive the outcome of the game to the most profitable situation for herself, and in this sense, she acts as the leader of a Stackelberg (or leader-follower) game [1]. In a real implementation of the mechanism, the operator may not perfectly know the user reactions, but an iterative tâtonnement of prices can converge to those profit-maximizing prices.

D. User behavior related to availability

In the game we study, a peer $i \in \mathcal{I}$ has four strategic variables, namely her offered C_i and stored C_i^s capacities, and her mean online t_i^{on} and offline t_i^{off} period durations. Equivalently, we can also consider that the four strategic variables are $C_i^s, C_i^o, t_i^{\text{on}}$, and t_i^{off} . From (1), when C_i^s and C_i^o are fixed, the utility of each user is increasing in t_i^{on} , so t_i^{on} will be set by user i to a maximum value. We denote by \bar{t}_i^{on} that maximum value, which is only limited by uncontrolled events (power black-out, accidents, hardware failures, etc) that may force the user off the network.

Notice that this selfish decision is profitable to the whole network: longer online periods mean fewer data protection transfers and therefore smaller costs for the system (the parameter $\bar{\mu}$ in (1) being small). Remark also that since t_i^{off} does not appear in (1), there remain only two decision variables, namely C_i^s and C_i^o (that equals $C_i \bar{t}_i^{\text{on}} / (\bar{t}_i^{\text{on}} + t_i^{\text{off}})$). From now we will therefore write $P_i(C_i^o)$ instead of $P_i(C_i, t_i^{\text{on}}, t_i^{\text{off}})$, and will also use the notation

$$p_i^{\text{min}} := \delta_i / \bar{t}_i^{\text{on}} + \gamma_i \bar{\mu}, \quad (2)$$

so that the transfer costs simply write $C_i^o p_i^{\text{min}}$.

E. User supply and demand functions

Supply and demand functions are classically used in economics [12], and are respectively derived from the valuation of consumers and cost functions of providers. Notice however the

particularity here that peers can be consumers and providers at the same time.

Definition 2: For a user $i \in \mathcal{I}$, we call *supply function* (resp. *demand function*) the function $s_i(\cdot)$ (resp. $d_i(\cdot)$) such that for all $p \in \mathbb{R}_+$,

$$\begin{aligned} s_i(p) &:= \inf\{q \geq 0 : P'_i(q) \geq p\}, \\ d_i(p) &:= \inf\{q \geq 0 : V'_i(q) \leq p\}, \end{aligned}$$

where g' stands for the derivative function of g , and with the convention $\inf \emptyset = +\infty$.

For a given $p \geq 0$, $s_i(p)$ (resp. $d_i(p)$) is the amount of storage capacity that user i would choose to sell (resp. buy) if she were paid (resp. charged) a unit price p for it.

For the sake of simplicity, our main results in the following consider a particular form of supply and demand functions described in the assumption below.

Assumption A: For all $i \in \mathcal{I}$, the supply and demand functions of user i are affine. More precisely, there exist nonnegative values a_i, b_i , and p_i^{\max} such that

$$s_i(p) = a_i[p - p_i^{\min}]^+, \quad (3)$$

$$d_i(p) = b_i[p_i^{\max} - p]^+, \quad (4)$$

where p_i^{\min} is given in (2), $x^+ := \max(0, x)$, and we assume that $\max_i p_i^{\min} < \min_i p_i^{\max}$.

This actually corresponds to quadratic functions for the valuation and opportunity cost (with \wedge denoting the min):

$$O_i(C_i^o) = \frac{1}{a_i} \frac{C_i^{o2}}{2},$$

$$V_i(C_i^s) = \frac{1}{b_i} \left(-\frac{(C_i^s \wedge b_i p_i^{\max})^2}{2} + b_i p_i^{\max} (C_i^s \wedge b_i p_i^{\max}) \right).$$

Under Assumption A, a user i is entirely described by four parameters (see Figure 1):

- two price thresholds, namely p_i^{\min} and p_i^{\max} , that respectively represent the minimum value of the unit price p^o such that user i sells some of her disk space and the maximum value of the unit price p^s such that she buys some storage space,
- two price sensitivities a_i and b_i , that respectively correspond to the increase of sold capacity with the unit price $p^o \geq p_i^{\min}$ and the decrease of bought storage space with the unit price $p^s \leq p_i^{\max}$.

Consequently, the total supply function $S := \sum_{i \in \mathcal{I}} s_i$ is a (piecewise affine) nondecreasing convex function on the interval $[\min_i p_i^{\min}, \max_i p_i^{\min}]$, and is affine on $[\max_i p_i^{\min}, +\infty)$.

Likewise, the total demand function $D := \sum_{i \in \mathcal{I}} d_i$ is nonincreasing, affine on $[0, \min_i p_i^{\max}]$ and convex on $[\min_i p_i^{\max}, \max_i p_i^{\max}]$, as illustrated in Figure 2 displayed in subsection III-C.

III. SOCIAL WELFARE PERFORMANCE OF INCENTIVE MECHANISMS

In this section we introduce the performance measure used in this paper to compare incentive schemes, and study its value for the social optimum and the outcomes of the two incentive schemes that are the object of this paper.

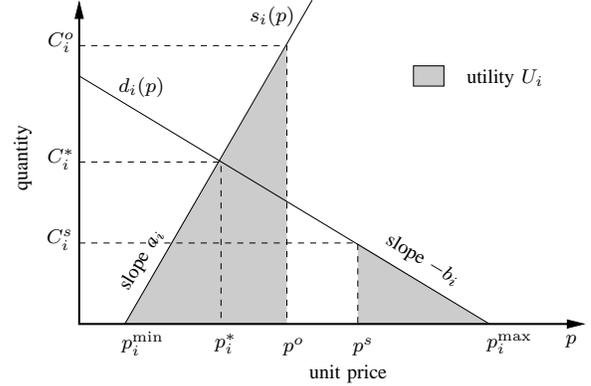


Fig. 1. Reactions to prices and utility of a user $i \in \mathcal{I}$ under Assumption A.

Definition 3: We call *social welfare* (or *welfare*) and denote by W the sum of the utilities of all agents in the system:

$$W := \sum_{i \in \mathcal{I}} V_i(C_i^s) - P_i(C_i^o). \quad (5)$$

Notice that no prices appear in (5), since all system agents are considered, including the operator that receives or gives payments, if any, and whose utility is her revenue. The operator being a member of the society, all money it exchanges with the users stays within the system and therefore does not influence social welfare.

A. Optimal value of social welfare

The optimal situation (in terms of social welfare) that the system can attain corresponds to the maximization problem $\max_{C_i^s, C_i^o} \sum_{i \in \mathcal{I}} V_i(C_i^s) - P_i(C_i^o)$, subject to the feasibility constraints $C_i^o \geq 0$, $C_i^s \geq 0$ for $\forall i$ and $\sum_i C_i^o \geq \sum_i C_i^s$.

This classical convex optimization problem can be solved by the Lagrangian method: if p^* and C^* are the (unique) solutions of the demand-supply equation

$$C^* := \sum_i s_i(p^*) = \sum_i d_i(p^*), \quad (6)$$

then the maximum social welfare is attained when $C_i^s = d_i(p^*)$, and $C_i^o = s_i(p^*)$ for all $i \in \mathcal{I}$. Under Assumption A, the optimal social welfare W^* is then

$$W^* = \frac{1}{2} \sum_i b_i (p_i^{\max 2} - p^{*2}) - a_i (p^{*2} - p_i^{\min 2}). \quad (7)$$

This maximal value W^* as well as the so-called ‘‘shadow price’’ p^* are illustrated in Figure 2 displayed in Subsection III-C. Remark that this optimal situation can be attained with a payment-based scheme where $p^o = p^s = p^*$.

B. Performance of symmetric schemes

Under a symmetry-based management scheme, each user i chooses C_i^o and C_i^s so as to maximize $V_i(C_i^s) - P_i(C_i^o)$, subject to $C_i^o \geq C_i^s$. As $P_i(\cdot)$ is increasing in C_i^s , it is in each user’s best interest to choose a strategy with $C_i^o = C_i^s$. User i then maximizes her utility at the point $C_i^s = C_i^o = C_i^*$

where $V'_i(C_i^*) = P'_i(C_i^*)$, as illustrated in Figure 1. Under Assumption A, this corresponds to every user “exchanging” capacity $C_i^* = \frac{a_i b_i}{a_i + b_i} (p_i^{\max} - p_i^{\min})$ at the *virtual unit price* $p_i^* := \frac{a_i p_i^{\min} + b_i p_i^{\max}}{a_i + b_i}$. Compared to the socially optimal situation, each and every user loses $\frac{1}{2}(p^* - p_i^*)^2(a_i + b_i)$ of utility if $\max_i p_i^{\min} \leq p^* \leq \min_i p_i^{\max}$. In that case, the welfare loss of the system is

$$W^* - W_{sym} = \sum_i \frac{a_i + b_i}{2} (p^* - p_i^*)^2. \quad (8)$$

Remark that $p^* = \frac{\sum_i (a_i + b_i) p_i^*}{\sum_i (a_i + b_i)}$ is then the weighted mean of p_i^* , therefore the loss of welfare only depends on the heterogeneity of users’ p_i^* . In particular, in the case when all users have the same p_i^* , then symmetric management schemes maximize social welfare.

C. Performance of profit-oriented pricing schemes

We now study a pricing mechanism where the system operator strives to extract the maximum profit out of the business by playing on prices p^s and p^o . Knowing that each user i will sell $s_i(p^o)$ and buy $d_i(p^s)$, the operator faces the following maximization problem.

$$\max_{p^s, p^o} \left(p^s \sum_i d_i(p^s) - p^o \sum_i s_i(p^o) \right), \quad (9)$$

subject to $p^s \geq 0$, $p^o \geq 0$ and the feasibility constraint $\sum_i s_i(p^o) \geq \sum_i d_i(p^s)$.

Let us examine the best choices for such a profit-driven monopoly. Figure 2 plots two curves: the total supply $S = \sum_i s_i$ and the total demand $D = \sum_i d_i$ as functions of the unit price p . First remark that p^o and p^s must be chosen such that $S(p^o) = D(p^s)$: otherwise it is always possible for the operator to decrease p^o (if $S(p^o) > D(p^s)$) or increase p^s (if $S(p^o) < D(p^s)$) to strictly improve its revenue. The operator revenue with such prices is then the area of the rectangle displayed in the left hand side of Figure 2, embedded within a zone whose area is the maximum value of social welfare. While $p^o > \max_i p_i^{\min}$ and $p^s < \min_i p_i^{\max}$, the largest revenue is attained when $S(p^o) = D(p^s) = C^*/2$. However we are not guaranteed that such p^o and p^s indeed verify $p^o > \max_i p_i^{\min}$ and $p^s < \min_i p_i^{\max}$, nor are we assured that such a choice yields the maximum revenue (that maximum might actually be attained with $p^o < \max_i p_i^{\min}$ or $p^s > \min_i p_i^{\max}$).

To be able to predict the choices of the profit-oriented monopoly, we therefore make the following assumption regarding user price thresholds, that fixes those two points.

Assumption B: The repartition of price thresholds p^{\min} and p^{\max} is such that

$$\max_i p_i^{\min} \leq \frac{p^* + \min_i p_i^{\min}}{2}, \quad (10)$$

$$\min_i p_i^{\max} \geq \frac{p^* + \max_i p_i^{\max}}{2}, \quad (11)$$

where $p^* = \frac{\sum_i a_i p_i^{\min} + b_i p_i^{\max}}{\sum_i (a_i + b_i)}$ from (6). Moreover user profile values a_i (resp. b_i) of all users $i \in \mathcal{I}$ are independent and identically distributed, and a_i and b_i are independent.

Remark that this straightforwardly imply that $\max_i p_i^{\min} \leq p^* \leq \min_i p_i^{\max}$, so under Assumptions A and B, (8) holds as noticed in the previous subsection.

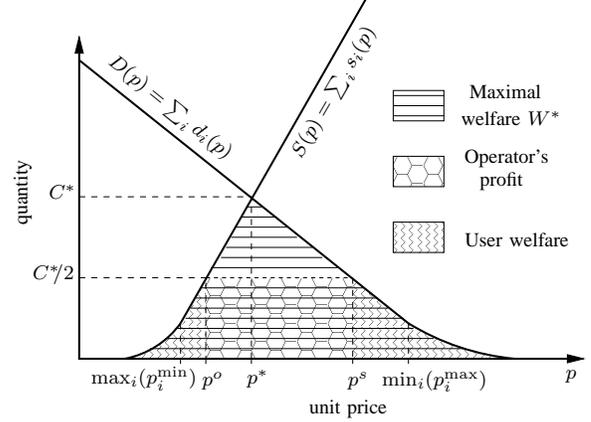


Fig. 2. Total supply S and demand D functions, maximum social welfare and surplus repartition with a revenue-driven monopoly under Assumption A.

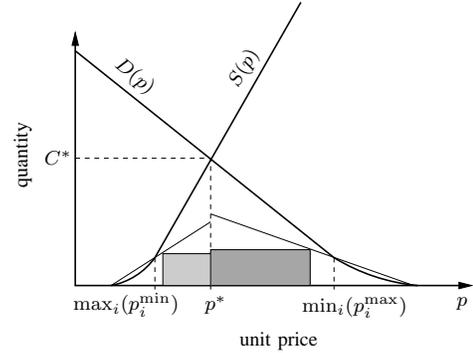


Fig. 3. Illustration of the proof of Proposition 1.

We can now quantify the performance of pricing mechanisms designed to maximize revenue.

Proposition 1: Under Assumptions A and B, a profit-oriented pricing yields a social welfare W_{mon} such that (with C^* given in (6)):

$$W^* - W_{mon} = \frac{1}{8} C^{*2} \left(\frac{1}{\sum_i a_i} + \frac{1}{\sum_i b_i} \right). \quad (12)$$

Proof: We first establish that the monopoly chooses the profit-maximizing unit prices p^o and p^s such that $S(p^o) = D(p^s) = C^*/2$, where p^* is the welfare-maximizing price given in (6). To do so, we compute an upper bound of the revenue that can be attained when choosing the prices in the non-linear part of S or D . Since both functions are convex, we can upper bound them by their chords on $[\min_i p_i^{\min}, \max_i p_i^{\min}]$ for the supply function, and on $[\min_i p_i^{\max}, \max_i p_i^{\max}]$ for the demand function. We extend these segments until the vertical line $p = p^*$ to form two triangles (with the abscissa axis). Under Assumption B, the largest rectangle embedded in each triangle is indeed embedded in the triangle formed by

extending the affine parts of S and D , as illustrated in Figure 3. Therefore the sum of their areas is smaller than the revenue yielded by the prices verifying $S(p^o) = D(p^s) = S(p^*)/2$, which are thus the profit-maximizing prices.

As illustrated in Figure 2, the difference $W^* - W_{mon}$ is then simply the area of the hatched triangle above the horizontal line $C^*/2$, which gives (12). ■

IV. WHICH MANAGEMENT TO PREFER?

In this section we compare the outcomes of the two practical schemes, i.e symmetric and payment-based schemes. From (8) and (12) we immediately have the following result.

Proposition 2: Under Assumptions A and B, symmetric schemes socially outperform profit-oriented pricing mechanisms if and only if

$$\frac{1}{4}C^{*2} \left(\frac{1}{\sum_i a_i} + \frac{1}{\sum_i b_i} \right) \geq \sum_i (a_i + b_i) (p^* - p_i^*)^2,$$

where C^* and p^* are given in (6), and $p_i^* = \frac{a_i p_i^{\min} + b_i p_i^{\max}}{a_i + b_i}$. That condition is equivalent to

$$\frac{1}{4} \left(p^* - \sum_i \alpha_i p_i^{\min} \right) \left(\sum_i \beta_i p_i^{\max} - p^* \right) \geq \sum_i \omega_i (p_i^* - p^*)^2, \quad (13)$$

with the weights for all $i \in \mathcal{I}$: $\alpha_i := \frac{a_i}{\sum_i a_i}$, $\beta_i := \frac{b_i}{\sum_i b_i}$, and $\omega_i := \frac{a_i + b_i}{\sum_i a_i + b_i}$.

Proof: Relation (13) comes after some algebra, using the equalities $p^* = \sum_i \omega_i p_i^*$ and $C^* = \sum_i a_i (p^* - p_i^{\min}) = \sum_i b_i (p_i^{\max} - p^*)$. ■

Proposition 2 combines the four user heterogeneity factors, namely the price thresholds p^{\min} , p^{\max} and price sensitivities a, b , to determine the best mechanism in terms of social welfare. Whereas the right-hand term of (13) is the variance of the p_i^* with weights ω_i , the left-hand term is hard to interpret. We thus suggest to have a look at the particular cases where user heterogeneity lies entirely on prices sensitivities (resp. on price thresholds).

A. Homogeneous price thresholds

We consider here that users only differ by their price sensitivities a_i and b_i . That simplified model has been studied in a previous work [13], we therefore recall the main results and refer the interested reader to [13] for details.

Assumption C: All users $i \in \mathcal{I}$ have the same price thresholds p_i^{\min} and p_i^{\max} . Without loss of generality (via a change of abscissa in Figure 2), we can therefore assume that

$$\forall i \in \mathcal{I}, \quad p_i^{\min} = 0 \quad \text{and} \quad p_i^{\max} = p^{\max}.$$

Notice that under Assumptions A and C, Assumption B always holds. It can then be proved (see [13]) that

$$\begin{cases} W_{sym} &= \left(\frac{1}{\sum_i a_i} + \frac{1}{\sum_i b_i} \right) \sum_i \left[\frac{1}{\frac{1}{a_i} + \frac{1}{b_i}} \right] W^*, \\ W_{mon} &= \frac{3}{4} W^*. \end{cases}$$

This yields the following comparison (which can also be directly obtained from Proposition 2 after some algebra).

Proposition 3: Under Assumptions A and C, symmetric schemes socially outperform profit-oriented pricing mechanisms if and only if

$$\left(\frac{1}{\sum_i a_i} + \frac{1}{\sum_i b_i} \right) \sum_i \frac{1}{\frac{1}{a_i} + \frac{1}{b_i}} \geq \frac{3}{4}. \quad (14)$$

Moreover, if the couples (a_i, b_i) are independently chosen for all users and identically distributed, then when the number of users tends to infinity, (14) writes

$$\frac{\mathbb{E}[f(a, b)]}{f(\mathbb{E}[a], \mathbb{E}[b])} \geq \frac{3}{4}, \quad \text{with } f : (x, y) \mapsto \frac{1}{1/x + 1/y}. \quad (15)$$

Since the function f is strictly concave, from Jensen's inequality the left-hand term of (15) is always smaller than 1, and decreases as the dispersion of (a, b) increases. Remark that when (a, b) are deterministic then the left-hand term of (15) equals 1 and symmetric schemes are better than profit-oriented ones, as we remarked in subsection III-B.

Let us have a look at (15) for two simple examples of distributions for (a, b) , assuming that a and b are independent variables.

• *Uniform distribution.* If a (resp. b) is uniformly distributed over $[0, a_{\max}]$ (resp. $[0, b_{\max}]$),

$$\begin{aligned} \frac{\mathbb{E}[f(a, b)]}{f(\mathbb{E}[a], \mathbb{E}[b])} &= \frac{2}{3} \left(\frac{1}{a_{\max}} + \frac{1}{b_{\max}} \right) \left(a_{\max} + b_{\max} \right. \\ &\quad \left. - \frac{a_{\max}^2}{b_{\max}} \ln\left(1 + \frac{b_{\max}}{a_{\max}}\right) - \frac{b_{\max}^2}{a_{\max}} \ln\left(1 + \frac{a_{\max}}{b_{\max}}\right) \right). \end{aligned} \quad (16)$$

This expression is minimum when $a_{\max} = b_{\max}$, in which case it equals $8(1 - \ln(2))/3 \simeq 0.82$. Consequently inequality (15) always holds.

• *Exponential distribution.* If a (resp. b) follows an exponential distributions with parameter μ_a (resp. μ_b), i.e. $\mathbb{P}(a > x) = e^{-\mu_a x}$, then we obtain after some calculation

$$\frac{\mathbb{E}[f(a, b)]}{f(\mathbb{E}[a], \mathbb{E}[b])} \geq \frac{3}{4} \quad \Leftrightarrow \quad \alpha \leq \frac{\mu_a}{\mu_b} \leq \frac{1}{\alpha},$$

where $\alpha \simeq 0.179$ is the smallest positive root of $x \mapsto \frac{1+x}{(1-x)^3} (1-x^2+2x \ln(x)) - 3/4$. In that case, either a symmetric or a profit oriented mechanism is socially preferable depending on the relative values of μ_a and μ_b .

B. Homogeneous price sensitivities

We now consider the case where the price thresholds p_i^{\min} and p_i^{\max} can be user specific, but the price sensitivities a_i and b_i are identical for every user.

Assumption D: All users have the same price sensitivity of supply (resp. demand), i.e. $\forall i \in \mathcal{I}, \quad a_i = a$ and $b_i = b$. Moreover, the couples (p_i^{\min}, p_i^{\max}) are independent and identically distributed among users, and p_i^{\min} is independent of p_i^{\max} for all $i \in \mathcal{I}$.

In that case, we establish that one mechanism is always preferable to the other.

Proposition 4: Under Assumptions A, B and D, management mechanisms based on symmetry are always socially better than profit-oriented pricing mechanisms.

Proof: Assumption D implies that for all i , the weights α_i , β_i and ω_i introduced in (13) all equal $\frac{1}{n}$, where n is the number of users. Moreover we have $p^* = \frac{a\tilde{p}^{\min} + b\tilde{p}^{\max}}{a+b}$, where $\tilde{p}^{\min} := \frac{\sum_i p_i^{\min}}{n}$ and $\tilde{p}^{\max} := \frac{\sum_i p_i^{\max}}{n}$. So when n tends to infinity, (13) is equivalent to

$$\frac{ab}{4} (\tilde{p}^{\max} - \tilde{p}^{\min})^2 \geq a^2 \text{Var}(p^{\min}) + b^2 \text{Var}(p^{\max}),$$

where Var denotes the variance, and where we used the independence assumption of p^{\max} and p^{\min} to develop the right-hand term.

Since the variance of a real variable with support length y is always smaller than $y^2/4$, and using (10)

$$\text{Var}(p^{\min}) \leq \frac{(p^* - \max p_i^{\min})^2}{4} \leq \frac{(p^* - \tilde{p}^{\min})^2}{4},$$

where the last inequality comes from $\tilde{p}^{\min} \leq \max p_i^{\min} \leq p^*$. Likewise, $\text{Var}(p^{\max}) \leq (\tilde{p}^{\max} - p^*)^2/4$. Therefore by replacing the optimal shadow price p^* by $(a\tilde{p}^{\min} + b\tilde{p}^{\max})/(a+b)$ and applying the inequality $(a+b)^2 \geq 2ab$, we get

$$a^2 \text{Var}(p^{\min}) + b^2 \text{Var}(p^{\max}) \leq \frac{ab}{4} (\tilde{p}^{\max} - \tilde{p}^{\min})^2,$$

Therefore Relation (13) is always satisfied and symmetric schemes always outperform profit-maximizing pricing schemes. ■

V. CONCLUSIONS AND FUTURE WORK

In this work we have addressed the problem of user incentives in a peer-to-peer storage system. Using a game theoretical model to describe selfish reactions of all system actors (users and the operator), we have studied and compared the outcomes of two possible managing schemes, namely symmetry-based and profit oriented payment-based. Not only the size of the offered storage space was targeted with incentives, but as the availability and reliability are particularly important issues in storage systems, the model also aimed to reduce churn. By comparing the social welfare level at the outcome in the two cases, under some assumptions on user preferences we exhibited a necessary and sufficient condition for a type of management to be preferable to the other: it appears that profit oriented payment-based schemes may be socially better than symmetric ones under some specific circumstances, namely if the heterogeneity among user profiles is high.

There are different ways to extend the results we have obtained. First of all, in reality the perceived utility of a user should not only depend on the amount of stored data and the associated availability, but also on the rapidity to access those data. Therefore the available bandwidth of a storage space offerer should be taken into account in addition to the amount of space proposed. Another interesting direction would be to consider demand and supply functions that are not affine, but can have any form, or eventually to carry out experiments to estimate the form of those functions. Finally, since a more complete and realistic model may not be solvable analytically, a simulation testbed could be built in order to study the behaviour of a peer-to-peer storage system in a more complex setting and eventually exhibit other phenomena that are not captured by our model.

REFERENCES

- [1] D. Fudenberg and J. Tirole, *Game Theory*. MIT Press, Cambridge, Massachusetts, 1991.
- [2] E. Adar and B. Huberman, "Free riding on Gnutella," *First Monday*, vol. 5, no. 10, Oct 2000.
- [3] P. Antoniadis, C. Courcoubetis, and R. Mason, "Comparing economic incentives in peer-to-peer networks," *Computer Networks*, vol. 46, no. 1, pp. 133–146, 2004.
- [4] C. Courcoubetis and R. Weber, "Incentives for large peer-to-peer systems," *IEEE JSAC*, vol. 24, no. 5, pp. 1034–1050, May 2006.
- [5] P. Golle, K. Leyton-Brown, I. Mironov, and M. Lillibridge, "Incentives for sharing in peer-to-peer networks," in *Proc. of 3rd ACM conference on Electronic Commerce (EC'01)*, Tampa, Florida, USA, Oct 2001, pp. 264–267.
- [6] K. Lai, M. Feldman, I. Stoica, and J. Chuang, "Incentives for cooperation in peer-to-peer networks," in *Proc. of 1st Workshop on Economics of Peer-to-Peer Systems (P2PECON'03)*, Berkeley, CA, USA, Jun 2003.
- [7] C. Batten, K. Barr, A. Saraf, and S. Treptin, "pStore: A secure peer-to-peer backup system," MIT Laboratory for Computer Science, Tech. Rep. MIT-LCS-TM-632, Dec 2001.
- [8] P. Druschel and A. Rowstron, "PAST: A large-scale, persistent peer-to-peer storage utility," in *HotOS VIII*, Schloss Elmau, Germany, May 2001, pp. 75–80.
- [9] M. Lillibridge, S. Elnikety, A. Birrell, M. Burrows, and M. Isard, "A cooperative internet backup scheme," in *Proc. of 1st Workshop on Economics of Peer-to-Peer Systems (P2PECON'03)*, Berkeley, CA, USA, Jun 2003.
- [10] L. Cox and B. Noble, "Samsara: Honor among thieves in peer-to-peer storage," in *Proc. of 19th ACM Symposium on Operating Systems Principles (SOSP'03)*, Bolton Landing, NY, Oct 2003.
- [11] B. Stefansson, A. Thodis, A. Ghodsi, and S. Haridi, "MyriadStore," Swedish Institute of Computer Science, Tech. Rep. T2006:09, May 2006.
- [12] D. Besanko and R. R. Braeutigam, *Microeconomics*. Wiley, 2005.
- [13] L. Toka and P. Maillé, "Managing a peer-to-peer backup system: Does imposed fairness socially outperform a revenue-driven monopoly?" in *Proc. of 4th International Workshop on Grid Economics and Business Models (Gecon 2007)*, Aug 2007.
- [14] R. Rodrigues and B. Liskov, "High availability in DHTs: Erasure coding vs.replication," in *Proc. of 4th International Workshop on Peer-to-Peer Systems (IPTPS)*, Ithaca, USA, Feb 2005.
- [15] A. Duminuco, E. Biersack, and T. En-Najjary, "Proactive replication in distributed storage systems using machine availability estimation," in *Proc. of 3rd International Conference on emerging Networking Experiments and Technologies (CoNEXT)*, New York, USA, Dec 2007.



Patrick Maillé graduated from Ecole polytechnique and TELECOM ParisTech, France, in 2000 and 2002, respectively. He has been an assistant professor at the Network, Security, Multimedia department of TELECOM Bretagne since 2002, where he obtained his PhD in applied mathematics in 2005. His research focuses on game theory applied to networks: resource pricing, routing games, consequences of user selfishness on network performance.



László Toka graduated in 2007 at Budapest University of Technology and Economics and received his MSc degree in Telecommunications. He also obtained the engineer diploma of Telecom Bretagne and Eurecom in France. After graduating he enrolled as a PhD candidate at Telecom Paris and at Budapest University of Technology and Economics. His research interests are on economic modeling of distributed systems, with special focus on incentives and mechanism design.

Appendix B

Main publications evoked in Chapter 3

Price War in Heterogeneous Wireless Networks

Patrick Maillé^a, Bruno Tuffin^b

^a*Institut Telecom; Telecom Bretagne
2 rue de la Châtaigneraie CS 17607, 35576 Cesson-Sévigné Cedex, FRANCE*

^b*INRIA Rennes - Bretagne Atlantique
Campus Universitaire de Beaulieu, 35042 Rennes Cedex, FRANCE*

Abstract

Wireless users have the opportunity to choose between heterogeneous access modes, such as 3G, WiFi or WiMAX for instance, which operate with different distance ranges. Due to the increasing commercial interest in access networks, those technologies are often managed by competing providers. The goal of this paper is to study the price war occurring in the case of two providers, with one provider operating in a sub-area of the other. A typical example is that of a WiFi operator against a WiMAX one, WiFi being operated in the smaller area. Using a simple model, we discuss how, for fixed prices, (elastic) demand is split among providers, and then characterize the Nash equilibria for the price war. We derive the conditions on provider capacities and coverage areas under which providers share demand on the common area. A striking additional result is that among the Nash equilibria, the one for which providers set the largest price corresponds to the case when the competitive environment does not bring any loss in terms of social welfare with respect to the socially optimal situation: at equilibrium, the

Email addresses: patrick.maill@telecom-bretagne.eu (Patrick Maillé),
btuffin@irisa.fr (Bruno Tuffin)

overall utility of the system is maximized. The price of stability is one.

Key words: Wireless networks, Pricing, Competition, Game theory

1. Introduction

1.1. Context

Broadband access networks are becoming prominent in nowadays life, with various applications such as Internet access, wired or wireless telephony, television... One of the main trends is the convergence of all those services in a single network. At the same time, personal devices such as laptops or cellular phones are reliant on ubiquitous connectivity: there is now the possibility to access the network by different means in terms of provider and technology. Each user may have the opportunity to choose his access mode depending on the service availability first, and then the feasible quality of service (QoS), pondered by the corresponding access charge. Among the numerous network access technologies, we can mention

- cable modem, fiber optic links and digital subscriber line (xDSL), that require fixed access from houses or offices,
- 3G (for third generation) wireless that may be accessed from most inhabited areas,
- WiFi (for Wireless Fidelity) technology, that has been developed by working group IEEE 802.11 to provide wireless access from local area networks or hotspots [1],

- WiMAX [2, 3] (for Worldwide Interoperability for Microwave Access), that has been more recently standardized by working group IEEE 802.16, in order to reach devices at further distances.

With respect to WiFi, WiMAX is a long-range system, covering many kilometers, while WiFi typically covers tens of meters, but WiMAX and WiFi also provide different Quality of Service (QoS).

Apart from this diversity in access technologies, another trend in networking is the transition from monopolies to oligopolies. Since the Internet has moved from an academic network to a commercial one with providers fighting for customers by choosing the appropriate access price, competition issues in Internet access are highly relevant. Providers have to charge for access as a return on investment and want to maximize their profits. On the other hand, they have to take care of prices of competitors, since users can find a better combination of QoS and price with a competitor, and change providers. This kind of interaction is typical of non-cooperative game theory [4], and one usually tries to look for a *Nash equilibrium*, representing here a state where no provider can increase his revenue by an unilateral price change.

1.2. Goal

In this paper, we consider two providers in competition for customers. Users are assumed non-atomic, in the sense that their *individual* actions have no influence on the QoS of others. They are charged a fixed price per sent packet, so that the average price per served packet is the packet price charged divided by the probability of successful transmission. This way, a

congestion cost is imposed thanks to the loss probability. Indeed, losses are frequently an issue in wireless networks, such as when dealing with WiFi for instance. Total demand, in terms of effective throughput, is assumed to be a decreasing function of the average price per served packet, that we call the *perceived* price. Each customer chooses the provider with the best -i.e., cheapest- perceived price. This results in a customer distribution equilibrium satisfying the Wardrop principle. That principle is widely used in transportation theory, an area closely related to telecommunications [5], and states in our context that within an area of competition between providers, the perceived price has to be the same at both providers provided they attract some demand; otherwise the highest charged users would have an interest in switching to the cheapest provider. The providers (which will be called provider 1 and provider 2) are assumed to have fixed (but possibly different) capacities, and operate in different areas. We assume that provider 2 operates in a sub-domain of provider 1's access area. Provider 1 could typically represent a WiMAX operator while the other proposes WiFi access. WiMAX can reach customers at a much longer distance than WiFi, and therefore runs a larger coverage area. We can then think of a WiMAX provider enduring competition on a fraction only of his customers, since the other part is not reachable from his competitor. The questions we aim at answering are:

- What is the strategy of each provider in terms of price setting, knowing what the user distribution would be (the Wardrop equilibrium) for any given couple of prices?
- Shall the (WiMAX) provider compete for demand on the common market, or shall he just focus on revenue on the monopolistic area to de-

termine his price so that all users in the common area could prefer to go to the (WiFi) competitor?

- Is there a (Nash) equilibrium in the price war? If it is the case, is it unique?
- What is the *price of anarchy* due to non-cooperation? The price of anarchy is a measure of the loss of efficiency due to actors' selfishness. This loss has been defined in [6] as the worst-case ratio comparing the global efficiency measure (that has to be chosen) at an outcome of the noncooperative game played among actors, to the optimal value of that efficiency measure. Similarly, what is the *price of stability*, measuring the loss of efficiency when the *best* Nash equilibrium is reached [7] (i.e., if we consider the socially optimal situation such that no actor will defect)?

1.3. Related work

Our work uses game theory to model competition among providers. Game theory [4] is a powerful tool for representing the interactions of selfish actors, and has been quite recently introduced in telecommunication networks; see [8] for a survey on the different types of problems that can be encountered. More specifically, our goal is to study pricing issues. Pricing [9, 10, 11] has been used in telecommunications to cope with congestion due to more and more demanding applications and an increasing number of customers; here typically, game theory is the natural tool to describe the interplay of selfish customers in front of a given pricing scheme. Providers use pricing

to better control demand, differentiate services for different QoS-requiring users/applications, and/or provide return on investment.

On the other hand, most of the studies investigate the case of a single provider, a *monopoly*, and it is only recently that modeling the competition among providers has been introduced in networking. Competition may disrupt the monopoly-case behavior of some schemes such as the very promising *Paris Metro Pricing* (PMP) scheme, consisting in separating the network into disjoint networks served in the same manner but with different access prices. In that case, there is no guarantee that the QoS will be better at a subnetwork than at another, but it is expected that most expensive ones will be less congested due to the higher price. It is shown in [12] that such a simple and attractive scheme to differentiate service actually does not allow service differentiation under competition, since at equilibrium no provider has an interest in offering several classes. Other competition models, with less complexity than ours, have been studied. For example, [13] models competitive providers playing both on price and on a QoS parameter, but demand is there driven by an arbitrary function which does not depend on price and QoS at competitors and therefore does not cover the fact that users could switch to more attractive providers, if any. The Wardrop's principle we consider here precisely encompasses that aspect. [14] considers on the other hand a Wardrop equilibrium among users, but QoS does not depend on demand, a simplifying assumption we do not make here. In wireless networks, competition has been analyzed by several works in the case of a shared spectrum, in order to lead to a more efficient utilization than with potentially unused fixed licenses. For instance, [15] uses a more specific model than ours and

shows that competition may increase users' acceptance probability for offered service. In [16], competition among selfish wireless providers is also considered, but their strategy space is only on the power of the pilot signals of their base stations, and does not include any pricing activity, a lever that should require attention. [17] studies the case where an operator can lease part of the bandwidth he owns from his license; a learning automaton is used to converge to an equilibrium, while in our model a direct proof of existence and uniqueness of an equilibrium is obtained. In a more general context, [18] studies competition in the case of uncertainty on demand, whereas in our case demand repartition among providers is obtained through a (deterministic) equilibrium among users. In [19], the pricing competition between a WiMAX and a WiFi community is investigated, but the externality is coverage instead of QoS here: the more customers the WiFi community has, the more connectivity it has. A model more closely related to ours is in [20], where atomic users can choose between two technologies operating on different ranges, typically a WAN and WiFi hotspots. Using a stochastic geometric model for the locations of customers and providers' access points and a greedy algorithm for the decision about which technology to use, multiple equilibria are found for a final selection. WAN and WiFi competition is analyzed in an asymptotic scenario where the service zones of WAN provider are much larger than those of WiFi access providers. Our model is different from the fact that users are assumed non-atomic. This drives to an analytical characterization of the equilibrium. Moreover, no asymptotic scenario is required for the analysis and we are able to precisely determine when both providers will attract customers in the common area.

Studying competition for customers when demand is distributed according to Wardrop's principle, was considered in [21, 22, 23], where the QoS externality is the expected delay, while it is the loss probability here, which seems more relevant for some wireless contexts. The price of anarchy, measuring the loss of efficiency due to competition with respect to cooperation, is determined, for fixed demand in [21] and random demand but linear delay in [23]; we do look at the price of anarchy too (that is unbounded here), but we also have a look at the price of stability and prove it equals one, i.e., competition can lead the system to the socially optimal situation. Moreover, we consider a more comprehensive model, by including the fact that part of customers are not accessible from one of the providers, thus competition is only on one part of demand. On the other hand, setting a price too high would also reduce (elastic) demand in the part where the WiMAX provider has a monopoly.

Note that competition can also occur in interdomain or multihop networks, where selfish providers need to send their traffic through competitors' networks to ensure end-to-end delivery, and pricing is a mean to produce such incentives [24, 25, 26]. The goal is different in this paper because we only look at direct competition for users between providers.

We have studied competition among providers in a previous work [27] using also loss probability as the externality, but for a specific network topology where all users have the choice among all providers. That could represent competition among access providers using the same technology, say WiFi, at a given hotspot or hotzone.

In this paper, we intend to model the competition in heterogeneous net-

works, i.e. for providers using different wireless technologies. Those technologies correspond to different coverage areas, and it therefore results in a model drastically different from the one studied in [27]. Indeed, the mathematical characterization of the user equilibrium changes completely since the different coverage zones have to be taken into account. Consequently, the higher level game played on prices by competing providers is much more complicated to study and all the required proofs are of different nature. On the other hand, we believe that studying the heterogeneousness we introduce here is a primal need, because it is a very important aspect of nowadays wireless networks.

1.4. Organization of the paper

The paper is organized as follows. Section 2 presents the mathematical model we will use to represent provider competition in heterogeneous networks, while Section 3 defines our social welfare measure as the sum of utilities of all actors (customers plus providers) and compute its maximum value; this will provide a reference to investigate the loss of efficiency due to competition. Section 4 discusses how demand is split -according to Wardrop’s principle- between providers in both zones, the common one and the one where provider 1 is a monopoly. Section 5 then shows what the Nash equilibria are for the pricing game between providers, with an explicit characterization depending on the proportion of demand that is common. It is also shown that with social welfare as a global performance measure, the price of stability is one, meaning that there is no loss of efficiency by introducing competition when using the “best” Nash equilibrium. An argument is provided in favor of that particular equilibrium. The price of anarchy, when

comparing social welfare at the optimal value and at the worst Nash equilibrium, is also computed. Finally, Section 6 concludes and gives directions for future research.

2. Model

2.1. Network topology and perceived prices

Consider two providers, denoted by 1 and 2, with provider 2 operating in a subdomain of provider 1, as illustrated in Figure 1. This is a typical situation

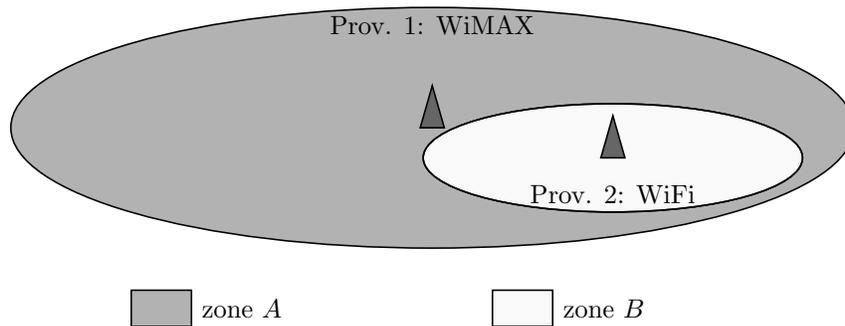


Figure 1: The competition framework

of a WiFi provider operating on smaller distances -tens of meters- than a WiMAX one -covering many kilometers-. As a consequence, competition only occurs in the domain of operator 2, while operator 1 has a monopoly in the remaining area. But operator 1 having a unique price, competition influences the optimal price in the monopoly area. As illustrated in Figure 1, we partition the total domain in

- zone A , the domain where only provider 1 operates, and
- zone B , the domain where both providers operate.

In order to analyze the outcome of competition, we consider a model where time is discretized, divided into slots. Provider i ($i \in \{1, 2\}$) is assumed to be able to serve C_i packets (or units, seen as a continuous number) per slot. Congestion is experienced at provider i if demand exceeds capacity, and demand in excess is lost, lost packets being chosen uniformly over the set of submitted ones. Formally, let d_i be the total demand at provider i . Then the number of packets served is $\min(d_i, C_i)$, meaning that packets are actually served with probability $\min(C_i/d_i, 1)$, i.e., packets are eventually served after a random time following a geometric distribution with parameter $\min(C_i/d_i, 1)$. Following an idea first introduced in [28], prices are per *submitted* packet rather than received one in order to prevent users from sending as many packets as possible, which would maximize their chance to be served. Charging on sent packets instead of successfully transmitted ones may seem unrealistic. However, that mechanism can be seen as a volume-based pricing scheme, with a congestion-dependent charge. Somewhat equivalently, it can also be seen as a consequence of the more frequently used time-based charging with a fixed price per time unit. Indeed, when congestion occurs on a network i and packets are lost, having to send them again multiplies the total transfer time (and thus the price paid) by $\max(1, d_i/C_i)$, the mean number of transmissions per packet. If each packet sent to provider i is charged p_i , the expected price \bar{p}_i to successfully send a packet is therefore given by

$$\bar{p}_i = p_i / \min(C_i/d_i, 1) = p_i \max(d_i/C_i, 1), \quad (1)$$

which will from now be called the *perceived price* per served traffic unit at provider i . Figure 2 plots that perceived price \bar{p}_i depending on the demand d_i : \bar{p}_i is constant while provider i is not saturated, and increases linearly when

demand exceeds capacity C_i . Demand for provider 1 is decomposed into $d_{1,A}$,

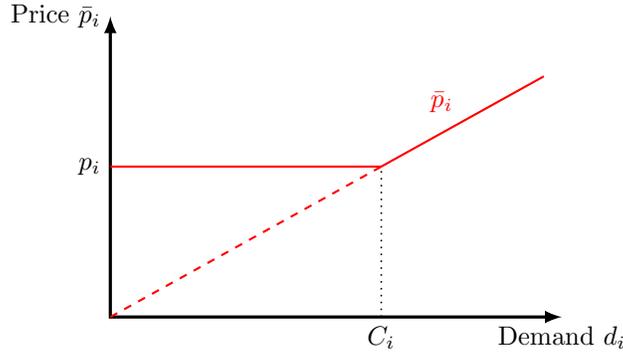


Figure 2: Perceived price at provider i versus demand d_i .

the demand in zone A , and $d_{1,B}$, the demand in zone B , with $d_1 = d_{1,A} + d_{1,B}$. Remark that our work does not deal with customer mobility: we assume that the (wireless) users do not move as soon as connected, a situation typical of most current WiFi users. Therefore coverage is not an issue for customers.

In a given zone $z \in \{A, B\}$ where the subset of operating service providers is $\mathcal{I}^z \subset \{1, 2\}$, the perceived price can be defined as

$$\bar{p}^z := \min_{i \in \mathcal{I}^z} \bar{p}_i.$$

This models the fact that users are only sensitive to the lowest perceived price available, since they choose the least expensive network.

2.2. User demand and valuation

In this paper, we assume that users are sensitive to the perceived price, in the sense that they reduce their demand when the perceived price increases. We model that effect using an aggregated demand function $D(\cdot)$.

Definition 1. *If the perceived price \bar{p} were the same on the whole domain, then the total demand is a function $D(\cdot)$ of that perceived price \bar{p} . Let us denote by $[0, p_{\max})$ the support of D .*

The demand function D is assumed to be continuous and strictly decreasing on its support, with $D(p_{\max}) = 0$ and possibly $p_{\max} = +\infty$, meaning that there is demand starvation when price is sufficiently high.

In other words, $D(\bar{p})$ represents the number of users/packets having a willingness to pay larger than or equal to \bar{p} . To deal with the case where there actually is competition, we assume that there is not enough resource to satisfy all demand, i.e., $D(0) > C_1 + C_2$.

A useful function in the rest of the paper is the *marginal valuation function*, that is the generalized inverse of the demand function.

Definition 2. *The maximum unit price at which a given quantity of traffic units can be sold is called the marginal valuation for that quantity. The marginal valuation is thus the application $v : q \mapsto \min\{p : D(p) \leq q\}$, with the convention $\min \emptyset = 0$.*

The sum of the marginal valuations of the q units of users with largest willingness-to-pay is denoted by $V(q)$, and $V(\cdot)$ is called the global valuation function. Formally,

$$V(q) := \int_{x=0}^q v(x) dx.$$

Notice that v is a nonincreasing function since D is nonincreasing. It is easy to see that

$$v(q) = \begin{cases} D^{-1}(q) & \text{if } q \in (0, D(0)) \\ p_{\max} & \text{if } q = 0 \\ 0 & \text{if } q \geq D(0). \end{cases} \quad (2)$$

Consequently, the valuation function V is nondecreasing and concave. $V(q)$ measures the “value” that the service has for the whole population, since it is the total price that the q units of demand with highest marginal valuation (i.e., those that actually accept to pay the unit price $v(q)$) are willing to pay to be served.

Since perceived prices on both zones may be different, we introduce a new parameter (namely, the proportion of the population covered by zone B) to express separately the demand in each zone, still using the aggregated demand function D .

Definition 3. *Let us denote by α the proportion of the population in zone B . We consider that users’ willingness-to-pay across sub-domains A and B are equidistributed. Therefore, total demand in zone A (resp., zone B) is $(1 - \alpha)D(\bar{p}_A)$ (resp., $\alpha D(\bar{p}_B)$) if the perceived price on that zone is \bar{p}_A (resp., \bar{p}_B).*

If users are uniformly distributed over the domain, α is simply the proportion of the surface covered by provider 2 with respect to provider 1, but it can be more general if we assume a non-uniform repartition.

Most of our results hold under the following assumption on the influence of price on demand.

Assumption A. Demand function D is differentiable, and price elasticity of demand $-\frac{D'(p)p}{D(p)}$ is strictly larger than 1 for all $p \in [\hat{p}, p_{\max})$, with $\hat{p} \leq \min\left(v\left(\frac{C_1}{1-\alpha}\right), v\left(\frac{C_2}{\alpha}\right)\right)$.

The price elasticity of demand measures the percentage change in demanded quantity implied by a percentage change in perceived price. Values larger than 1 (leading to *relatively elastic* demand in economic terms) correspond to a quite high reactivity to a perceived price change.

Under Assumption A, the function $p \mapsto pD(p)$ is strictly decreasing on $[\hat{p}, p_{\max})$; this is a typical assumption in telecommunications ($\hat{p} = 0$ is often considered, our assumption here is weaker), confirmed by operators¹. This property will be used in this paper to characterize the Nash equilibrium of the pricing game.

Assumption A can be interpreted as follows: if all users in zone B always choose provider 2 (or equivalently, if both zones were disjoint), then both providers have an interest in setting a price such that all of their capacity is used. Indeed, otherwise the revenue of provider i covering a proportion α_i of the population is $\alpha_i p_i D(p_i)$, which is strictly decreasing in p_i , thus provider i should decrease its price to maximize its revenue.

2.3. Methodology

Our analysis of the pricing game is decomposed into three steps:

1. We first study how, for fixed prices p_i ($i \in \{1, 2\}$), total demand is split among providers. This is described and discussed in Section 4 in terms

¹From discussions at Orange Labs.

of a Wardrop equilibrium. The output, also called user equilibrium, consists in a demand distribution $\mathbf{d} := (d_{1,A}, d_{1,B}, d_2)$. Notice that we do have to consider the two different zones, each one impacting the other, when computing that equilibrium. As we will see, we may end up with different perceived prices on the two different zones.

2. Knowing how demand is distributed for fixed prices, each provider $i \in \{1, 2\}$ tries to maximize his revenue

$$R_i(p_1, p_2) := p_i d_i$$

by playing with the price charged to customers. The strategy of a provider has an impact on the demand distribution, and therefore on the revenue of the other. In Section 5 we determine the Nash equilibria for the price game. Recall here the definition of a Nash equilibrium when applied to our problem

Definition 4. *A Nash equilibrium is a price vector $\mathbf{p}^* := (p_1^*, p_2^*)$ such that no provider can increase his own benefit by unilaterally changing his access price, i.e., $\forall p \geq 0$,*

$$R_1(p_1^*, p_2^*) \geq R_1(p, p_2^*) \text{ and } R_2(p_1^*, p_2^*) \geq R_2(p_1^*, p).$$

3. In the same section, we show that among all the Nash equilibria, there is one corresponding to the socially-optimal situation, so that there is no loss of efficiency due to competition (the price of stability is one). We actually argue that this equilibrium is the most likely if (even if negligible and not counted here) management costs are involved. We also compute the price of anarchy if the worst Nash equilibrium, in terms of social welfare, is chosen.

3. Social welfare and optimal value

We define here social welfare (SW) as the sum of utilities of all agents (customers plus providers) in our specific context, and then study the optimal value that can be obtained.

If we consider only zone B , $v(q)$ is the price a user would pay to buy the $\alpha \times q$ -th unit since only a proportion α of the population is in that zone. A customer buying the q -th unit of resource in zone B is therefore willing to pay $v(q/\alpha)$ to be served. If total demand in zone B is $d_{1,B} + d_2$, then the total price that users in zone B are willing to pay is

$$\int_{x=0}^{d_{1,B}+d_2} v(x/\alpha)dx = \alpha V\left(\frac{d_{1,B} + d_2}{\alpha}\right).$$

However, the demand $d_{1,B} + d_2$ might not totally be served due to capacity limitations. Consequently, reasonably assuming that packet loss are independent of user willingness-to-pay, the value that the service has to zone B users should include the average transmission success probability in zone B : that overall value is then

$$\frac{d_{1,B}\pi_1 + d_2\pi_2}{d_{1,B} + d_2} \alpha V\left(\frac{d_{1,B} + d_2}{\alpha}\right),$$

where $\pi_1 := \min\left(1, \frac{C_1}{d_{1,B}+d_{1,A}}\right)$ and $\pi_2 := \min\left(1, \frac{C_2}{d_2}\right)$ are the transmission success probabilities with provider 1 and provider 2, respectively. Similarly, the total value that the service has for zone A users is

$$\pi_1(1 - \alpha)V\left(\frac{d_{1,A}}{1 - \alpha}\right),$$

leading to the following definition of social welfare.

Definition 5. For a demand configuration $(d_{1,A}, d_{1,B}, d_2)$, social welfare (sum of utilities of all actors) is

$$SW(d_{1,A}, d_{1,B}, d_2) = \pi_1(1 - \alpha)V\left(\frac{d_{1,A}}{1 - \alpha}\right) + \frac{d_{1,B}\pi_1 + d_2\pi_2}{d_{1,B} + d_2}\alpha V\left(\frac{d_{1,B} + d_2}{\alpha}\right), \quad (3)$$

where $\pi_1 := \min\left(1, \frac{C_1}{d_{1,B} + d_{1,A}}\right)$ and $\pi_2 := \min\left(1, \frac{C_2}{d_2}\right)$.

Remark that social welfare depends only on $(d_{1,A}, d_{1,B}, d_2)$, but not on prices paid by users since SW is the sum of the utilities of all actors: customers (with their willingness to pay minus price paid) and providers (with the revenue they get from prices).

From Definition 5, maximizing social welfare can be formally written as

$$\begin{aligned} \max SW(d_{1,A}, d_{1,B}, d_2) & \quad (4) \\ \text{s.t. } d_{1,A} \geq 0, d_{1,B} \geq 0, d_2 \geq 0. & \end{aligned}$$

We now solve that optimization problem.

Proposition 1. The maximal value SW^* of social welfare is

$$SW^* = \begin{cases} V(C_1 + C_2) & \text{if } \alpha C_1 \geq (1 - \alpha)C_2, \\ (1 - \alpha)V\left(\frac{C_1}{1 - \alpha}\right) + \alpha V\left(\frac{C_2}{\alpha}\right) & \text{otherwise.} \end{cases}$$

The proof is provided in Appendix A.

4. Demand distribution

Let us now describe more clearly how demand distributes itself among providers. As in several other works where the number of users is large and no user has a significant weight with respect to the others [29, 30], we assume

that users are infinitely small: their choices do not individually affect the demand levels (and therefore the perceived costs) of the different providers. Games involving infinitesimal users are called *nonatomic games* [31]. Under that nonatomicity assumption, an equilibrium among users follows Wardrop's principle [5] taken from road transportation: demand is distributed in such a way that all users choose the available provider with the least perceived price, and none if this perceived price is too expensive. That principle is formalized below.

Definition 6. *A Wardrop (or user) equilibrium is a triple $(d_{1,A}, d_{1,B}, d_2)$ that verifies the following system, where \bar{p}_i stands for the perceived price at provider $i \in \{1, 2\}$.*

$$\bar{p}_1 = p_1 \max\left(1, \frac{d_{1,A} + d_{1,B}}{C_1}\right) \quad (5)$$

$$\bar{p}_2 = p_2 \max\left(1, \frac{d_2}{C_2}\right) \quad (6)$$

$$d_{1,A} \min\left(1, \frac{C_1}{d_{1,A} + d_{1,B}}\right) = (1 - \alpha)D(\bar{p}_1) \quad (7)$$

$$d_{1,B} \min\left(1, \frac{C_1}{d_{1,A} + d_{1,B}}\right) + d_2 \min(1, C_2/d_2) = \alpha D(\min(\bar{p}_1, \bar{p}_2)) \quad (8)$$

$$\bar{p}_1 > \bar{p}_2 \Rightarrow d_{1,B} = 0 \quad (9)$$

$$\bar{p}_1 < \bar{p}_2 \Rightarrow d_2 = 0. \quad (10)$$

We now give the interpretations for those relations. (5) and (6) are simply (1) applied to provider 1 and 2, respectively. Relations (7) and (8) link

demand (in terms of *effective* throughput, hence the multiplications by the success probabilities) to perceived prices in zones A and B . In zone B , where $100\alpha\%$ of the population is, the perceived price is $\bar{p}^B = \min(\bar{p}_1, \bar{p}_2)$. The other users (a proportion $100(1 - \alpha)\%$) are in zone A , with perceived price $\bar{p}^A = \bar{p}_1$. As suggested in the definition of perceived prices per zone, the min in the right-hand side of (8) reflects the fact that users in zone B choose the cheapest provider (only provider 1 is available in zone A). Finally, relations (9) and (10) also represent user choices in zone B : if one provider is strictly more expensive than the other, then he gets no demand in that zone. An example of the situation faced by users is illustrated in Figure 3. To see how things happen in each zone, we artificially consider that demand

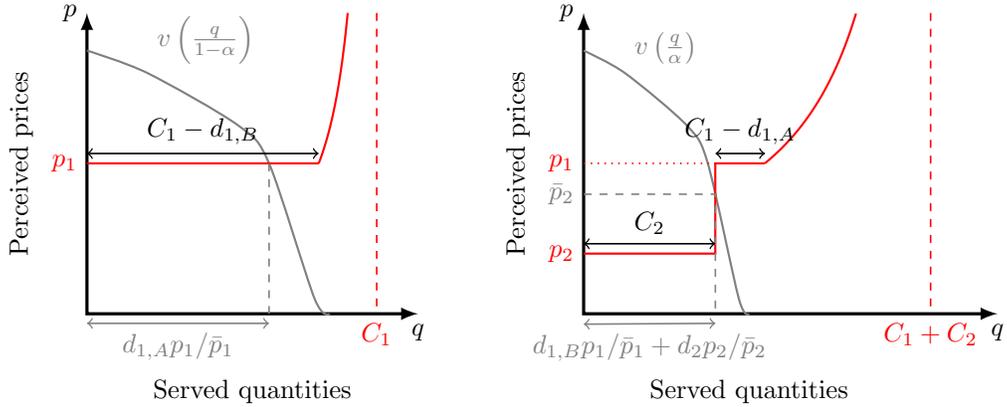


Figure 3: Demand repartition in zones A (left) and B (right).

is fixed in the other zone. In zone A , while demand $d_{1,A} \leq C_1 - d_{1,B}$, then from (5) the perceived price is $\bar{p}_1 = p_1$. Then when $d_{1,A} > C_1 - d_{1,B}$ we get $\bar{p}_1 = p_1 \frac{d_{1,A} + d_{1,B}}{C_1}$. In that case, losses occur, so that for a fixed $d_{1,B}$, the perceived price to actually get a service rate q on zone A (i.e., because

provider 1 is saturated, the total quantity served, C_1 , is decomposed into $C_1 = q + d_{1,B}p_1/\bar{p}_1$ among the two zones) is $\bar{p}_1 = p_1 \frac{d_{1,B}}{C_1 - q}$. At a Wardrop equilibrium, from (7) in zone A the pair $(d_{1,A}, \bar{p}_1)$ is the (unique) intersection point of the functions $q \mapsto p_1 \max\left(1, \frac{d_{1,B}}{C_1 - q}\right)$ and $q \mapsto v\left(\frac{q}{1-\alpha}\right)$.

In zone B , both providers are involved and users first choose the cheapest provider (here, provider 2) until it is saturated, then they continue choosing it, increasing the perceived price due to losses, until both providers have the same perceived price. Then some demand is served by provider 1 at a unit price p_1 , until it gets saturated. Afterwards, if $d_{1,A}$ is fixed, then the perceived price in zone B to be served at a rate q is $\bar{p}_1 = \bar{p}_2 = p_1 \frac{d_{1,A}}{C_1 + C_2 - q}$.

From (8), the pair $(d_{1,B} + d_2, \bar{p}_2)$ is the (unique) intersection point of that demand-price relation with the function $q \mapsto v(q/\alpha)$. In the example of Figure 3, we have $d_2 > C_2$ but $\bar{p}_2 < p_1$, thus $d_{1,B} = 0$ (i.e. all users in zone B choose provider 2 because demand is fulfilled in that zone before perceived price at provider 2 reaches p_1 , the price at provider 1).

The difficulty of the Wardrop equilibrium is that both zones have to be combined: the demand $d_{1,A}$ in zone A must correspond to the values of $d_{1,B}$ and d_2 in zone B and vice-versa.

The following proposition gives insightful results about the existence and characterization of a Wardrop equilibrium.

Proposition 2. *For every price profile (p_1, p_2) with strictly positive prices, there exists at least a Wardrop equilibrium. Moreover, the corresponding perceived prices (\bar{p}_1, \bar{p}_2) are unique.*

The proof is given in Appendix B.

Remark 1. *The uniqueness of perceived prices at a Wardrop equilibrium leads in most cases to a uniqueness of demands. Actually from (7) and (8), $d_{1,A}$ and $d_{1,B}p_1/\bar{p}_1 + d_2p_2/\bar{p}_2$ are unique. From (5) and (6), if $\bar{p}_1 > p_1$ or $\bar{p}_2 > p_2$ then demands are unique. Also, (9) and (10) imply that demands are also unique if $\bar{p}_1 \neq \bar{p}_2$. Therefore the only cases when demands might not be unique are when $\bar{p}_1 = p_1 = p_2 = \bar{p}_2$. Moreover, if $d_1 + d_2 = C_1 + C_2$ then demands are also unique (proof by contradiction: either $d_1 = C_1$ and $d_2 = C_2$, or from (1) one provider $i \in \{1, 2\}$ has $\bar{p}_i > p_i$). This will actually be the case for the Nash equilibrium of the pricing game: we will end up with $d_1 + d_2 = C_1 + C_2$, and a unique Wardrop equilibrium.*

We will see in the next section that even in a competitive context, situations with $\bar{p}_1 > \bar{p}_2$ can occur. In that case, all customers in zone B join provider 2, but the revenue that provider 1 gets from zone A exceeds what he could obtain by entering the price war on zone B .

5. Price war and Nash equilibrium

Knowing the above user equilibrium, we can discuss the pricing game between the two providers. Provider $i \in \{1, 2\}$ tries to maximize his revenue $R_i(p_1, p_2) = p_i d_i$ by playing with his price. Again, a price change modifies the Wardrop equilibrium, therefore the revenue of the competitor.

We give here a simple lemma regarding providers revenues.

Lemma 1. *For each provider i , $i = 1, 2$, we have at a Wardrop equilibrium $R_i \leq \bar{p}_i C_i$, and*

$$d_i \geq C_i \Leftrightarrow R_i = \bar{p}_i C_i. \quad (11)$$

As a consequence, we also have $\bar{p}_i > p_i \Rightarrow R_i = \bar{p}_i C_i$.

Proof: The lemma immediately follows from (5), (6) and the expressions of the revenues $R_i = p_i d_i$. \square

We now show our main result, characterizing the set of Nash equilibria.

Proposition 3. *Under Assumption A, in the price war between providers there is a set of Nash equilibria $(0, p_1^*] \times (0, p_2^*]$ for the price profile (p_1, p_2) , all yielding the same revenues $R_1^* = p_1^* C_1$ and $R_2^* = p_2^* C_2$. This set is characterized as follows.*

- If $\frac{C_1}{1-\alpha} \leq \frac{C_2}{\alpha}$, that set of Nash equilibria is such that

$$p_1^* = v \left(\frac{C_1}{1-\alpha} \right) \geq p_2^* = v \left(\frac{C_2}{\alpha} \right). \quad (12)$$

We then have $d_{1,B} = 0$, meaning that zone B is left to provider 2 by provider 1.

- If $\frac{C_1}{1-\alpha} > \frac{C_2}{\alpha}$, the set of Nash equilibria $(0, p_1^*] \times (0, p_2^*]$ is such that

$$p_1^* = p_2^* = p^* = v(C_1 + C_2). \quad (13)$$

In that case, zone B is shared by the providers.

The proof is given in Appendix C.

Remark 2. *The assumption $\frac{C_1}{1-\alpha} \leq \frac{C_2}{\alpha}$ means that the capacity per unit of surface for (the smaller-range) provider 2 is larger than that in the remaining area for provider 1. This can happen for fixed C_1 and C_2 if the proportion α of the common zone is small enough. As a consequence, at a Nash equilibrium,*

it is better for provider 1 to disregard potential revenue from zone B , and all users there go to the cheaper provider 2. We therefore end up with two monopolies in the different zones. On the other hand, if the assumption is not verified, zone B is too important for provider 1, and the price war is played. Both providers then share the area.

Remark 3. Among all the Nash equilibria, all yielding the same revenues, the price profile (p_1^*, p_2^*) is the one for which demand is the smallest, because price is the highest. We claim that it is the most likely situation since there is in this case less demand to manage, therefore less management costs, even if those costs are assumed negligible and not considered here. In the next proposition, we actually show that this equilibrium exactly corresponds to the socially-optimal situation.

Corollary 4. In this system, the Nash equilibrium (p_1^*, p_2^*) corresponds to the socially-optimal situation. As a consequence, the price of stability, defined as the best-case ratio comparing social welfare at the Nash equilibrium to the optimal value, is equal to one.

Proof: This corollary is a direct consequence of the Nash equilibrium (p_1^*, p_2^*) demand repartition, that exactly corresponds to the socially optimum one computed in Section 3. \square

However, if we consider any Nash equilibrium, then the performance of the system can be arbitrarily bad with respect to the socially optimal situation.

Corollary 5. In this system, the price of anarchy is unbounded. Indeed, social welfare tends to 0 when the prices fixed by providers tend to 0 (that situation being a Nash equilibrium of the pricing game).

Proof: As seen in the proof of Proposition 3, when prices (p_1, p_2) set by providers are sufficiently small then (p_1, p_2) is a Nash equilibrium, and $\bar{p}_i = p_i^*$ for p_i^* given in (12)-(13). From (5)-(6) and (7), this means that demands $d_{1,A}$ and d_2 tend to infinity. Now remark that due to the concavity and increasingness of V ,

$$\lim_{x \rightarrow \infty} V(x)/x = \lim_{x \rightarrow \infty} v(x) = 0,$$

where the last equality is a consequence of D being bounded for strictly positive prices.

Consequently, using $d_i \pi_i \leq C_i$ in the social welfare expression, we have when prices tend to 0:

$$\text{SW} \leq C_1 \underbrace{\frac{1-\alpha}{d_{1,A}} V\left(\frac{d_{1,A}}{1-\alpha}\right)}_{\rightarrow 0} + (C_1 + C_2) \underbrace{\frac{\alpha}{d_{1,B} + d_2} V\left(\frac{d_{1,B} + d_2}{\alpha}\right)}_{\rightarrow 0}, \quad (14)$$

which concludes the proof. \square

6. Conclusion

In this paper, we have studied a pricing game between two wireless access providers, one of the two (say, with WiFi technology) operating only in a sub-area of the other (say, with WiMAX technology). Demand is driven by the perceived price, being the price charged per packet sent divided by the probability to be served (i.e. the average price per served unit). Users are assumed to choose the cheapest available provider, or none if both are too expensive. We have explained how demand is distributed according to Wardrop's principle. Knowing this distribution, providers play a pricing game in order to maximize their revenue. We have characterized explicitly

all Nash equilibria for that game. Moreover, if the capacity per user offered in the WiFi hotzone exceeds the capacity per user of the WiMAX access in the remaining zone, then the WiMAX provider leaves the common area to the WiFi provider and only takes care of the region where he is the only provider available. Otherwise, the providers share the common area. A last contribution is to study whether competition brings a loss in terms of social welfare with respect to the cooperative case. We have shown that the price of stability (i.e., when looking at the Nash equilibrium yielding the largest welfare) equals one, and remarked that this situation is actually a likely one.

As directions for future research, we plan to look at several issues. First, the case of more than two providers would be interesting to study, but much more complex. Adding demand uncertainty, and/or other externalities than loss probability such as delay [22], to the model could highlight more complex provider strategies and increase the price of stability. Also, considering that providers can not only play with their price but also with their capacity or the area they can reach would be of interest: in wireless networks, this could typically mean playing with the transmission power of the antennas (or base stations), similarly to [33]. Those points would help understand better the providers behavior in a competitive wireless network environment.

Acknowledgment

This work was supported by Euro-NF Network of Excellence and the French research agency through the CAPTURES project for both authors, and the WINEM project for the second author.

A. Proof of Proposition 1

Proof: We are looking for nonnegative values $(d_{1,A}^*, d_{1,B}^*, d_2^*)$ that maximize the objective function $\text{SW}(d_{1,A}, d_{1,B}, d_2)$ given in (3). To do so, we prove two intermediate results.

a) We can take $d_2^* = C_2$, since SW is nondecreasing in d_2 if $d_2 \leq C_2$, and nonincreasing if $d_2 \geq C_2$.

Indeed, remark that only the second term in the sum in (3) depends on d_2 , so we only focus on that term. Remark also that π_1 does not depend on d_2 , while $\pi_2 = \min(1, C_2/d_2)$.

- If $d_2 \geq C_2$ then $\pi_2 = C_2/d_2$, so that the second term of SW is $\frac{d_{1,B}\pi_1 + C_2}{d_{1,B} + d_2} V\left(\frac{d_{1,B} + d_2}{\alpha}\right)$. The valuation function V being concave with $V(0) = 0$, $z \mapsto V(z)/z$ is nonincreasing, which implies SW being nonincreasing in d_2 .
- If $d_2 \leq C_2$ then the second term of SW is $\frac{d_{1,B}\pi_1 + d_2}{d_{1,B} + d_2} V\left(\frac{d_{1,B} + d_2}{\alpha}\right)$, which is a product of two terms that are nondecreasing in d_2 , since $\pi_1 \leq 1$ and V is nondecreasing.

b) We can also take $d_{1,A}^* + d_{1,B}^* = C_1$:

- If $d_{1,A} + d_{1,B} \leq C_1$ then $\pi_1 = 1$, so that only the first term in (3) depends on $d_{1,A}$. Thus SW is nondecreasing due to the nondecreasingness of V , and consequently we can consider that $d_{1,A}^* + d_{1,B}^* \geq C_1$.
- If $d_{1,A} + d_{1,B} \geq C_1$, then $\pi_1 = \frac{C_1}{d_1}$ with $d_1 = d_{1,A} + d_{1,B}$. We then define $\beta := \frac{d_{1,A}}{d_1}$, so that $d_{1,A} = \beta d_1$ and $d_{1,B} = (1 - \beta)d_1$. The objective

function SW can then be written

$$(1 - \alpha) \frac{C_1}{d_1} V \left(\frac{\beta d_1}{1 - \alpha} \right) + \frac{(1 - \beta)C_1 + d_2 \pi_2}{(1 - \beta)d_1 + d_2} \alpha V \left(\frac{(1 - \beta)d_1 + d_2}{\alpha} \right),$$

where both terms in the sum are nonincreasing in d_1 because V is concave and $V(0) = 0$.

As a result, we can find some nonnegative values $(d_{1,A}^*, d_{1,B}^*, d_2^*)$ maximizing SW, and that are such that $d_2^* = C_2$ and $d_{1,A}^* + d_{1,B}^* = C_1$. Remark that $\pi_1 = \pi_2 = 1$ in that case. There just remains one parameter to find, say $d_{1,B}^*$ (since $d_{1,A}^* = C_1 - d_{1,B}^*$) to obtain the maximum value of social welfare. That value is thus the solution of the problem

$$\begin{aligned} \max_y \quad & f(y) \\ \text{s.t.} \quad & 0 \leq y \leq C_1, \end{aligned}$$

where $f(y) := (1 - \alpha)V \left(\frac{C_1 - y}{1 - \alpha} \right) + \alpha V \left(\frac{C_2 + y}{\alpha} \right)$ is differentiable. The marginal valuation function v being strictly decreasing, $f'(y) = v \left(\frac{C_2 + y}{\alpha} \right) - v \left(\frac{C_1 - y}{1 - \alpha} \right)$ verifies

- $f'(y) > 0 \Leftrightarrow \frac{C_1 - y}{1 - \alpha} > \frac{C_2 + y}{\alpha}$, and
- $f'(y) < 0 \Leftrightarrow \frac{C_1 - y}{1 - \alpha} < \frac{C_2 + y}{\alpha}$.

Using the constraints over y , function f reaches its maximum at $y = \max(0, \alpha C_1 - (1 - \alpha)C_2)$.

We therefore have proved the proposition, by just inserting the values in the expression of social welfare. \square

B. Proof of Proposition 2

Proof: A very general proof of the existence of a Nash equilibrium in nonatomic games (what we call here a Wardrop equilibrium) was provided by Schmeidler [32]. Therefore a solution to the system (5)-(10) exists.

We now establish the uniqueness of perceived prices at a user equilibrium. We will use the fact that (5) and (6) respectively imply

$$\begin{aligned} \min\left(1, \frac{C_1}{d_{1,A} + d_{1,B}}\right) &= \frac{p_1}{\bar{p}_1} \\ \text{and} \quad \min\left(1, \frac{C_2}{d_2}\right) &= \frac{p_2}{\bar{p}_2}. \end{aligned}$$

Assume two user equilibria $(d_{1,A}, d_{1,B}, d_2)$ and $(\tilde{d}_{1,A}, \tilde{d}_{1,B}, \tilde{d}_2)$ with different perceived prices (\bar{p}_1, \bar{p}_2) and $(\tilde{p}_1, \tilde{p}_2)$ exist for a given price profile (p_1, p_2) , and suppose that $\tilde{p}_1 > \bar{p}_1$. Then (5) implies that

$$\tilde{d}_{1,A} \frac{p_1}{\tilde{p}_1} + \tilde{d}_{1,B} \frac{p_1}{\tilde{p}_1} = C_1 \geq d_{1,A} \frac{p_1}{\bar{p}_1} + d_{1,B} \frac{p_1}{\bar{p}_1}. \quad (15)$$

On the other hand, (7) yields² $\tilde{d}_{1,A} \frac{p_1}{\tilde{p}_1} < d_{1,A} \frac{p_1}{\bar{p}_1}$, therefore (15) gives

$$\tilde{d}_{1,B} \frac{p_1}{\tilde{p}_1} > \frac{p_1}{\bar{p}_1} d_{1,B}. \quad (16)$$

Thus $\tilde{d}_{1,B} > 0$, and from (9) we have $\tilde{p}_1 \leq \tilde{p}_2$. Now applying (8) twice gives

$$\tilde{d}_{1,B} \frac{p_1}{\tilde{p}_1} + \tilde{d}_2 \frac{p_2}{\tilde{p}_2} = \alpha D(\tilde{p}_1) < \alpha D(\bar{p}_1) \leq \alpha D(\min(\bar{p}_1, \bar{p}_2)) = d_{1,B} \frac{p_1}{\bar{p}_1} + d_2 \frac{p_2}{\bar{p}_2}.$$

Relation (16) then yields

$$\tilde{d}_2 \frac{p_2}{\tilde{p}_2} < d_2 \frac{p_2}{\bar{p}_2} \leq C_2, \quad (17)$$

²Notice that $D(\tilde{p}_1) > 0$, otherwise one can check that we would get $\tilde{d}_{1,A} = \tilde{d}_{1,B} = 0$, a contradiction with (15). Therefore D is strictly decreasing on $[\bar{p}_1, \tilde{p}_1]$ and $D(\tilde{p}_1) < D(\bar{p}_1)$.

where the second inequality comes from (6). This implies that $d_2 > 0$, and thus $\bar{p}_2 \leq \bar{p}_1$ from (10). Summarizing our results we get $\bar{p}_2 \leq \bar{p}_1 < \tilde{p}_1 \leq \tilde{p}_2$, thus $\tilde{p}_2 > \bar{p}_2$. From (6) this means $\tilde{d}_2 > C_2$ and therefore $\tilde{d}_2 \frac{p_2}{\tilde{p}_2} = C_2$, a contradiction with (17). Therefore the perceived price \bar{p}_1 is unique.

Likewise, knowing that $\tilde{p}_1 = \bar{p}_1$, assume $\tilde{p}_2 > \bar{p}_2$. Then from (6) we get $\tilde{d}_2 = C_2 \frac{\tilde{p}_2}{p_2}$. Therefore $\tilde{d}_2 > 0$, and (10) implies $\tilde{p}_2 \leq \tilde{p}_1 = \bar{p}_1$, thus

$$\bar{p}_2 < \bar{p}_1. \quad (18)$$

Now applying (8) we obtain³

$$\tilde{d}_{1,B} \frac{p_1}{\tilde{p}_1} + \underbrace{\tilde{d}_2 \frac{p_2}{\tilde{p}_2}}_{=C_2} = \alpha D(\tilde{p}_2) < \alpha D(\bar{p}_2) \leq \alpha D(\min(\bar{p}_1, \bar{p}_2)) = d_{1,B} \frac{p_1}{\bar{p}_1} + \underbrace{d_2 \frac{p_2}{\bar{p}_2}}_{\leq C_2},$$

therefore $d_{1,B} > 0$, and thus $\bar{p}_1 \leq \bar{p}_2$ from (9), which is a contradiction with (18) and proves the uniqueness of \bar{p}_2 . \square

C. Proof of Proposition 3

We distinguish the two cases that appeared when computing the welfare maximizing situation.

C.1. Case $\alpha C_1 \leq (1 - \alpha)C_2$

Lemma 2. *Consider that Assumption A holds, and assume that $\alpha C_1 \leq (1 - \alpha)C_2$. For any price $p_1 > 0$, any price $p_2 \in (0, p_2^*]$ ensures provider 2 a revenue $R_2 = p_2^* C_2$, while any other price yields a strictly lower revenue.*

³Again, since $\tilde{d}_2 > 0$, from (8) we are in the zone where D is strictly positive, thus strictly decreasing.

Proof: We consider a strictly positive price p_1 and a price $p_2 \in (0, p_2^*]$, and we proceed in several steps to prove that $R_2 = p_2^* C_2$.

1. We have $\bar{p}_1 \geq p_2^*$: if not, (5) and (7) would imply

$$C_1 \geq d_{1,A} \frac{p_1}{\bar{p}_1} = (1 - \alpha)D(\bar{p}_1) > (1 - \alpha)D(p_2^*) \geq (1 - \alpha)D(p_1^*) = C_1,$$

a contradiction.

2. Also, $\bar{p}_2 \geq p_2^*$: otherwise from step 1 and (9) we would have $d_{1,B} = 0$, and (8) and (6) would give

$$C_2 \geq d_2 \frac{p_2}{\bar{p}_2} = \alpha D(\bar{p}_2) > \alpha D(p_2^*) = C_2,$$

another contradiction.

3. But on the other hand, $\bar{p}_2 \leq p_2^*$: otherwise we would have $\bar{p}_2 > p_2$, and thus $d_2 p_2 / \bar{p}_2 = C_2$ from (6). Then (10) would yield $\bar{p}_1 \geq \bar{p}_2$, and applying (8) would give

$$d_{1,B} \frac{p_1}{\bar{p}_1} + C_2 = \alpha D(\bar{p}_2) < \alpha D(p_2^*) = C_2,$$

another contradiction. As a consequence of this and of previous result, $\bar{p}_2 = p_2^*$.

4. Finally, $R_2 = p_2^* C_2$: we use results from the previous steps, and distinguish two cases.

- if $\bar{p}_1 = p_2^*$, then adding (7) and (8) gives

$$\underbrace{d_1 \frac{p_1}{\bar{p}_1}}_{\leq C_1 \text{ from (5)}} + \underbrace{d_2 \frac{p_2}{\bar{p}_2}}_{\leq C_2 \text{ from (6)}} = \alpha D(p_2^*) + (1 - \alpha)D(p_2^*) \geq C_1 + C_2,$$

thus all inequalities are equalities, and in particular $d_2 p_2 = p_2^* C_2$.

- if $\bar{p}_1 > p_2^*$ then (9) and (8) directly give $p_2 d_2 = p_2^* C_2$.

Now we prove that any price $p_2 > p_2^*$ corresponds to a revenue $R_2 < p_2^* C_2$.

From (10):

- either $d_2 = 0$, and therefore $R_2 = 0$;
- or $\bar{p}_2 \leq \bar{p}_1$, which from (8) implies that

$$d_2 p_2 / \bar{p}_2 \leq \alpha D(\bar{p}_2) \leq \alpha D(p_2) < \alpha D(p_2^*) = C_2,$$

and from (6) yields $\bar{p}_2 = p_2$. Then applying (8) again, we have $d_2 \leq \alpha D(p_2)$, and

$$R_2 \leq \alpha p_2 D(p_2) < \alpha p_2^* D(p_2^*) = p_2^* C_2,$$

where the last inequality comes from Assumption A. This concludes the proof. □

Lemma 3. *Consider that Assumption A holds. For any fixed price $p_2 \in (0, p_2^*]$, any price $p_1 \in (0, p_1^*]$ ensures provider 1 a revenue $R_1 = p_1^* C_1$, while any other price yields a strictly lower revenue.*

Proof: Fix $p_2 \in (0, p_2^*]$. As seen in the proof of Lemma 2 we have $\bar{p}_2 = p_2^*$ and $d_2 p_2 / \bar{p}_2 = C_2$ whatever the value of p_1 , which from (10) gives $\bar{p}_2 \leq \bar{p}_1$, and from (8) implies that $d_{1,B} = 0$. As a result, the total demand for provider 1 is in zone A, and is given by (7). Then,

- If provider 1 sets $p_1 \leq p_1^*$, then (5) and (7) give

$$\underbrace{C_1}_{(1-\alpha)D(p_1^*)} \geq d_1 \frac{p_1}{\bar{p}_1} = (1-\alpha)D(\bar{p}_1), \quad (19)$$

thus $\bar{p}_1 \geq p_1^*$. But actually $\bar{p}_1 = p_1^*$, otherwise (5) and (7) would give

$$C_1 = d_1 \frac{p_1}{\bar{p}_1} = (1 - \alpha)D(\bar{p}_1) < (1 - \alpha)D(p_1^*) = C_1,$$

a contradiction. As a result, from (19) we have $R_1 = p_1 d_1 = p_1^* C_1$.

• If provider 1 sets $p_1 > p_1^*$, then $\bar{p}_1 > p_1^*$ from (5), and (7) yields

$$d_1 \frac{p_1}{\bar{p}_1} = (1 - \alpha)D(\bar{p}_1) > (1 - \alpha)D(p_1^*) = C_1,$$

thus from (5), $d_1 < C_1$ and $\bar{p}_1 = p_1$. As a result, (5) implies

$$d_1 p_1 = (1 - \alpha)p_1 D(p_1) < (1 - \alpha)p_1^* D(p_1^*) = p_1^* C_1,$$

where the inequality comes from Assumption A. □

The fact that any price (p_1, p_2) with $p_i \in (0, p_i^*]$, $i = 1, 2$ is a Nash equilibrium of the price game is a direct consequence of Lemmas 2 and 3.

C.1.1. Case $\alpha C_1 > (1 - \alpha)C_2$

Recall that in that case we have

$$\frac{C_2}{\alpha} < D(p^*) = C_1 + C_2 < \frac{C_1}{1 - \alpha}. \quad (20)$$

Lemma 4. *Consider that Assumption A holds. All price profiles $(p_1, p_2) \in (0, p^*]^2$ form a Nash equilibrium. Those profiles are the only Nash equilibria of the pricing game, and the corresponding revenue for each provider $i = 1, 2$ is $R_i = p^* C_i$.*

Proof: The proof follows three steps:

1. We first prove that when both providers set a price below p^* then each provider k gets a revenue $p^* C_k$.

2. Then we show that if only one provider j sets a price $p_j > p^*$ he gets a strictly smaller revenue, while his opponent i gets at least the same revenue p^*C_i .
3. Finally, we prove that if both providers were to set a price strictly above p^* , then at least one provider i would obtain strictly less than p^*C_i , and thus from the previous point he would be better off reducing his price below p^* .

Step 1. Consider a price profile (p_1, p_2) with $p_k \in (0, p^*]$, $k = 1, 2$. Then adding (7) and (8) gives

$$\underbrace{d_1 \frac{p_1}{\bar{p}_1}}_{\leq C_1 \text{ from (5)}} + \underbrace{d_2 \frac{p_2}{\bar{p}_2}}_{\leq C_2 \text{ from (6)}} \geq D(\bar{p}_1),$$

thus $\bar{p}_1 \geq p^*$, due to the nonincreasingness of D . Now, we also have $\bar{p}_2 \geq p^*$, otherwise (9) would imply $d_{1,B} = 0$, and (6) and (8) would give

$$C_2 \geq d_2 \frac{p_2}{\bar{p}_2} = \alpha D(\bar{p}_2) > \alpha D(p^*) > C_2,$$

a contradiction. Now we prove that we actually have $\bar{p}_1 = \bar{p}_2 = p^*$. Assume $\bar{p}_1 > p^*$: then $d_1 p_1 / \bar{p}_1 = C_1$ from (5), and adding (7) and (8) would give

$$C_1 + d_2 \frac{p_2}{\bar{p}_2} < D(p^*) = C_1 + C_2,$$

thus $d_2 p_2 / \bar{p}_2 < C_2$, and $\bar{p}_2 = p_2$ from (6). Since $p_2 \leq p^*$ and $\bar{p}_2 \geq p^*$, we would have $\bar{p}_2 = p^*$. But then (10) and (7) would imply

$$C_1 = (1 - \alpha)D(\bar{p}_1) < (1 - \alpha)D(p^*) < C_1,$$

a contradiction. We therefore have $\bar{p}_1 = p^*$, which implies $\bar{p}_2 \leq p^*$ (otherwise (10) and (6) would lead to a contradiction). Summarizing, we have $\bar{p}_1 = \bar{p}_2 = p^*$. Adding again (7) and (8) now gives

$$\underbrace{d_1 \frac{p_1}{p^*}}_{\leq C_1 \text{ from (5)}} + \underbrace{d_2 \frac{p_2}{p^*}}_{\leq C_2 \text{ from (6)}} = C_1 + C_2,$$

implying $d_1 p_1 / p^* = C_1$ and $d_2 p_2 / p^* = C_2$, which gives $d_k p_k = p^* C_k$ for $k = 1, 2$ and establishes the first step of the proof.

Step 2. Now consider a provider i setting $p_i \in (0, p^*]$, while his opponent j sets a price $p_j > p^*$. Then we prove that $R_j < p^* C_j$ and $R_i \geq p^* C_i$.

- If $\bar{p}_i < \bar{p}_j$, then using (9) or (10), and adding (7) and (8), we have

$$\underbrace{d_i \frac{p_i}{\bar{p}_i}}_{\leq C_i \text{ from (1)}} = D(\bar{p}_i).$$

Thus $\bar{p}_i \geq v(C_i) > p^* \geq p_i$, and therefore from Lemma 1, $R_i = \bar{p}_i C_i > p^* C_i$. To study R_j we distinguish two cases.

- Case $i = 1$: (10) directly gives $R_2 = 0$;
- Case $i = 2$: (9) implies $d_{1,B} = 0$, and (7) yields

$$d_1 p_1 = (1 - \alpha) \bar{p}_1 D(\bar{p}_1) < (1 - \alpha) p^* D(p^*) < p^* C_1,$$

where the first inequality comes from Assumption A, and the second one from (20).

- If $\bar{p}_i \geq \bar{p}_j$, then from (1) this implies $d_i p_i / \bar{p}_i = C_i$ and directly gives $R_i > p^* C_i$. Moreover, adding (7) and (8) we obtain

$$d_j \frac{p_j}{\bar{p}_j} + C_i \leq D(\bar{p}_j),$$

which gives

$$d_j p_j \leq \bar{p}_j (D(\bar{p}_j) - C_i) < p^* (D(p^*) - C_i) = p^* C_j,$$

where we used Assumption A and the fact that $\bar{p}_j > p^*$.

This concludes the second step of the proof.

Step 3. Assume now that both providers set a price strictly above p^* . We index the providers such that $\bar{p}_i \geq \bar{p}_j$ at the Wardrop equilibrium.

- If $\bar{p}_i > \bar{p}_j$ then $R_i < p^* C_i$ for the same reasons as in the previous step ($R_i = 0$ if $i = 2$, and $R_i = (1 - \alpha) p^* D(p^*) < p^* C_1$ if $i = 1$).
- If $\bar{p}_i = \bar{p}_j$, then adding (7) and (8) we have

$$d_i \frac{p_i}{\bar{p}_i} + d_j \frac{p_j}{\bar{p}_j} = \frac{R_i + R_j}{\bar{p}_i} = D(\bar{p}_i),$$

and from Assumption A we obtain

$$R_i + R_j = \bar{p}_i D(\bar{p}_i) < p^* D(p^*) = p^* (C_i + C_j),$$

which implies that either $R_i < p^* C_i$, or $R_j < p^* C_j$.

Those three steps completely characterize the Nash equilibria of the pricing game: if a provider (or both) sets his price strictly above p^* then at least one provider is strictly better off reducing his price, while when both providers set their price below p^* no provider can strictly improve his revenue by a unilateral price change. \square

References

- [1] IEEE Standard for Information technology-Telecommunications and information exchange between systems-Local and metropolitan area networks-Specific requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, IEEE Std 802.11-2007 (Revision of IEEE Std 802.11-1999) (2007) C1–1184.
- [2] IEEE 802.16-2004, IEEE Standard for Local and Metropolitan Area Networks - Part 16: Air Interface for Fixed Broadband Wireless Access Systems (October 2004).
- [3] R. Piggin, WiMAX in depth: Broadband wireless access, *IEEE Communications Engineer* 2 (5) (2004) 36–39.
- [4] M. Osborne, A. Rubenstein, *A Course on Game Theory*, MIT Press, 1994.
- [5] J. Wardrop, Some theoretical aspects of road traffic research, *Proc. of the Institute of Civil Engineers* 1 (1952) 325–378.
- [6] E. Koutsoupias, C. Papadimitriou, Worst-case equilibria, in: *Proc. of 16th Annual Symposium on Theoretical Aspects of Computer Science (STACS 1999)*, Vol. 1563 of *Lecture Notes in Computer Science*, 1999, pp. 404–413.
- [7] E. Anshelevich, A. Dasgupta, J. Kleinberg, E. Tardos, T. Wexler, T. Roughgarden, The price of stability for network design with fair cost allocation, in: *Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2004, pp. 295–304.

- [8] E. Altman, T. Boulogne, El-Azouzi, T. Jiménez, L. Wynter, A survey on networking games in telecommunications, *Computers and Operations Research* 33 (2) (2006) 286–311.
- [9] C. Courcoubetis, R. Weber, *Pricing Communication Networks—Economics, Technology and Modelling*, Wiley, 2003.
- [10] L. DaSilva, Pricing of QoS-Enabled Networks: A Survey, *IEEE Communications Surveys & Tutorials* 3 (2).
- [11] B. Tuffin, Charging the Internet without bandwidth reservation: an overview and bibliography of mathematical approaches, *Journal of Information Science and Engineering* 19 (5) (2003) 765–786.
- [12] R. Gibbens, R. Mason, R. Steinberg, Internet service classes under competition, *IEEE Journal on Selected Areas in Communications* 18 (12) (2000) 2490–2498.
- [13] R. El-Azouzi, E. Altman, L. Wynter, Telecommunications network equilibrium with price and quality-of-service characteristics, in: *Proc. of 18th International Teletraffic Congress, Berlin, Germany, 2003*.
- [14] Z. Liu, L. Wynter, C. Xia, Pricing information services in a competitive market: avoiding price wars, *Tech. rep., INRIA* (2002).
- [15] O. Ileri, D. Samardzija, T. Sizer, N. Mandayam, Demand responsive pricing and competitive spectrum allocation via a spectrum server, in: *Proc. of IEEE DySpan 2005, 2006*.

- [16] M. Felegyhazi, J. Hubaux, Wireless operators in a shared spectrum, in: Proc. of IEEE INFOCOM, 2006.
- [17] Y. Xing, R. Chandramouli, C. Cordeiro, Price dynamics in competitive agile spectrum access markets, IEEE Journal on Selected Areas in Communications 25 (3) (2007) 613–621.
- [18] F. Bernstein, A. Federgruen, A general equilibrium model for industries with price and service competition, Management Science 52 (6) (2004) 868–886.
- [19] M. Manshaei, J. Freudiger, M. Felegyhazi, P. Marbach, J.-P. Hubaux, On wireless social community networks, in: Proc. of IEEE INFOCOM, Phoenix, AZ, USA, 2008.
- [20] A. Zemlianov, G. de Veciana, Cooperation and decision-making in a wireless multi-provider setting, in: Proc. of IEEE INFOCOM, Miami, FL, USA, 2005.
- [21] D. Acemoglu, A. Ozdaglar, Competition and efficiency in congested markets, Mathematics of Operations Research.
- [22] D. Acemoglu, A. Ozdaglar, Price competition in communication networks, in: Proc. of IEEE INFOCOM, Barcelona, Spain, 2006.
- [23] A. Hayrapetyan, E. Tardos, T. Wexler, A network pricing game for selfish traffic, in: Proc. of IEEE PODC, 2006.
- [24] L. He, J. Walrand, Pricing internet services with multiple providers, in:

Proc. of the 41st Allerton Conference on Communication, Control and Computing, 2003.

- [25] L. He, J. Walrand, Pricing and revenue sharing strategies for internet service providers, in: Proc. of IEEE INFOCOM, 2005.
- [26] S. Shakkottai, R. Srikant, Economics of network pricing with multiple ISPs, in: Proc. of IEEE INFOCOM 2005, Miami, FL, USA, 2005.
- [27] P. Maillé, B. Tuffin, Analysis of price competition in a slotted resource allocation game, in: Proc. of IEEE INFOCOM 2008, Phoenix, AZ, USA, 2008.
- [28] P. Marbach, Analysis of a static pricing scheme for priority services, IEEE/ACM Transactions on Networking 12 (2).
- [29] R. Cole, Y. Dodis, T. Roughgarden, How much can taxes help selfish routing?, Journal of Computer and System Sciences 72 (3) (2006) 444–467.
- [30] G. Perakis, The “Price of Anarchy” Under Nonlinear and Asymmetric Costs, Mathematics of Operations Research 32 (3) (2007) 614–628.
- [31] R. J. Aumann, L. S. Shapley, Values of Non-Atomic Games, Princeton University Press, 1974.
- [32] D. Schmeidler, Equilibrium points of nonatomic games, Journal of Statistical Physics 7 (4) (1973) 295–300.
- [33] P. Maillé, B. Tuffin, Optimization of transmission power in competitive wireless networks, in: P. Reichl, B. Stiller, B. Tuffin (Eds.), Network

Economics for Next Generation Networks. Proceedings of the sixth International Workshop on Advanced Internet Charging and QoS technologies (ICQT'09), Vol. 5539 of Lecture Notes in Computer Science, Springer Verlag, 2009, pp. 2–10.

Price war with migrating customers

Patrick Maillé

TELECOM Bretagne
2, rue de la Châtaigneraie
CS 17607

35576 Cesson Sévigné Cedex, France

Email: patrick.maill@telecom-bretagne.eu

Maurizio Naldi

Università di Roma Tor Vergata
Dip. di Informatica Sistemi Produzione
Via del Politecnico 1

00133 Roma, Italy

Email: naldi@disp.uniroma2.it

Bruno Tuffin

INRIA Rennes - Bretagne Atlantique
Campus Universitaire de Beaulieu
35042 Rennes Cedex, France

Email: Bruno.Tuffin@inria.fr

Abstract—In the telecommunication world, competition among providers to attract and keep customers is fierce. On the other hand customers churn between providers due to better prices, better reputation or better services. We propose in this paper to study the price war between two providers in the case where users' decisions are modeled by a Markov chain, with price-dependent transition rates. Each provider is assumed to look for a maximized revenue, which depends on the strategy of the competitor. Therefore, using the framework of non-cooperative game theory, we show how the price war can be analyzed and show the influence of various parameters.

I. INTRODUCTION

The migration of customers from a service provider to another, a.k.a. as *churn*, has become a relevant phenomenon since the liberalization of telecommunications service and the ensuing proliferation of network operators. Churn is especially large in mobile networks, where yearly migration rates as high as 25% are not uncommon [1]. The migration of each customer, to the benefit of another service provider, implies both the loss of the stream of future revenues associated to that customer and of the acquisition cost. Service providers are therefore very keen on retaining their customers as well as on attracting new ones. In doing so they can rely both on preventive and on reactive strategies. An example of the latter is given by unfair practices such as the malicious introduction of delays in the migration process [2] [3]. Preventive strategies rely instead on the identification of the factors having a major influence on the churning decision (the churn determinants) [4] [5] and on successive actions on those factors. Among the identified churn determinants, price plays the most relevant role. Price always appears as a major factor: in the context of mobile number portability (a mechanism allowing to switch provider with minimal discontinuity, since the telephone number is retained), price is stated as a key element in spurring churn [6]. Another example is provided in [7], where retention and attrition phenomena are studied in an experimental setting by proposing different pricing plans to test customers. We can then expect that providers may compete for retaining customers by acting primarily on price. In this paper we propose a model for the competition among service providers based on price, the competition being here limited to two service providers. To our knowledge, it is the first time that such a competition model is introduced in telecommunications to model the price war. While most

research efforts on telecommunication pricing are concerned with congestion externality for usage-based pricing [8], here we focus more on subscription-fee based pricing, where users are charged for the amount of time they stay with a provider, regardless of their usage (e.g., the Internet subscription fee incorporated in most of the current pricing packages). Our model makes use of a Markov chain to mimic the churn behaviour of a customer in terms of prices and other parameters. Basically, the user can be with any of the providers or none of them if not satisfied with their combination of price and services. The per-user revenue of each provider can then be easily computed from steady-state probabilities, considering a single user without loss of generality. Indeed, assuming that customers behave independently and according to the same Markov chain, the expected revenue of providers is exactly the expected revenue per customer times the total population. Those state probabilities depending on both prices, so are the revenues of providers. As a consequence, the natural framework for analyzing the competition between providers seeking to maximize their revenue is non-cooperative game theory. We show how to solve this game, and illustrate the influence of parameters such as impact of other churn determinants, and (social or financial) cost for not getting any service due to excessive prices.

The paper is organized as follows. Section II presents the Markov chain model representing the user behaviour and computes the associated steady-state probabilities. Section III explains in full generality why non-cooperative come into play and how it can be solved, either analytically or numerically. Section IV then shows how, in a simplified setting, the game can be solved analytically. Section V on the other hand makes use of a description of rates issued from logit models. In this case, a numerical analysis is performed. In addition, we investigate the sensitivity of the resulting equilibria to the degree of asymmetry between the two providers in attracting customers and to the relevance of the price factor for the customers. Finally, Section VI describes our conclusions.

II. MARKOV CHAIN MODEL OF USERS' BEHAVIOR

We assume that the behavior of a customer is represented by the continuous time Markov chain¹ that is depicted in

¹We therefore implicitly assume that all events leading to a state change occur after an exponentially distributed time.

$$Q := \begin{pmatrix} -2\alpha & & & \\ \lambda_{10}(p_1, p_2) & -(\lambda_{10}(p_1, p_2) + \lambda_{12}(p_1, p_2)) & & \\ \lambda_{20}(p_1, p_2) & \lambda_{21}(p_1, p_2) & & \\ & & \lambda_{12}(p_1, p_2) & \\ & & -(\lambda_{20}(p_1, p_2) + \lambda_{21}(p_1, p_2)) & \end{pmatrix}.$$

Figure 2: Infinitesimal generator of the Markov chain

Figure 1². Here state 1 means that the customer is with provider 1, state 2 that he is with provider 2 and state 0 that he does not use any service. The parameter α is a constant rate independent of prices. Remark on the other hand that other rates will be considered in the numerical section, but the resulting steady-state probabilities will be computed in the same way. The resulting infinitesimal generator Q is given in

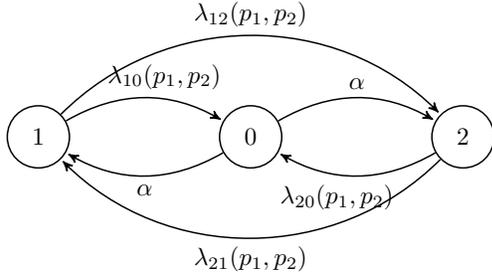


Figure 1: Continuous time Markov chain model of the customer's switching behaviour

Figure 2. From standard Markov chain analysis, the steady-state probabilities for each of the three states grouped in the row vector $\pi = (\pi_i)_{i=0, \dots, 2}$ exist and are given by the solution of equations

$$\pi Q = 0, \quad \sum_{i=0}^2 \pi_i = 1.$$

If

$$c = \alpha (2\lambda_{12}(p_1, p_2) + 2\lambda_{21}(p_1, p_2) + \lambda_{10}(p_1, p_2) + \lambda_{20}(p_1, p_2)) + \lambda_{10}(p_1, p_2)\lambda_{21}(p_1, p_2) + \lambda_{20}(p_1, p_2)\lambda_{12}(p_1, p_2) + \lambda_{10}(p_1, p_2)\lambda_{20}(p_1, p_2),$$

we have

$$\begin{aligned} \pi_0 &= \frac{\lambda_{10}(p_1, p_2)\lambda_{21}(p_1, p_2) + \lambda_{10}(p_1, p_2)\lambda_{20}(p_1, p_2)}{c} \\ &\quad + \frac{\lambda_{12}(p_1, p_2)\lambda_{20}(p_1, p_2)}{c} \\ \pi_1 &= \frac{(\lambda_{20}(p_1, p_2) + 2\lambda_{12}(p_1, p_2))\alpha}{c} \\ \pi_2 &= \frac{\alpha(\lambda_{10}(p_1, p_2) + 2\lambda_{21}(p_1, p_2))}{c}. \end{aligned}$$

²Remark that other distributions for sojourn times can be used as well. In that case the model requires to be handled by simulation, while here steady-state probabilities can be computed analytically and as a consequence the game is solved using simple numerical analysis tools.

III. NON-COOPERATIVE GAME FROM THE PROVIDERS' SIDE

The previous section describes the behaviour of a customer as a function of prices set by providers. The question is now to define the best pricing strategy for each provider knowing that behaviour. We therefore have a so-called Stackelberg game [9], with leaders (the providers) choosing their prices knowing the consequences they would have on users' behaviour, and the followers (the users) whose reaction is a direct consequence of providers' prices. This means that providers play first, but using backward induction, they anticipate the resulting strategy of end users who actually make the last move.

It is important to stress that our model, considering a single customer in front of two providers, is sufficient if assuming that each user has a behaviour independent of others'. The case of N users can then indeed be easily derived by multiplying the expected revenue by N (thanks to the independence).

In the first step of the Stackelberg game, each provider tries to maximize its revenue. There is a trade-off to be analyzed between the fact that increasing the price will increase the revenue per customer, but on the other hand potentially reduce the number of customers (i.e., the probability of having the user as customer in our case). The revenue per customer R_i for provider $i \in \{1, 2\}$ is therefore expressed formally as the price charged multiplied by the probability that this customer is indeed with provider i , i.e., $R_i = p_i \pi_i \forall i \in \{1, 2\}$, or more exactly using the expressions for the steady-state probabilities of the Markov chain:

$$\begin{aligned} R_1(p_1, p_2) &= p_1 \frac{\alpha(\lambda_{20}(p_1, p_2) + 2\lambda_{12}(p_1, p_2))}{c} \\ R_2(p_1, p_2) &= p_2 \frac{\alpha(\lambda_{10}(p_1, p_2) + 2\lambda_{21}(p_1, p_2))}{c}. \end{aligned}$$

From those expressions, it is clear that the revenue of a provider depends on the price strategy of the concurrent. Indeed, steady-state probabilities are functions of rates which themselves depend on both prices. As a consequence, this fits the framework of *non-cooperative game theory* [9]. Each provider strives to find its best strategy, i.e., its price maximizing its revenue, which can be modified by the strategy of the competitor. The solution concept is that of a *Nash equilibrium*: a Nash equilibrium is a price profile (p_1^*, p_2^*) such that no provider can unilaterally increase its revenue, i.e.,

$$\begin{aligned} R_1(p_1^*, p_2^*) &= \max_{p_1 \geq 0} R_1(p_1, p_2^*) \\ R_2(p_1^*, p_2^*) &= \max_{p_2 \geq 0} R_2(p_1^*, p_2). \end{aligned}$$

In general the existence of a Nash equilibrium cannot be ensured without assumptions, nor its uniqueness when existence is shown. In the case where rate functions are simple enough in terms of prices, we may find the form of the Nash equilibria analytically (see next section). Otherwise, the computations can be performed numerically using the following algorithm. We define the best response of each provider as a function of the strategy of its opponent by

$$\begin{aligned} \text{BR}_1(p_2) &:= \arg \max_{p_1 \geq 0} R_1(p_1, p_2) \text{ and} \\ \text{BR}_2(p_1) &:= \arg \max_{p_2 \geq 0} R_2(p_1, p_2). \end{aligned}$$

In this setting, a Nash equilibrium is just a point (p_1^*, p_2^*) such that $\text{BR}_1(p_2^*) = p_1^*$ and $\text{BR}_2(p_1^*) = p_2^*$ (if best responses are not unique, it means that p_1^* in the set of best responses when provider's price is p_2^* , and reciprocally). Algorithm 1 describes how to algorithmically and graphically determine Nash equilibria (if any)

Alg. 1 Graphically finding the Nash equilibria of the game

Input: transition rates of the Markov chain in Figure 1, as functions of prices p_1 and p_2 .

- 1) For all possible values of $p_2 \geq 0$, find the set $\text{BR}_1(p_2)$ of p_1 values maximizing $R_1(p_1, p_2)$.
 - 2) For all possible values of $p_1 \geq 0$, find the set $\text{BR}_2(p_1)$ of p_2 values maximizing $R_2(p_1, p_2)$.
 - 3) On a same graphic, plot the best response functions $p_1 = \text{BR}_1(p_2)$ and $p_2 = \text{BR}_2(p_1)$, as illustrated Figure 3.
 - 4) The set of Nash equilibria is the (possibly empty) set of intersection points of those functions.
-

Two practical remarks can be made:

- Instead of assuming the set $[0, \infty)$ for each price, we will limit ourselves to $[0, p_{max}]$ since customers are unlikely to come to the provider if price is too high.
- When analytical derivation cannot be performed and solved to determine the best response functions, only a finite number of values can be tried in practice in each case, and the best responses determined at these points, and the solution investigated on the corresponding lattice.

IV. ANALYSIS FOR SPECIFIC BUT REALISTIC RATES

This section is dedicated to the case where rates of the Markov chain are simple, but realistic enough, to derive analytical expressions for the Nash equilibria. We more specifically assume that

$$\begin{aligned} \lambda_{10}(p_1, p_2) &= p_1 \\ \lambda_{20}(p_1, p_2) &= p_2 \\ \lambda_{12}(p_1, p_2) &= \zeta p_1 / p_2 \\ \lambda_{21}(p_1, p_2) &= p_2 / p_1, \end{aligned}$$

with ζ a strictly positive real number. Expressions for λ_{10} and λ_{20} mean that a customer is more likely to leave a provider

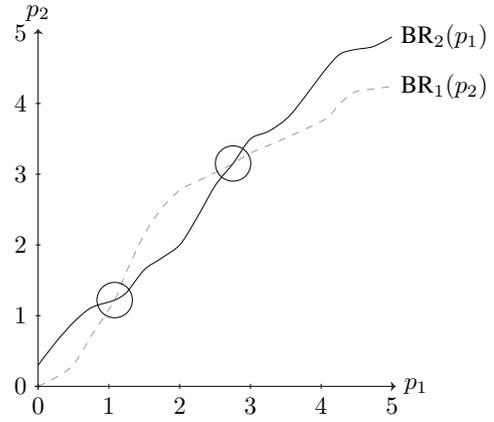


Figure 3: Graphical determination of Nash equilibria. Two solutions here (circled points).

for no service if its price is high, and depend linearly (and only) on the price at the incumbent provider. Values for λ_{12} and λ_{21} mean that swapping between providers depend on the ratios of prices. The insertion of parameter a is to introduce some asymmetry, because provider 1 may have a better (worse) reputation and it is therefore less (more) likely to be left for provider 2 when $\zeta < 1$ ($\zeta > 1$).

In that case, we end up, after simple computations, with

$$\begin{aligned} R_1 &= \frac{\alpha p_2^2 p_1 (p_1 + 2)}{2\alpha p_2^2 + (\alpha + 1)p_2 p_1^2 + (1 + \alpha)p_1 p_2^2 + p_1^2 p_2^2 + 2\zeta \alpha p_1^2} \\ R_2 &= \frac{\alpha p_1^2 p_2 (p_2 + 2\zeta)}{2\alpha p_2^2 + (\alpha + 1)p_2 p_1^2 + (1 + \alpha)p_1 p_2^2 + p_1^2 p_2^2 + 2\zeta \alpha p_1^2}. \end{aligned}$$

In order to determine the existence of a Nash equilibrium, we compute the derivatives. Using $D = 2\alpha p_2^2 + (\alpha + 1)p_2 p_1^2 + (1 + \alpha)p_1 p_2^2 + p_1^2 p_2^2 + 2\zeta \alpha p_1^2$, we have

$$\begin{aligned} \frac{\partial R_1}{\partial p_1} &= -\alpha p_2^2 \left[\frac{p_1^2 ((1 - \alpha)p_2^2 + 2(\alpha + \zeta)p_2 + 4\zeta\alpha)}{D^2} \right. \\ &\quad \left. + \frac{p_1(-4\alpha p_2^2) + (-4\alpha p_2^2)}{D^2} \right] \\ \frac{\partial R_2}{\partial p_2} &= \alpha p_1^2 \left[\frac{p_2^2 (-2\zeta(1 + \alpha)p_1 + (\alpha - \zeta)p_1^2 - 4\zeta\alpha)}{D^2} \right. \\ &\quad \left. + \frac{p_2(4\zeta\alpha p_1^2) + (4\zeta^2\alpha p_1^2)}{D^2} \right]. \end{aligned}$$

Note that $(p_1 = 0, p_2 = 0)$ is always a solution of the system $\frac{\partial R_1}{\partial p_1} = 0$ and $\frac{\partial R_2}{\partial p_2} = 0$, and a Nash equilibrium. On the other hand, equating the numerators to zero, any solution with $(p_1, p_2) \neq (0, 0)$, i.e., any non-degenerate strictly positive and finite Nash equilibrium, then solves the system of second degree equations

$$\begin{aligned} p_1^2 ((1 - \alpha)p_2^2 + 2(\alpha + \zeta)p_2 + 4\zeta\alpha) \\ + p_1(-4\alpha p_2^2) + (-4\alpha p_2^2) &= 0 \quad (1) \end{aligned}$$

$$\begin{aligned} p_2^2 (-2\zeta(1 + \alpha)p_1 + (\alpha - \zeta)p_1^2 - 4\zeta\alpha) \\ + p_2(4\zeta\alpha p_1^2) + (4\zeta^2\alpha p_1^2) &= 0. \quad (2) \end{aligned}$$

The unique strictly positive solution of this system is given by

$$p_1 = -2 \frac{\alpha(-\zeta^2 + \zeta + \alpha^2\zeta - \alpha^2)}{\zeta\alpha^3 - 2\alpha^2\zeta - \alpha^2 + 3\zeta\alpha - \alpha^2} \quad (3)$$

$$p_2 = 2 \frac{\zeta\alpha(-\zeta^2 + \zeta + \alpha^2\zeta - \alpha^2)}{3\zeta^2\alpha - \zeta^2 - \zeta^2\alpha^2 - 2\alpha^2\zeta + \alpha^3}, \quad (4)$$

provided those values are positive. At most one Nash equilibrium with strictly positive prices is possible in that case.

V. NUMERICAL RESULTS

A. Churn rates and prices in the literature

In Section II the transition rates that mark the passage from a provider to the other are shown to depend on the prices offered by the providers and Section IV illustrates that, in very simplified cases, the Nash equilibrium can be obtained analytically, though not easily. In order to adopt a model as close as possible to reality, we briefly review the related literature on the mathematical relationship between churn rates and prices in this sub-section, that which will be adopted during our numerical analysis.

Significant efforts have been spent to identify the most relevant factors in determining churn (often named churn determinants). In order to model the relationship between prices and churn rates in a quantitative fashion both parametric and non-parametric approaches have been proposed in the literature. Among the non-parametric approaches we can cite [10], where neural networks and decision trees are employed, and [11] where a novel evolutionary learning algorithm is proposed. Since we need a closed form relationship here we are more interested in parametric approaches. The most widespread model adopted in the literature to represent that relationship is the logit model, which employs a logistic probability distribution function [12] [13] [14]. The argument of the logistic function is a linear function of a number of churn determinants. The most general expression of the probability that a user churns in the next period (e.g., a year as in [12]) is then

$$\mathbb{P}_{\text{churn}} = \frac{1}{1 + e^{-I}}, \quad (5)$$

where I is the logit factor, in turn given by

$$I = \sum_{i=1}^n \beta_i X_i, \quad (6)$$

where X_i , $i = 1, \dots, n$ are the explanatory variables (churn determinants) and β_i , $i = 1, \dots, n$, are the coefficients representing the relative importance of those determinants. In this paper we have focussed on the price factor, so that we can group the impact of the other churn determinants in the overall term γ , arriving at the simpler expression

$$\mathbb{P}_{\text{churn}} = \frac{1}{1 + \gamma e^{-\beta p_1/p_2}} \quad (7)$$

for the probability that in the specified period the user switches from Provider 1 to Provider 2. We may employ that expression for a time period of any duration, so we can adopt it in the

Markov chain model described in Section II. We note that, according to expression (7), there is a non zero probability, namely $\mathbb{P}_{\text{churn}} = \frac{1}{1+\gamma}$, that the user switches provider due to the ensemble of other dissatisfaction factors, even when the service offered by the losing provider is free.

B. Transition rates

The transition rates that we assume now are chosen to reflect the conclusions of the previous subsection. However, the data and conclusions drawn from the literature do not provide us with a complete description of all the transition rates we need, in particular the state where users do not subscribe to the service is not encompassed in previous results. Moreover, the literature considers discretized time, whereas we focus here on a continuous-time model.

That latter difficulty is addressed here by assuming that time periods considered in Subsection V-A, are short with respect to the mean sojourn time in a given state. This implies that the discrete time transition probabilities are approximately the continuous-time transition rates multiplied by the period duration. Consequently, we would like to consider transition rates from state $i \in \{1, 2\}$ to state $j \in \{1, 2\} \setminus \{i\}$ of the form $\frac{\kappa}{1 + \gamma_i e^{-\beta p_i/p_j}}$, where $\kappa > 0$ represents the inverse of the period duration. Since β represents the user sensitivity to prices, we consider it is the same for the different states of the model. However, such an expression would imply that all transition rates be in an interval $[\kappa/(1 + \gamma_i), \kappa]$ regardless of the price values. This is not realistic, since it would imply that a provider could ensure an arbitrarily large revenue by setting a very large price. We therefore need that the transition rates to a provider i tend to 0 and/or that the rates from provider i tend to ∞ , when $p_i \rightarrow \infty$. To that end, we slightly modify the previous expression, and take transition rates of the form

$$\lambda_{ij}(p_i, p_j) = \frac{\kappa}{\gamma_i e^{-\beta p_i/p_j}} = \frac{\kappa}{\gamma_i} e^{\beta p_i/p_j}, \quad (8)$$

We introduce asymmetry among providers through the parameter γ_i : as explained before, this parameter encompasses the reasons other than price (e.g., Quality of Service, reputation, ...), why a user should leave state i .

We propose to address the former difficulty (no hints regarding the transitions to/from our state 0) by assuming that being in state 0 corresponds to perceiving a *cost* p_0 , that reflects the inconvenience for not benefitting from the service. We therefore treat state 0 as the two other states, but considering p_0 as a fixed value instead of a strategic variable. As a result, the transition rate we assume from any state $i \in \{0, 1, 2\}$ to state $j \in \{0, 1, 2\} \setminus \{i\}$ is given by (8).

The model parameters that we consider are then:

- the user sensitivity to price β ,
- the likeliness γ_i to stay in current state i , $i = 0, 1, 2$,
- the inverse of period duration κ (this parameter should not play a role in our model, since by a time unit change we can assume $\kappa = 1$),
- the user perceived cost p_0 for not benefitting from the service.

C. Numerical analysis of the game

In this subsection, we suggest to study the game while considering the previous expressions of the transition rates. The dependence of those rates on provider prices are too complicated to solve the problem analytically, therefore we perform here a numerical study. Unless otherwise stated, the parameter values considered in this section are the following: $p_0 = 1, \kappa = 1, \beta = 0.5, \gamma_1 = 1, \gamma_2 = 2, \gamma_0 = 1$. We will refer to this set of parameter values as S .

Figure 4 plots the steady-state revenue R_1 of provider 1 when its price p_1 varies, for different values of the opponent price p_2 . We remark that the revenue of provider 1 is first

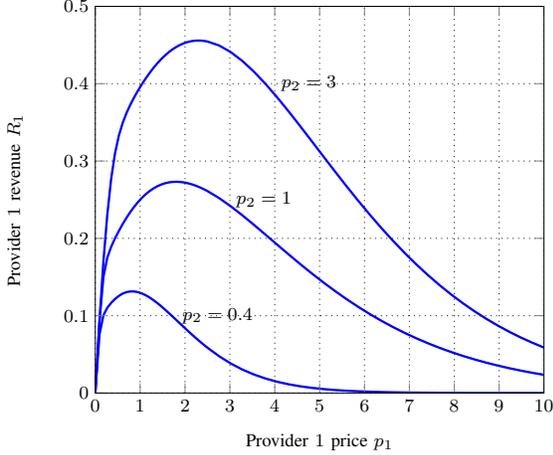


Figure 4: Provider 1 revenue versus p_1 for fixed p_2 .

increasing, then decreasing in p_1 . However, this is not always the case: for some parameter values, the revenue of a provider as a function of its price may have two local maxima, and depending on the opponent's price, the first or the second local optimum is a global one. This is exemplified in Figure 5 for $\gamma_2 = 7$ and the other parameter values in S . Nevertheless, it appears that the revenue of provider 1 tends to 0 as its price tends to infinity, which implies that there exists a finite price p_1 maximizing R_1 . That revenue-maximizing price constitutes the best reply of provider 1 to the price set by provider 2.

As explained in Section III, plotting the best-reply curves of both providers on the same graph highlights the Nash equilibria of the game. Those curves are shown in Figure 6 for the parameter values in S . Figure 6 shows that there are two Nash equilibria of the game, namely $(0, 0)$ and $p^* \approx (2.29, 2.84)$. However, we notice that $(0, 0)$ is not a satisfying situation, since it brings no revenue to the providers, and moreover it is not a stable Nash equilibrium: if any of the two providers slightly deviates from that situation by setting a strictly positive price, then successive best replies lead to the other (stable) Nash equilibrium p^* . We will consequently focus on that equilibrium in the following, when it exists. Indeed, notice that as the example of Figure 5 illustrates, the best reply correspondence $p_2 \mapsto BR_1(p_2)$ may not be continuous

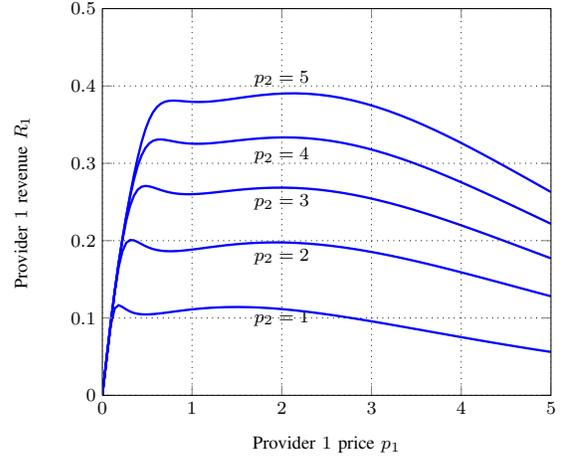


Figure 5: Provider 1 revenue versus p_1 for fixed p_2 and $\gamma_2 = 7$ (other parameters taken from S).

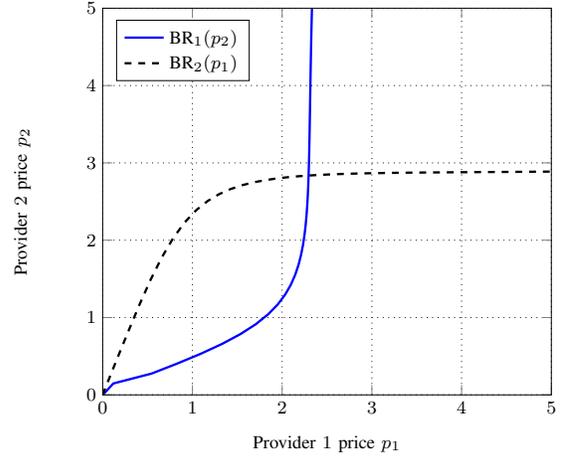


Figure 6: Best-reply curves of both providers.

due to the fact that the global maximum can switch from one local maximum to the other. Figure 7 plots the best-reply correspondences with the same parameter values as for Figure 5. Interestingly, for that set of value there are two stable Nash equilibria: one around $(p_1, p_2) = (0.14, 0.7)$ and the other one near $(p_1, p_2) = (2, 3.9)$. Nevertheless, those cases were rarely met in our numerical computations, and were not met with the “reasonable” values that we used.

D. Influence of the parameter p_0

In Subsection V-B we have interpreted p_0 as the user perceived cost for not benefiting from the service. When the service in question concerns a rapidly time-evolving sector such as telecommunications, it is very likely that this perceived cost p_0 changes. For example, Internet access or cellular telephony are now almost priceless services for many users, because those tools are extensively used and becoming

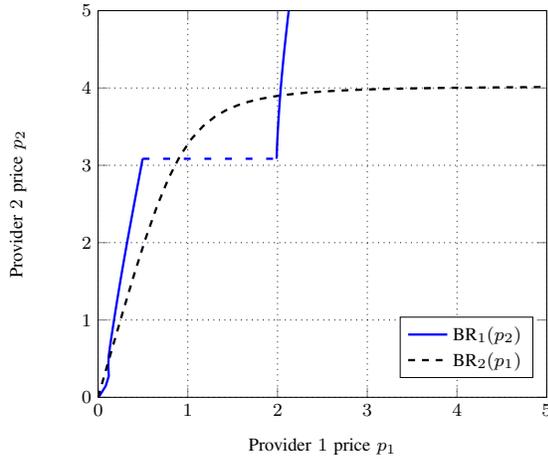


Figure 7: Best-reply curves of both providers when $\gamma_2 = 7$ (other parameter values taken from S).

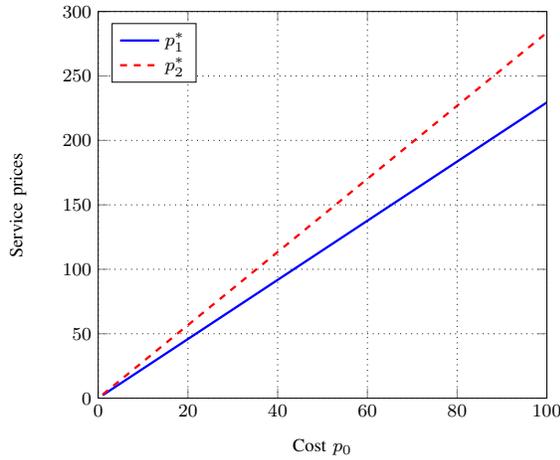


Figure 8: Nash equilibrium prices when the no-service cost p_0 varies.

mandatory for regular business. This was not the case at the very beginning of those technologies. On the other hand, some services/technologies can lose value because they get abandoned or can be replaced by other ones. Services with higher importance/impact can therefore be modeled by a larger p_0

For those reasons, we investigate now the effect of the no-service cost p_0 on the outcomes of the game. We assume that the variations of p_0 are on a longer time scale than the game on prices and user behavior, so that we still compute the Nash equilibrium of the pricing game as described in the previous subsection.

Figure 8 plots the *Nash equilibrium prices* (p_1^*, p_2^*) versus p_0 , while the corresponding user repartition at steady-state is plotted in Figure 9. We remark that the steady-state distribution (again, at Nash equilibrium depending on p_0) is almost

constant when p_0 varies, and equilibrium prices increase close to linearly with p_0 . Therefore, if providers are aware of an increase in p_0 , they should raise their prices correspondingly to benefit from the increased value of the service. Interestingly, this price increase compensates the service value increase from the point of view of the users, since the proportion of users choosing the service or not is unchanged.

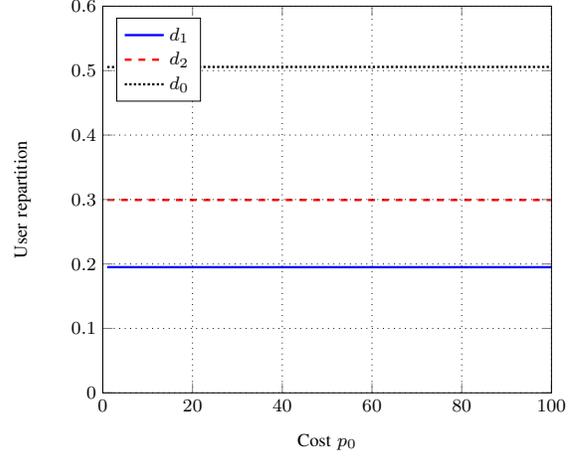


Figure 9: User repartition at Nash equilibrium when the no-service cost p_0 varies.

E. Influence of the price sensitivity β

We study here the effect of the parameter β , that represented users' sensitivities to price differences between the different states of the Markov chain. When β increases, we expect providers to decrease their prices so as to attract more efficiently a maximum of customers. The Nash equilibrium prices (p_1^*, p_2^*) when β varies are shown in Figure 10. We give the corresponding user repartition and provider revenues in Figure 11 and 12, respectively. It appears that when the

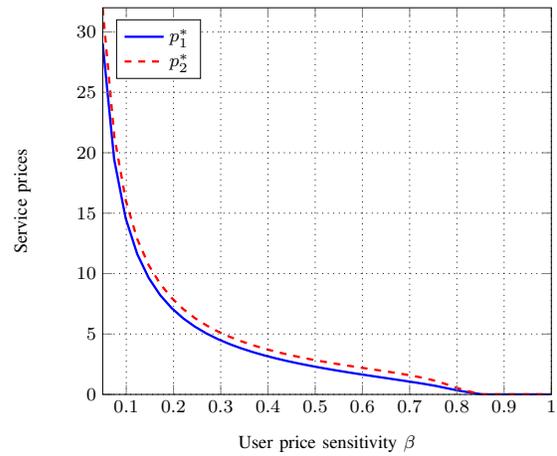


Figure 10: Nash equilibrium prices when β varies.

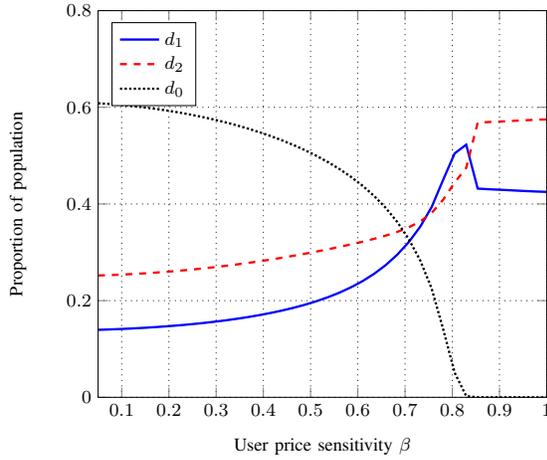


Figure 11: Steady-state user repartition at Nash equilibrium when β varies.

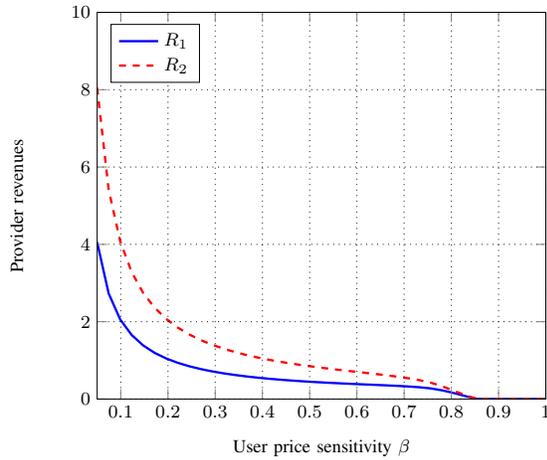


Figure 12: Provider revenues at Nash equilibrium when β varies.

price sensitivity exceeds a given threshold (that is around 0.85), providers engage in a price war that makes their price tend to 0, and they finally get no revenue from providing the service. This benefits users who all end up with one of the two providers. Below that threshold, an increased price sensitivity already implies a provider price reduction, and a revenue decrease for them. Remark also that provider 2, being more attractive than its opponent (since $\gamma_2 > \gamma_1$ in our parameter values), takes benefit of that advantage by setting higher prices.

F. Influence of the asymmetry between providers

In our model, providers only differ for their parameter γ_i , that reflects the user likeliness to stay with him: from (8), a larger γ_i means lower transition rates to the other states of the Markov chain. That asymmetry between providers may for example come from users being more reluctant to leave

the incumbent operator than another one, because they trust less the newly-arrived operators in terms of honesty and/or QoS. Therefore, those parameters γ_i may vary due to word of mouth, advertisement, and are consequently difficult to evaluate.

We therefore investigate here the influence of the asymmetry in providers' γ_i value. To do so, we fix all values but γ_2 from the parameter set S , and make γ_2 vary. We assume that provider 2 is the one with an advantage, i.e. $\gamma_2 \geq \gamma_1$. In Figures 13, 14, and 15, we respectively plot the Nash prices, user repartition and provider revenues versus $1/\gamma_2$ (therefore abscissa also gives the ratio γ_1/γ_2). Notice that the curves are only given for the values of $1/\gamma_2 > 0.18$, because for values below that threshold the game may have several stable Nash equilibria (as in the case shown in Figure 7) and we cannot predict which one will be chosen. Nevertheless, those extreme values may seem unrealistic, since they mean an asymmetry magnitude larger than 5.

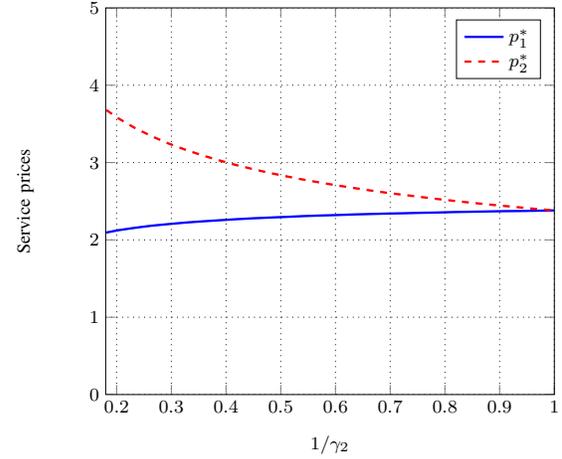


Figure 13: Nash equilibrium prices when $1/\gamma_2$ varies.

We remark that provider 2 takes benefit from his advantage by setting a higher price than his opponent, while still having more customers. That difference in price and user repartition increases with the game asymmetry, and vanishes when the game becomes symmetric, i.e. when γ_2 tends to 1.

VI. CONCLUSIONS AND FUTURE WORK

In this paper, we have presented a competition game model, where providers take into account the user churn behavior to determine the price they fix for the service, so as to maximize their steady-state revenue. Through a numerical analysis, we remarked that the game has a Nash equilibrium, that might not be unique if the asymmetry between providers is very large. When providers are not too different in terms of attractivity or reputation, we investigate the effect of user sensitivity to prices (that is seen to exacerbate the price war), and the effect of an increase in the need for the service (that is observed to benefit only to providers).

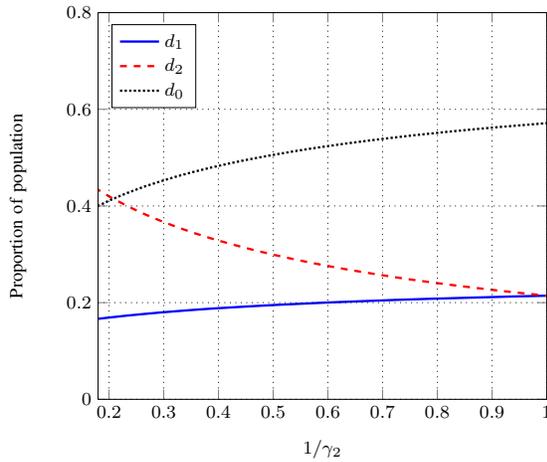


Figure 14: Steady-state user repartition at Nash equilibrium when $1/\gamma_2$ varies.

One interesting direction for future work is to have a different approach as to the considered time scales. In this paper, we have considered several time scales: at the smallest time scale users are assumed to react to prices, while those prices are fixed at a larger time scale, reasoning on the user behavior steady-state outcome. Finally, the value of the service may vary at an even larger time scale. Those assumptions can be justified, but it would also be interesting to relax them, for example by considering the user dynamics within the pricing game: an incumbent provider may start the game with more customers than his opponent, and may therefore be better off beginning with a large price since not all users will immediately churn to the opponent. Likewise, the competitor may have an incentive to start with low prices so as to attract customers, before possibly raising its price.

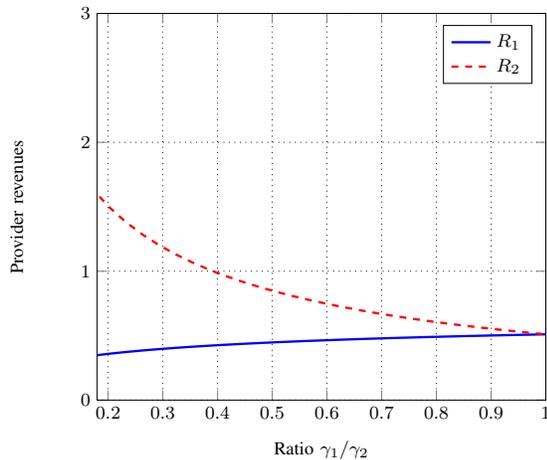


Figure 15: Provider revenues at Nash equilibrium when $1/\gamma_2$ varies.

ACKNOWLEDGMENTS

The authors would like to acknowledge the support of Euro-NF Network of Excellence through the specific research project PRECO, and the French *Agence Nationale pour la Recherche* through the CAPTURES project.

REFERENCES

- [1] K. Wieland, "The customer retention challenge," *Telecommunications*, vol. 40, no. 10, pp. 14–17, October 2006.
- [2] M. Naldi, "A simple model for the effectiveness of delaying strategies for telecommunications churn reduction," in *10th International Conference on Computer Modelling and Simulation EUROSIM 2008*, Cambridge, 1-3 April 2008.
- [3] P. Maillé, M. Naldi, and B. Tuffin, "Competition for migrating customers: A game-theoretic analysis in a regulated regime," in *Proceedings of IEEE GLOBECOM 2008*, Nov. 30 2008-Dec. 4 2008.
- [4] J. Ahn, S. Han, and Y. Lee, "Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry," *Telecommunications Policy*, vol. 30, no. 10-11, pp. 552–568, November-December 2006.
- [5] A. Eshghi, D. Haughton, and H. Topi, "Determinants of customer loyalty in the wireless telecommunications industry," *Telecommunications Policy*, vol. 31, no. 2, pp. 93–106, March 2007.
- [6] S. Buehler, R. Dewenter, and J. Haucap, "Mobile number portability in Europe," *Telecommunications Policy*, vol. 30, no. 7, pp. 385–399, August 2006.
- [7] P. Danaher, "Optimal pricing of new subscription services: Analysis of a market experiment," *Marketing Science*, vol. 21, no. 2, pp. 119–138, 2002.
- [8] 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.
- [9] M. Osborne and A. Rubinstein, *A Course on Game Theory*. MIT Press, 1994.
- [10] M. Mozer, R. Wolniewicz, D. Grimes, E. Johnson, and H. Kaushansky, "Predicting Subscriber Dissatisfaction and Improving Retention in the Wireless Telecommunications Industry," *IEEE Transactions on Neural Networks*, vol. 11, no. 3, pp. 690–696, May 2000.
- [11] W. Au, K. Chan, and X. Yao, "A Novel Evolutionary Data Mining Algorithm With Applications to Churn prediction," *IEEE Transactions on Evolutionary Computation*, vol. 7, no. 6, pp. 532–545, December 2003.
- [12] H.-S. Kim and C.-H. Yoon, "Determinants of subscriber churn and customer loyalty in the Korean mobile telephony market," *Telecommunications Policy*, vol. 28, pp. 751–765, 2004.
- [13] M.-K. Kim, J.-H. Park, J.-H. Paik, S.-H. Seol, and H.-J. Park, "An empirical study on the influencing factors of customer churning intention after introducing mobile number portability in Korea," in *15th ITS Biennial Conference*, Berlin, 4-7 September 2004.
- [14] J. Qi, Y. Zhang, Y. Zhang, and S. Shi, "Treelogit model for customer churn prediction," in *APSCC '06: Proceedings of the 2006 IEEE Asia-Pacific Conference on Services Computing*. Washington, DC, USA: IEEE Computer Society, 2006, pp. 70–75.

Technological Investment Games Among Wireless Telecommunications Service Providers ‡

Patrick Maillé¹, Bruno Tuffin² and Jean-Marc Vigne^{1,2} *†

¹*Institut Telecom; Telecom Bretagne; 2, rue de la Châtaigneraie CS 17607 35576 Cesson Sévigné Cedex, France*

²*Dionysos team, INRIA Rennes - Bretagne Atlantique, Campus Universitaire de Beaulieu 35042 Rennes, France*

SUMMARY

With the development of new technologies in a competitive context, infrastructure investment and licence purchase as well as existing technology maintenance are crucial questions for current and emerging operators. This paper presents a three-level game analyzing this problem. At the highest level, the operators decide on which technologies to invest, given that some may already own licences or infrastructures. We limit ourselves to the realistic case where technologies are 3G, WiFi and WiMAX. At the intermediate level, with that set of operated technologies fixed, operators determine their service price. Finally, at the lowest level, customers choose their provider depending on the best combination of price and available quality of service. At each level, the best decision of actors depends on the actions of others, the interactions hence requiring to be studied as a (noncooperative) game. The model is analyzed by backward induction, meaning that decisions at a level depend on the equilibria reached at the lower levels. Different real-life cost scenarios are studied. Our model aims at helping both the operators to make their final decision on technological investments, and the regulator to determine a proper licence fee range for a better competition among providers. Copyright © 2008 John Wiley & Sons, Ltd.

KEY WORDS: Network pricing, Nash equilibrium, Wardrop equilibrium, technological game, telecommunication investments

1. INTRODUCTION

Telecommunications are becoming omnipresent, with all kinds of services, telephony, television, web browsing, email, etc., now available on the same terminals. Similarly, due to the market liberalization, customers may choose among different providers, their choice being based on different parameters such as price, quality of service (QoS), coverage or used technology for instance. Those providers are therefore competitors, fighting to attract customers in order to maximize their revenue. In this context, providers need to carefully determine not only the access price they will impose, but also on which set of technologies to operate. Basically when talking about wireless providers, the choice is among 3G, WiFi and WiMAX (or LTE).

*Correspondence to: Jean-Marc Vigne, Telecom Bretagne, 2 rue de la Châtaigneraie CS 17607 35576 Cesson-Sévigné Cedex, France

†Email: jm.vigne@telecom-bretagne.eu

‡This paper is an extended version of a paper [1] submitted to IEEE Globecom 2010

We will assume in this paper that customers have at their disposal terminals supporting multiple interfaces, and the technology used is the one providing the best combination of price and QoS meaning that there is no coverage issue (all technologies can be potentially used by all customers). Our goal is to model, understand and propose to operators strategic choices in terms of technology investment, as well as to suggest rules that a regulator could apply to induce a more efficient competition (from a global or a user perspective). Examples of questions we aim at answering are:

- is it worth for an operator paying a licence and an infrastructure for being present in a new technology? Will the return on investment be sufficient? That question is typical of what operators ask themselves with regard to the implementation of WiMAX [2] or LTE.
- Does this investment help to attract more customers, or is it just at the expense of other technologies already implemented? This has to be studied in a competitive environment, given that other operators may make similar strategic moves. Some operators may already be present on some technologies, and therefore their costs are limited; this potential heterogeneity has also to be taken into account.
- Why investing on a technology where a competitor is installed and dominant? The total cost has to be pondered with the revenue from customers. A regulator, in order to break such dominant positions, may decide to compensate that unbalance through the licence fees. An illustration comes from the third generation wireless licences in France, where the regulator wants to open an additional licence to increase competition, this new licence being offered at a lower cost than the initial ones. Our paper helps understanding the range of licence fees allowing a candidate to enter the market and make benefits.

We introduce a model made of three levels of game, corresponding to three different time scales. At the lowest level, given fixed operated technologies and service costs, customers spread themselves among available operators in order to get the “best” combination of price and QoS, where the QoS (or the congestion) they get depends on the choice of the other users. To simplify the analysis, users are assumed infinitesimal. As a consequence, the (selfish) decision of a single individual does not have any influence on the system behavior. The equilibrium analysis is therefore provided by the so-called *Wardrop’s principle* [3] coming from transportation theory: at equilibrium, all providers with a positive demand have the same perceived combination of price and QoS, otherwise customers would switch to the “best”. When choosing their price at the intermediate level, providers can anticipate what the resulting equilibrium (and therefore their revenue) would be for a given price profile, i.e., for given and fixed prices for all providers, which induces a pricing game among providers at this second level. The general framework is here again that of non-cooperative game theory, and the equilibrium notion is now that of *Nash equilibrium* with atomic players [4]. Finally, at the largest time scale, providers can decide which technologies to operate. In order to make that decision, they have to compute what their revenue would be at the equilibrium or equilibria (if any) of the lower-level pricing game, and compare it with their costs. That choice, which also depends on the strategy of competitors, will be made in order to reach again a Nash equilibrium for this “technology game”. While there exist papers on the first two levels of game (see the literature review subsection below), we are not aware of any other paper analyzing the technological game, especially when using the results of the two other levels. This paper seems to be the first in that important direction of modeling and understanding the complete chain of provider strategies. We additionally illustrate the interest of our methodology by modeling practical situations of competition and technology investment arising in France.

1.1. Literature review

Several papers can be found on the two lower levels of game, i.e., the game among users to find the best provider and the pricing game among providers [5, 6, 7, 8, 9, 10]. In [5], Acemoglu and Ozdaglar consider infinitesimal users choosing a provider so as to minimize their perceived price, that is the sum of a congestion cost and a financial charge. When an equilibrium on prices exists (and the authors show it does when congestion costs are linear), then competition is proved to lead to an overall social welfare not too far from the optimal one. Our model can be seen as a variation of theirs, by considering some technologies where the resource is shared among operators, and imposing operators to set a unique price for all the technologies they implement. The main extension though comes from the third level of game where providers can choose the operated technologies, and from the associated real-life scenarios. Another model in [6] describes several retailers periodically selling product units and competing on the initial fill rate (i.e., the fraction of available items to be sold), the periodical retail price and the stock-policy they choose. The authors show that an equilibrium on those parameters can be deduced from another equilibrium of a single-period game on prices and fill rates only. For some demand expressions, it is also shown that such equilibria exist. In our context, product units can be interpreted as data units or packets, and fill rates are supposed constant. Our model makes use of Wardrop principle to define the customer equilibrium, while attraction models are considered in [6]; above all, no technological game is considered there due to the different focus.

A model for the two lower levels of game has been again proposed in [7, 8], similarly to ours; but congestion is modeled there by losses instead of delay, and the higher-level technological game is still not considered. In [10], a two-sided competition model is proposed. Users choose operators offering the highest utility in term of price and QoS too, but two populations, with different sensitivity to the two parameters are considered. Providers offer different QoS because they operate on different frequency bands, an aspect which can be covered by our model. A new aspect is that some bandwidth can be sold to ad-hoc networks, serving as *secondary* users. The equilibrium on prices is searched numerically by means of a learning algorithm, but there again, no technological game is considered.

There is to our knowledge no other paper dealing with the technology game, especially using the result of the pricing game. But other multilevel games exist, the most notable ones being [11] and [12]. Though, [11] rather models the interactions among Internet service providers and content providers, while we rather aim at investigating which technologies a provider should implement, given the potential revenue and the potential infrastructure and licence costs. On the other hand, [12] considers investments to improve the quality of service, but not to implement a new technology. Moreover, that paper mixes the investment and the pricing decisions into a single game level, while we separate here those choices due to different time scales. We have not seen elsewhere that kind of study in a similar competitive context. We also provide typical and real-life competition situations of providers already installed but which could try to extend their technological range to increase their revenue.

1.2. Organization

This paper is organized as follows. In Section 2, we present the basic model that will be used: the user behavior, the set of providers and available technologies, and the three levels of game. The lowest level of game, that is the competition among users looking for the network with the best combination of price and QoS, is analyzed in Section 3; the equilibrium is characterized, and existence and uniqueness are discussed. Section 4 analyzes the pricing game for fixed technologies, anticipating what the reaction of users would be. The third level of game, the game on technologies, is described in Section 5. This game makes use of the revenues at the pricing-game level, and pastes the infrastructure and licence

costs. We show how this can be solved. Section 6 on the other hand considers practical situations of competition, and illustrates which equilibria can be found. Real-life scenarios are considered, typical of competition encountered in France, and we show how the model can help to propose relevant technological investment strategies for providers, but also to propose ranges of licence costs for the regulators towards a better use or sharing of the resource. We finally conclude and give directions for future research activities in Section 7.

2. THE MODEL

We describe in this section the definitions and assumptions on the model, as well as the three levels of game that will be analyzed later on. We assume that we have a set $\mathcal{N} = \{1, \dots, N\}$ of telecommunication providers trying to maximize their revenue. Each provider $i \in \mathcal{N}$ has to decide which set of technologies \mathcal{S}_i it will operate, and the access price p_i per unit of flow that a customer needs to pay if using her network. The set of technologies \mathcal{S}_i is to be chosen within a set \mathcal{T} of available technologies. In our practical scenario, considering wireless operators, this set will be

$$\mathcal{T} = \{3G, \text{WiMAX}, \text{WiFi}\}.$$

This set \mathcal{T} of technologies is partitioned into two subsets \mathcal{T}_p and \mathcal{T}_s . For a technology in \mathcal{T}_p , each operating provider owns a licence and a part of the radio spectrum, using it *alone*. In other words, congestion on this technology depends only on the level of demand the operator experiences on her own network. On the other hand, for a technology in \mathcal{T}_s , the spectrum is shared by the customers of the competing providers, so that congestion depends not only on the level of demand at the provider, but also on demand coming from competitors using this technology. Basically, we will consider

$$\mathcal{T}_p = \{3G, \text{WiMAX}\} \text{ and } \mathcal{T}_s = \{\text{WiFi}\}.$$

We assume that the price p_i charged by provider $i \in \mathcal{N}$ is independent of the technology used by customers. In other words, provider i fixes a price for network access that is the same for all the technologies she operates. We moreover assume that users have terminals with multiple interfaces, allowing them to use any technology, and that they can sense the available QoS. The choice of the technology is left to the user terminal, that will (selfishly) choose a couple (provider, technology) offering the best combination of price and QoS.

Users are modeled by the aggregate level of flow demand $d_{i,t}$ experienced at technology $t \in \mathcal{T}$ by provider $i \in \mathcal{N}$. Users are seen as infinitesimal (also said non-atomic), so that the action of a single user is considered having no impact, contrary to that of an (infinite) set of users. This kind of assumption is usual in the related transportation theory where a car has no influence on congestion, but where a flow of cars rather has to be dealt with. We call \mathbf{d} the vector of all flow demands.

QoS is modeled by a congestion cost function $\ell_{i,t}$ of $d_{i,t}$ for owned-spectrum technology t (i.e., $t \in \mathcal{T}_p$) operated by provider i , and ℓ_t of total demand $\sum_j d_{j,t}$ for shared-spectrum technology t (i.e., $t \in \mathcal{T}_s$). Those functions are assumed strictly increasing -more demand implies more congestion-, continuous and non-negative.

User behavior is modeled in the following way. We assume that each user tries to minimize her *perceived price* $\bar{p}_{i,t}$, that is for technology t and provider i defined by

$$\bar{p}_{i,t}(\mathbf{d}) = \begin{cases} p_i + \ell_{i,t}(d_{i,t}) & \text{if } t \in \mathcal{T}_p \\ p_i + \ell_t\left(\sum_j d_{j,t}\right) & \text{if } t \in \mathcal{T}_s. \end{cases} \quad (1)$$

This means that the perceived price is a linear combination of a monetary cost (the price charged) and a QoS cost (the congestion level). Any other combination form could be considered, but we follow here the representation proposed in [5, 13]. Our model is actually an extension of the one in [13], where *i*) different technologies can be operated by a single provider, and *ii*) some can also be shared. This addition is for introducing the technological game which is not addressed in [5, 13]. We will see in the next section how demand is distributed at equilibrium, which will result in a *system* perceived price \bar{p} , common to all technologies that do get some demand.

We also assume that total user demand $d = \sum_{i=1}^N \sum_{t \in \mathcal{S}_i} d_{i,t}$ is a continuous function $D(\cdot)$ of the perceived price \bar{p} , strictly decreasing on its support and with $\lim_{p \rightarrow \infty} D(p) = 0$. $D(p)$ represents the total amount of traffic that users would want to transmit at a fixed perceived unit price p , it may encompass (cumulated) individual demand variations with respect to price, and/or (cumulated) individual decisions to abandon the service. In addition, we call v the inverse function of D on its support.

The system is characterized by three different time scales, resulting in three different levels of game:

- at the shortest time scale, for fixed prices and sets of offered technologies, users choose their provider and technology in order to minimize their perceived price. This drives to a user equilibrium situation $(d_{i,t}^*)_{i,t}$ where for all technologies with positive demand, the perceived price is the same, other technologies having larger perceived prices (otherwise some users would have an interest to change to a cheaper option).
- At the intermediate time scale, providers compete for customers by playing with prices for fixed sets of implemented technologies. The goal of each provider i is to maximize her revenue

$$R_i = \sum_{t \in \mathcal{S}_i} p_i d_{i,t}^*, \quad (2)$$

playing on price p_i and making use of what the user equilibrium $\mathbf{d}^* = (d_{i,t}^*)_{i,t}$ would be for a given price profile. Since the revenue of a provider depends on the price strategy of competitors (through the user equilibrium), this is analyzed using non-cooperative game theory.

- At the larger time scale, providers have to choose which technologies to invest on. This is again analyzed thanks to non-cooperative game theory, using the equilibrium situation $(p_i^*)_{1 \leq i \leq N}$ of the intermediate level. The goal is here again to optimize

$$B_i = \sum_{t \in \mathcal{S}_i} (p_i^* d_{i,t}^* - c_{i,t})$$

(where $c_{i,t}$ represents the licence and infrastructure costs to provider i to operate on technology t), by playing with the set \mathcal{S}_i .

3. FIRST LEVEL OF GAME: COMPETITION AMONG USERS

To study our three-level game, we first need to determine how user demand is distributed among providers. Access prices are assumed fixed in this section, those prices only being changed at a larger time scale. Following our non-atomicity assumption about users, the equilibrium is driven by Wardrop's principle [3] coming from transportation theory: at a user equilibrium the perceived price at each provider getting some demand is the same, otherwise, if one is larger than another, then her customers would prefer to change and go to the cheapest. Some providers may also be too expensive, with an access price larger than the perceived price of competitors, and therefore get no demand.

Wardrop's principle can be formalized as follows. Consider a technology configuration $\mathcal{S} = (\mathcal{S}_1, \dots, \mathcal{S}_N)$ and a price configuration (p_1, \dots, p_N) . A Wardrop (or user) equilibrium is a family $(d_{i,t})_{i \in \mathcal{N}, t \in \mathcal{S}_i}$ of positive real numbers such that

$$\left\{ \begin{array}{l} \forall i \in \mathcal{N}, \forall t \in \mathcal{S}_i \quad \bar{p}_{i,t} = \begin{cases} p_i + \ell_{i,t}(d_{i,t}^*) & \text{if } t \in \mathcal{T}_p \\ p_i + \ell_t(\sum_{j:t \in \mathcal{S}_j} d_{j,t}^*) & \text{if } t \in \mathcal{T}_s \end{cases} \\ \forall i \in \mathcal{N}, \forall t \in \mathcal{S}_i \quad d_{i,t}^* > 0 \implies \bar{p}_{i,t} = \min_{j \in \mathcal{N}, \tau \in \mathcal{S}_j} (\bar{p}_{j,\tau}) \\ \sum_{i \in \mathcal{N}} \sum_{t \in \mathcal{S}_i} d_{i,t}^* = D(\min_{i \in \mathcal{N}, t \in \mathcal{S}_i} (\bar{p}_{i,t})). \end{array} \right. \quad (3)$$

The first equality presents the perceived price at each couple (operator, technology), separating the case where the technology is owned by the operator (in \mathcal{T}_p) and the case where it is shared (in \mathcal{T}_s). The second equality states that users choose the cheapest option in terms of perceived price, and that the perceived price at providers getting demand is necessarily the same - otherwise, again, some customers would have an interest in churning - . Finally, the last equality states that total demand, i.e., sum of demands at each network, is the demand function at the perceived price.

Remark that this kind of nonatomic game played among users falls into the widely-studied set of routing games [14, 15, 16]. Indeed, the user problem can be interpreted as finding a route (i.e., a pair provider-technology) to reach the global internet, while congestion effects occur. Several powerful results exist for that kind of games, that we apply to prove existence of a user equilibrium for our particular problem, and uniqueness of perceived prices.

Proposition 3.1. *There always exists a user equilibrium. Moreover, the corresponding perceived price at each provider-technology pair (i, t) is unique.*

Proof: For strictly positive prices $(p_i)_{i \in \mathcal{N}}$, the existence of a Wardrop equilibrium directly comes from Theorem 5.4 in [15], where existence is ensured when perceived price functions are strictly positive, which is the case when providers set strictly positive prices. Just a few extra verifications are needed for the specific case where some providers set their price to 0: by choosing $\varepsilon > 0$ and replacing $\ell_{i,t}(x)$ by $\bar{\ell}_{i,t}(x) := \max(\varepsilon, \ell_{i,t}(x))$ for all provider i with $p_i = 0$, we know that a solution of the system (3) with modified perceived price functions exists. But when ε tends to 0, the corresponding perceived price $\bar{p} = \min_{i \in \mathcal{N}, t \in \mathcal{S}_i} (\bar{p}_{i,t})$ does not tend to 0 since all congestion cost functions are strictly increasing, and demand is continuous and strictly positive at price 0. As a result, ε can be chosen sufficiently small such that for a Wardrop equilibrium with modified cost functions, modified and original cost functions coincide, which means the original system (3) has a solution.

We now focus on uniqueness. For a Wardrop equilibrium, we denote by \bar{p} the common perceived price of all options (i.e., pairs provider-technology) that get positive demand. Assume there exist two Wardrop equilibria \mathbf{d} and $\hat{\mathbf{d}}$ with different perceived prices, say \bar{p} and \hat{p} , and assume without loss of generality that $\bar{p} > \hat{p}$. Since the demand function is strictly decreasing on its support, then total demand for \mathbf{d} is strictly smaller than for $\hat{\mathbf{d}}$. This implies that either total demand on one of the shared technologies, or demand on one proprietary technology of a provider, is strictly smaller for \mathbf{d} than for $\hat{\mathbf{d}}$. But following Wardrop's principle, this would mean that the corresponding cost for that technology is the minimal cost \hat{p} for $\hat{\mathbf{d}}$, that is strictly larger (due to congestion cost increasingness) than for \mathbf{d} , itself being larger than \bar{p} , a contradiction. As a result, the perceived price \bar{p} for options with demand is

unique, and we necessarily have for each provider-technology pair (i, t) :

$$\begin{aligned} t \in \mathcal{T}_p &\Rightarrow \bar{p}_{i,t} = \max(\bar{p}, p_i + \ell_{i,t}(0)) \\ t \in \mathcal{T}_s &\Rightarrow \bar{p}_{i,t} = p_i + \max(\bar{p} - \underline{p}_t, \ell_t(0)), \end{aligned}$$

where $\underline{p}_t := \min\{p_i, t \in \mathcal{S}_i\}$. All perceived prices are unique, which concludes the proof. ■

In general, the Wardrop equilibrium is not unique, as illustrated by the following example.

Example 3.1. Consider a situation where two providers operate the same shared technology, with a linear congestion function $\ell(d) = d$ and a demand function $D(\bar{p}) = [2 - \bar{p}]^+$. If $p_1 = p_2 = 1$, then any demand profile $(d_1, d_2) = (x, \frac{1}{2} - x)$ with $x \in [0, \frac{1}{2}]$ is a solution to system (3), i.e., a Wardrop equilibrium.

Though, one can assume that when a shared technology is charged exactly the same price by several providers, the total demand splits among those providers according to some predefined proportions. This can for example be interpreted as inner preferences of users, involving reputation effects for example: if the perceived prices were to be the same, users have a propensity to follow those preferences, and therefore to distribute accordingly among providers. This is formalized in the following assumption.

Assumption 3.1. Define a weight $w_i > 0$ associated to each provider $i \in \mathcal{N}$, characterizing her reputation among the user population. If for a technology $t \in \mathcal{T}_s$ there is a set \mathcal{N}_t of providers with the same minimal price, the total demand d_t on t is shared such that

$$\forall i \in \mathcal{N}_t \quad d_{i,t} = d_t \frac{w_i}{\sum_{j \in \mathcal{N}_t} w_j}.$$

A strict priority among providers, which would correspond to a limit case of the previous formulation, could also be considered. We can then establish a uniqueness property for the Wardrop equilibrium demand distribution.

Proposition 3.2. Under Assumption 3.1, for any price profile, the Wardrop equilibrium is unique.

Proof: From Proposition 3.1, all perceived prices are unique at a Wardrop equilibrium, as well as the common price perceived by all users $\bar{p} = \min_{i \in \mathcal{N}, t \in \mathcal{S}_i} \bar{p}_{i,t}$. Then for $i \in \mathcal{N}$, the conditions in (3) imply that

- if $t \in \mathcal{S}_i$ is such that $\bar{p}_{i,t} > \bar{p}$, then $d_{i,t} = 0$,
- if $\bar{p}_{i,t} = \bar{p}$ for a $t \in \mathcal{S}_i \cap \mathcal{T}_p$, then $d_{i,t} = \ell_{i,t}^{-1}(\bar{p} - p_i)$,
- if $\bar{p}_{i,t} = \bar{p}$ for a $t \in \mathcal{S}_i \cap \mathcal{T}_s$, then from Assumption 3.1, $d_{i,t} = \ell_t^{-1}(\bar{p} - p_i) \frac{w_i}{\sum_{j \in \mathcal{N}_t} w_j}$.

In all possible cases, demand $d_{i,t}$ is uniquely determined. ■

4. INTERMEDIATE LEVEL: THE PRICING GAME AMONG OPERATORS

For the rest of this paper, we assume that Assumption 3.1 holds. Now, for a fixed profile of implemented technologies $(\mathcal{S}_1, \dots, \mathcal{S}_N)$, the operators need to choose at the intermediate level the price they will charge to users. They act strategically, in the sense that for any price profile $\mathbf{p} = (p_1, \dots, p_N)$ they predict the corresponding (unique) Wardrop equilibrium discussed in the previous subsection, noted here $(d_{i,t}(\mathbf{p}))_{1 \leq i \leq N, t \in \mathcal{S}_i}$.

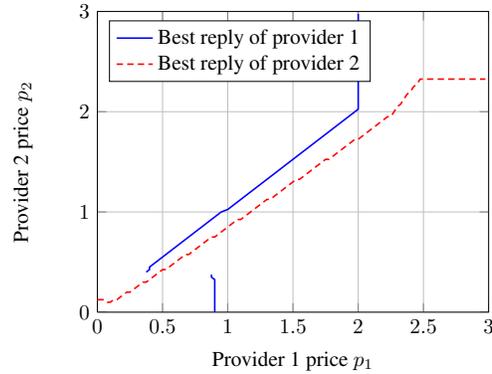


Figure 1. Curves of best-reply prices for Example 4.1.

Each provider $i \in \mathcal{N}$ tries to maximize her revenue

$$R_i(\mathbf{p}) := p_i \sum_{t \in \mathcal{S}_i} d_{i,t}(\mathbf{p}),$$

defined as the product of the price charged and the total demand at provider i .

The equilibrium notion in this case is the so-called *Nash equilibrium*, which is a price profile \mathbf{p}^* such that no provider, acting selfishly, can increase her revenue by an unilateral deviation. Formally, a Nash equilibrium is a price profile \mathbf{p}^* such that $\forall i \in \mathcal{N}$,

$$\forall p_i \geq 0 \quad R_i(p_i^*; \mathbf{p}_{-i}^*) \geq R_i(p_i; \mathbf{p}_{-i}^*)$$

where $(p_i; \mathbf{p}_{-i}^*)$ is vector \mathbf{p}^* with price p_i^* of operator i replaced by p_i .

A Nash equilibrium does not necessarily exist, as illustrated by the following example.

Example 4.1. Consider a scenario with two providers proposing both an owned-spectrum technology, and such that demand and congestion cost functions are defined by

$$\begin{aligned} D(x) &= [4 - x]^+ \\ \ell_1(d_1) &= \frac{1}{(5 - d_1)^5} - \frac{1}{5^5} \\ \ell_2(d_2) &= \frac{1}{(3 - d_2)^5} - \frac{1}{3^5} \end{aligned}$$

Then there is no Nash equilibrium for the pricing game. To illustrate this, Figure 1 displays the best response curves of providers (i.e., prices maximizing their revenue) in terms of the price of the competitor. A Nash equilibrium should be an intersection point of those two curves. Here, there is no intersection point, due to a discontinuity in Player 2's best response at around $p_2 = 0.39$. This comes from the fact that the revenue curve of Player 2 has two local maxima for a given p_1 , and the global maximum switches from one to the other. This transition is illustrated in Figure 2.

Though, in all practical scenarios that will be investigated in Section 6, when both operators propose owned spectrum technologies, a unique Nash equilibrium exists and is non null.

In the rest of this section, we characterize some properties of a Nash equilibrium, when it exists.

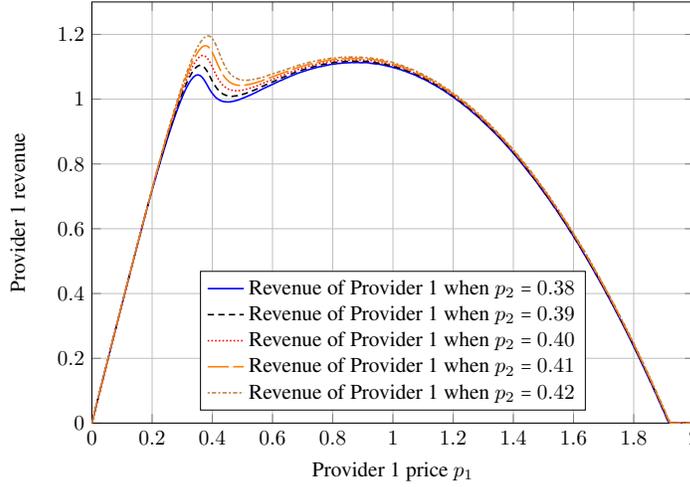


Figure 2. Provider 1 revenue as a function of prices of both providers.

Proposition 4.1. *Under Assumption 3.1, at a Nash equilibrium with strictly positive prices, each technology in \mathcal{T}_s (i.e., shared-spectrum) is used by at most a single operator.*

Proof: To establish the proposition, we apply a result from [17], where a model similar to ours is used, but no shared technologies are involved. In that paper, Hayrapetyan, Tardos, and Wexler establish that the Wardrop equilibrium repartition is continuous in the price profile. In our case, each shared technology $t \in \mathcal{T}_p$ can be seen as a single option (i.e., regardless of the provider chosen) with a charge price $\underline{p}_t := \min\{p_j, t \in \mathcal{S}_j\}$, and a congestion cost $\ell_t(\sum_j d_{j,t})$. Since $\min\{p_j, t \in \mathcal{S}_j\}$ is continuous in the price profile, then so is the total flow on each shared technologies, as well as flows on each owned-spectrum technology $t \in \mathcal{T}_s$.

Suppose that there exists a Nash equilibrium with a shared-spectrum technology $t \in \mathcal{T}_s$ for which at least two providers have a positive demand.

Remark first that in that case, the providers have declared the same price p . Indeed, from (3) the perceived price at those providers are necessarily the same due to the fact that they have a positive demand, and the congestion cost is the same for both providers on that shared technology.

Let us consider a provider i that does not obtain all the demand on technology t at the Wardrop equilibrium, i.e., $d_{i,t}^* < \sum_{j:t \in \mathcal{S}_j} d_{j,t}^* := d_t^*$. Then consider that operator decreasing her price by a small amount $\varepsilon > 0$. Consequently, by a small decrease of one's price, provider i would only slightly affect demand on her owned-spectrum technologies, but will be the only cheapest provider on technology t , and thus get all demand (itself being slightly modified) on that technology. Likewise, if provider i operates on other shared technologies, her demand does not decrease on those. Therefore, when ε tends to 0, the change in revenue for provider i tends to a value that is at least $p_i (d_t^* - d_{i,t}^*)$, which is strictly positive. Provider i can then choose ε small enough to strictly improve her revenue, which contradicts the price profile being a Nash equilibrium. ■

A Nash equilibrium of the pricing game does not always exist. However, when there is one, the prices set by providers satisfy a relation similar to the ones proved in [12, 13].

Proposition 4.2. *Assume Assumption 3.1 holds and that, on its support, the demand function D is*

concave and differentiable with bounded differential function D' . Further assume that all congestion functions ℓ_t and $\ell_{i,t}$ are such that $\ell_t(0) = \ell_{i,t} = 0$, i.e., no congestion means no cost. Then at any Nash equilibrium (p_1^*, \dots, p_N^*) of the pricing game, with corresponding demands $(d_{i,t}^*)_{1 \leq i \leq N, t \in \mathcal{S}_i}$ and perceived price \bar{p} , we have

$$p_i^* = \frac{d_i^*}{\sum_{k \in \mathcal{S}_i} \frac{1}{\ell'_{i,k}(d_{i,k}^*)}} + \frac{d_i^*}{\left(\sum_{j \neq i} \sum_{t \in \mathcal{S}_j} \frac{1}{\ell'_{j,t}(d_{j,t}^*)} \right) - D'(\bar{p})}, \quad (4)$$

where $d_i^* := \sum_{t \in \mathcal{S}_i} d_{i,t}^*$ is the total demand of provider i and $d^* := \sum_{i \in \mathcal{N}} d_i^*$ is the overall demand. Remark that the right-hand side of (4) can be completely expressed in terms of demands, since \bar{p} is the solution of $D(\bar{p}) = d$.

The proof is provided in the Appendix.

5. THIRD LEVEL OF GAME: COMPETITION ON TECHNOLOGIES

At an even larger time scale, providers have to decide which technologies to operate. Here again, their decisions will depend on the anticipation about what their revenue would be at the equilibrium of the intermediate level, for any profile $\mathcal{S} = (\mathcal{S}_1, \dots, \mathcal{S}_N)$ of strategies implemented by the operators. The goal is therefore for each provider i to determine the combination of technologies $\mathcal{S}_i \subset \mathcal{T}$ which will maximize her revenue, taking into account the implementation (infrastructure plus licence) costs.

Formally, each provider can choose her (finite) subset of technologies, resulting in a (multidimensional) matrix of revenues $(R_1(\mathcal{S}), \dots, R_N(\mathcal{S}))_{\mathcal{S} \in \mathcal{T}^N}$. Similarly, let $c_{i,t}$ represent the licence and infrastructure costs to provider i to operate on technology t . For simplicity of the presentation, those costs are assumed additive, such that the cost of implementing \mathcal{S}_i for i is $\sum_{t \in \mathcal{S}_i} c_{i,t}$, but one can without any added complexity consider a general cost function $c_i(\mathcal{S}_i)$. Those costs $c_{i,t}$ can be highly asymmetric because some providers may already own an infrastructure, or a part of it, and/or a licence, when it is required. The goal of each provider i is at this time scale also to maximize her net benefit

$$B_i(\mathcal{S}) = R_i(\mathcal{S}) - \sum_{t \in \mathcal{S}_i} c_{i,t} = \sum_{t \in \mathcal{S}_i} (p_i^* d_{i,t}^* - c_{i,t}),$$

given that competitors proceed similarly. The equilibrium notion is here again that of a Nash equilibrium, which is a profile \mathcal{S}^* of implemented technologies, such that no provider can improve her benefit by changing unilaterally her set of technologies:

$$\forall i \in \{1, \dots, N\}, \forall \mathcal{S}_i \subset \mathcal{T}, \quad B_i(\mathcal{S}^*) \leq B_i(\mathcal{S}_i; \mathcal{S}_{-i}^*)$$

where $(\mathcal{S}_i; \mathcal{S}_{-i}^*)$ is vector \mathcal{S}^* with \mathcal{S}_i^* replaced by \mathcal{S}_i . Note that the set of strategies is finite here instead of continuous in the previous section.

In the next section, we consider specific situations and analyze the existence and uniqueness of an equilibrium, which cannot be guaranteed in general. We will consider the situation of $N = 2$ providers in competition. We will therefore end up with two matrices:

- a matrix of revenues from subscribers

$$R := (R_1(\mathcal{S}), R_2(\mathcal{S}))_{\mathcal{S}_1, \mathcal{S}_2 \subset \mathcal{T}}$$

giving for each combination of technology choices, the respective revenues of the two providers at the equilibrium of the pricing game, obtained from Section 4;

- and a cost matrix

$$C = (c_1(\mathcal{S}), c_2(\mathcal{S}))_{\mathcal{S}_1, \mathcal{S}_2 \in \mathcal{T}}.$$

From those two matrices, the net benefit matrix

$$R - C = (R_1(\mathcal{S}) - c_1(\mathcal{S}), R_2(\mathcal{S}) - c_2(\mathcal{S}))_{\mathcal{S}_1, \mathcal{S}_2 \in \mathcal{T}}$$

is deduced. A Nash equilibrium, if any, is then an element of that last matrix such that the first coordinate $R_1 - c_1$ is maximal over the lines, and the second coordinate $R_2 - c_2$ is maximal over the columns.

If in general, we may have several and non-symmetric Nash equilibria, one of them may be more relevant, be it from users point of view, or regarding the overall wealth generated by the resources (here, the radio spectrum). We therefore introduce corresponding measures, that can be computed for each technology profile.

Define the valuation function V as the total money that customers are ready to spend in order to buy q flow units, i.e.,

$$V(q) = \int_0^q v(q) dq, \quad (5)$$

where v is the generalized inverse of the demand function D , i.e., $v(q) = \min\{p : D(p) \leq q\}$. Define also the user welfare as the difference between the money customers are ready to spend, and the total price they actually pay to buy q flow units,

$$UW = V(q) - v(q) \times q.$$

As a last definition, we call social welfare, noted SW , the sum of utilities of all actors in the game:

- customers, with aggregate utility represented by the user welfare,
- providers, with utility $R_j - c_j$ for provider j ,
- and licence sellers and infrastructure sellers, with respective total revenues R_{ls} and R_{is}

Hence,

$$\begin{aligned} SW &= UW + R_{ls} + R_{is} + \sum_{j=1}^N (R_j - c_j) \\ &= UW + \sum_{j=1}^N R_j, \end{aligned}$$

the last equality coming from $R_{ls} + R_{is} = \sum_{j=1}^N c_j$ because revenues of licence and infrastructure sellers are exactly the sum of costs of providers. From the customers point of view (resp. from the point of view of the society as a whole: users, providers, licence seller, infrastructure sellers) the most interesting Nash equilibrium is the one maximizing user welfare (resp. social welfare).

6. PRACTICAL EXAMPLES

We now apply our model to some particular contexts of telecommunication operator competition. In the case of the competition among french wireless providers in 2010, we present a technology choice analysis with a WiMAX deployment option for two operators.

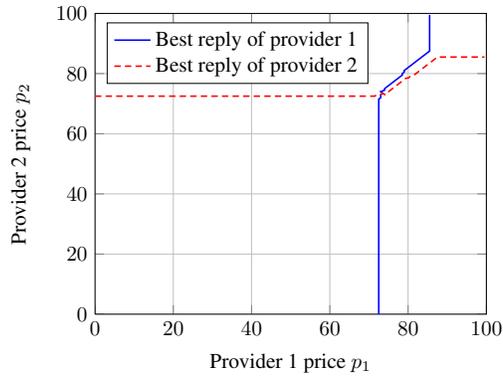


Figure 3. Curves of best-reply prices when both players propose the WiMAX technology.

6.1. Intermediate level pricing game

We first focus on the game on prices at the intermediate level, ignoring for the moment the licence and infrastructure costs. Formally, we build the revenue matrix R defined in the previous section.

We assume that a set of users is positioned in a bounded predefined zone, and that all users have a terminal with multiple interfaces. We have in mind a specific zone covered by a 3G UMTS base station in France. There are approximately 10^4 such zones on the French territory. We additionally assume that the zone is covered by a WiFi 802.11g access point and a single 802.16e (WiMAX) base station. We only consider downlink for convenience and choose realistic values of demand and capacities:

- 28 Mb/s per operator for 3G [18];
- 40 Mb/s, still in downlink, for WiMAX technology [19];
- 25 Mb/s for WiFi [20].

We moreover assume the demand function D to be linear on its support, given by (in Mb/s)

$$D(\bar{p}) = [300 - 3\bar{p}]^+,$$

with \bar{p} in €/month. Hence no user is willing to pay more than 100 € monthly to benefit from the service. In our numerical computations, the congestion function $\ell_{i,t}$ of a couple demand-technology (i, t) is supposed to be the average waiting time of a M/M/1 queue of parameters $(d_{i,t}, C_{i,t})$ if the technology t belongs to the set \mathcal{T}_p , and of parameters $(\sum_{t \in \mathcal{S}_i} d_{i,t}, C_t)$ if t belongs to \mathcal{T}_s . Recall that the average waiting time of an M/M/1 queue with parameters (λ, μ) is $1/(\mu - \lambda) - 1/\mu$.

To illustrate how those Nash equilibria are found, consider for instance the case where the two operators decide to propose only a WiMAX access to the users. Figure 3 displays the best responses (i.e., prices maximizing their revenue) of providers in terms of the price of the other. A Nash equilibrium is an intersection point of those two curves. One can check that there exists a single non-null intersection point between the two curves, here approximately equal to $(72.5, 72.5)$. For the practical examples studied in this paper, we proceeded numerically to compute best-replies of both providers, and find the Nash equilibria on prices (actually, we always found either one unique Nash equilibrium, or no equilibrium at all).

Table I displays the monthly revenues in euros of both operators at the Nash price profile, for every technology profile. On each element of the table, the first number is the revenue of Provider 1, while the

$I \setminus J$	\emptyset	3G	WiMAX	3G,WiMAX	WiFi	WiFi,3G	WiFi,WiMAX	WiFi,3G,WiMAX
\emptyset	0;0	0;2470	0;3379	0;5081	0;2228	0;4220	0;4921	0; 6193
3G	2470;0	2200;2220	2090;3022	1843;4506	2241;1985	1970;3752	1859;4367	1623; 5415
WiMAX	3379;0	3022;2090	2864;2864	2489;4240	3061;1887	2689;3545	2532;4112	2192; 5043
3G,WiMAX	5081;0	4506;1843	4240;2489	3638;3638	4572 ;1666	3957 ;3079	3691 ;3538	3092 ;4232
WiFi	2228;0	1985;2241	1887;3061	1666; 4572	0;0	-	-	-
WiFi,3G	4220;0	3752;1970	3545;2689	3079; 3957	-	-	-	-
WiFi,WiMAX	4921;0	4367;1859	4112;2532	3538; 3691	-	-	-	-
WiFi,3G,WiMAX	6193 ;0	5415 ;1623	5043 ;2192	4232 ;3092	-	-	-	-

Table I. Revenues matrix (from users) for providers depending on the implemented technologies.

$I \setminus J$	\emptyset	3G	WiMAX	3G,WiMAX	WiFi	WiFi,3G	WiFi,WiMAX	WiFi,3G,WiMAX
\emptyset	0.0	74	182	662	54	384	600	1396
3G	74	486	726	1442	408	938	1262	2282
WiMAX	182	726	975	1785	662	1368	1755	2709
3G,WiMAX	662	1442	1785	2799	1368	2305	2799	4087
WiFi	54	408	662	1368	103	-	-	-
WiFi,3G	384	938	1368	2305	-	-	-	-
WiFi,WiMAX	600	1262	1755	2799	-	-	-	-
WiFi,3G,WiMAX	1396	2282	2709	4087	-	-	-	-

Table II. User welfare depending on the implemented technologies.

second is the revenue of Provider 2. We can notice a direct consequence of Proposition 4.1: when both operators choose to implement the WiFi technology, then if a Nash exists (which is when they operate WiFi only), then they both end up with a null revenue. The price equilibrium is when both providers set their price to 0: from that situation, no provider can unilaterally change her price and get a strictly positive revenue, since setting a strictly positive price implies that all users on the shared technology -here thus, all the demand- go to the competitor. In all other cases where the operators implement WiFi but at least one provider operates another technology, no Nash equilibrium actually exists, because at the only possibility $(0, 0)$, any provider operating an unshared technology could increase her price and make profit on that technology. We thus end up with the “-” symbol in Table I to represent that no Nash equilibrium exists.

Best responses of providers, in terms of technology sets, are displayed in bold in the Table. A Nash equilibrium is therefore easily spotted as cells with both numbers on bold. If we considered the game on technologies without any implementation cost, we would then have two possible (and symmetric) Nash equilibria, $(\{\text{WiFi}, 3\text{G}, \text{WiMAX}\}, \{3\text{G}, \text{WiMAX}\})$ and $(\{3\text{G}, \text{WiMAX}\}, \{\text{WiFi}, 3\text{G}, \text{WiMAX}\})$.

We now aim at investigating for different scenarios the outcome of the game on technologies, as well as the selection of those equilibria from a user and social welfare optimization point of view.

6.2. Symmetric game

Consider again, as well as in the rest of the paper, a zone covered by a single base station, for a period of one month. Estimated infrastructure plus licence costs, if any, are therefore also divided by the 10^4 zones in France and by the duration in months of the licence rights. As presented in Section 5, we define a cost per zone and per month at provider i for technology t and a cost matrix $(c_1(\mathcal{S}_1), c_2(\mathcal{S}_2))_{\mathcal{S}_1, \mathcal{S}_2 \subset \mathcal{T}}$.

$I \setminus J$	\emptyset	3G	WiMAX	3G,WiMAX	WiFi	WiFi,3G	WiFi,WiMAX	WiFi,3G,WiMAX
\emptyset	0	2544	3561	5742	2282	4604	5521	7588
3G	2544	4886	5838	7791	4634	6660	7488	9320
WiMAX	3561	5838	6703	8514	5610	7602	8399	9944
3G,WiMAX	5743	7791	8514	10075	7606	9341	10028	11411
WiFi	2282	4634	5610	7606	103	-	-	-
WiFi,3G	4604	6660	7602	9341	-	-	-	-
WiFi,WiMAX	5521	7488	8399	10028	-	-	-	-
WiFi,3G,WiMAX	7588	9320	9944	11411	-	-	-	-

Table III. Social welfare depending on the implemented technologies.

$I \setminus J$	\emptyset	3G	WiMAX	3G,WiMAX	WiFi	WiFi,3G	WiFi,WiMAX	WiFi,3G,WiMAX
\emptyset	0;0	0;762	0;2555	0;2549	0;2178	0;2462	0; 4047	0;3611
3G	762;0	492;512	382;2198	135;1974	533;1935	262;1994	151; 3493	-85;2833
WiMAX	2555;0	2198;382	2040;2040	1665;1708	2237 ;1837	1865 ;1787	1708 ; 3238	1368 ;2461
3G,WiMAX	2549;0	1974;135	1708;1665	1106;1106	2040;1616	1425;1321	1159; 2664	560;1650
WiFi	2178;0	1935;533	1837; 2237	1616;2040	-50;-50	-50;-1758	-50;-874	-50;-2582
WiFi,3G	2462;0	1994;262	1787; 1865	1321;1425	-1758;-50	-1758;-1758	-1758;-874	-1758;-2582
WiFi,WiMAX	4047 ;0	3493 ;151	3238 ; 1708	2664 ;1159	-874;-50	-874;-1758	-874;-874	-874;-2582
WiFi,3G,WiMAX	3611;0	2833;-85	2461; 1368	1650;560	-2582;-50	-2582;-1758	-2582;-874	-2582;-2582

Table IV. Benefits matrix for the symmetric game.

In this symmetric game, we consider two incoming providers without any wireless infrastructure paying the same infrastructure and licence costs. A total 3G licence of 649 M€ [21] needs to be paid to the regulation authority, and 3G infrastructure of 1.4B€ (value inspired from [22]) has to be purchased, both monthly paid over 10 years. Hence, the licence cost (resp. the infrastructure cost) is then evaluated to 541 € (resp. 1167 €) per month and per zone, giving $c_{1,3G} = c_{2,3G} = 1708$ €. We also assume that a licence costs 649 M€ for WiMAX and the infrastructure costs 340 M€ (inspired from [23]), yielding $c_{1,WiMAX} = c_{2,WiMAX} = 541 + 283 = 824$ €. We assume that every WiFi access point is renewed each year and is bought at the average price of 600 € per year. In France, since only a declaration to the regulation authority, negligible taxes and no licence purchase are necessary to deploy and use a WiFi infrastructure [24, 25], we then choose $c_{1,WiFi} = c_{2,WiFi} = 50$ €.

The resulting benefits matrix (revenue matrix minus cost matrix) is displayed in Table IV. It can be readily checked that there exists two Nash equilibria, $(\{WiFi,WiMAX\},\{WiMAX\})$ and $(\{WiMAX\},\{WiFi,WiMAX\})$.

In that case, we remark in Table IV that there are two Nash equilibria, $(\{WiFi,WiMAX\},\{WiMAX\})$ and $(\{WiMAX\},\{WiFi,WiMAX\})$. With respect to the previous situation, 3G is actually too expensive with the proposed licence cost to be implemented. Since those Nash equilibria on technologies are symmetric, user and social welfare values are the same and no preference can be defined.

6.3. A WiFi-positioned provider against a 3G one

Consider a WiFi-installed provider, named Provider 1, wishing to extend her position against a 3G-installed provider, noted Provider 2. We suppose that Provider 1 already owns a complete WiFi infrastructure over the 10^4 zones (basically like the provider called Free in France) and that Provider 2

1 \ 2	\emptyset	3G	WiMAX	3G,WiMAX	WiFi	WiFi,3G	WiFi,WiMAX	WiFi,3G,WiMAX
\emptyset	0;0	0;1929	0;2555	0;3716	0;2178	0;3629	0;4047	0; 4778
3G	1437;0	1167;1679	1057;2198	810;3141	1208;1935	937;3161	826;3493	590; 4000
WiMAX	2555;0	2198;1549	2040;2040	1665;2875	2237;1837	1865;2954	1708;3238	1368;3628
3G,WiMAX	3224;0	2649;1302	2383;1665	1781;2273	2715;1616	2100;2488	1834;2664	1235; 2817
WiFi	2228;0	1985;1700	1887;2237	1666; 3207	0;-50	-	-	-
WiFi,3G	3187;0	2719;1429	2512;1865	2046; 2592	-	-	-	-
WiFi,WiMAX	4097;0	3543;1318	3288;1708	2714;2326	-	-	-	-
WiFi,3G,WiMAX	4336;0	3558;1082	3186;1368	2375; 1727	-	-	-	-

Table V. Benefits matrix for the WiFi-3G game.

similarly owns a complete 3G infrastructure over the same zones (Bouygues Telecom for instance). Bouygues Telecom already owning an infrastructure, only its licence cost of 649 M€ accounts, giving $c_{2,3G} = 541$ €. The cost of the fourth licence in France (the one Free is buying) is fixed to 240 M€ [21], and of the new infrastructure estimated at 1.0 B€ (value inspired from [22]), so that $c_{1,3G} = 1033$ € per month and per site. For WiFi, we choose $c_{1,WiFi} = 0$ € and $c_{2,WiFi} = 50$ € to illustrate the better position of Provider 1, while WiMAX costs are the same as in the previous subsection (that technology being a new one). The benefits matrix is given in Table V, with again best responses highlighted in bold. For this game, there are two non symmetric Nash equilibria. The first one is ($\{WiFi, WiMAX\}, \{3G, WiMAX\}$): each operator chooses the technology on which she is already present, and additionally goes to the new WiMAX technology. The second Nash equilibrium is ($\{WiMAX\}, \{WiFi, 3G, WiMAX\}$) and corresponds to a situation where Provider 2 proposes all technologies and Provider 1 only proposes the WiMAX technology. Again, it is better not to fight on (the low-cost) WiFi.

Since the ($\{WiFi, WiMAX\}, \{3G, WiMAX\}$) results in a computed social welfare $SW = 10028$, and the ($\{WiMAX\}, \{WiFi, 3G, WiMAX\}$) equilibrium yields $SW = 9944$, the first one would be better-suited in terms of social welfare. Similarly, user welfare values are respectively 2799 and 2709, the first equilibrium is more advised from the users point of view too.

It is possible to vary the Nash equilibria set by changing licences prices. Indeed, obtaining a Nash equilibrium with 3G implemented by Provider 1 requires the cost $c_{1,3G}$ to be reduced to 900 € (the situation ($\{WiMAX\}, \{WiMAX, WiFi, 3G\}$)) is not an equilibrium anymore if this cost is reduced even more). This would mean a licence fee of 67 € (or equivalently a global 3G licence selling price of 80 M€). In that case, the situation ($\{3G, WiMAX\}, \{WiFi, 3G, WiMAX\}$) would be a third Nash equilibrium of the technological game: Provider 1 would focus on licenced technologies, giving up on WiFi. Social welfare and user welfare values are in this case respectively equal to 11411 and 4087. Those values are the maximal ones that can be attained, as could be expected: it is indeed in the interest of the community to use the maximum of resources (i.e., all the radio spectrum), and this also benefits to users since more available resources correspond to a harder competition for customers and less congestion.

6.4. A single technology-positioned provider against a dominant one

This kind of game would for instance correspond in France to Free (strongly established in the Internet and WiFi networks), named Provider 1 again, against Orange, named Provider 2, already positioned on almost all technologies, except WiMAX, for which we keep the costs of previous subsections. The 3G costs are also considered the same as in the previous subsection, but WiFi costs are here

1 \ 2	\emptyset	3G	WiMAX	3G,WiMAX	WiFi	WiFi,3G	WiFi,WiMAX	WiFi,3G,WiMAX
\emptyset	0;0	0;1929	0;2555	0;3716	0;2228	0;3679	0;4097	0; 4828
3G	1437;0	1167;1679	1057;2198	810;3141	1208;1985	937;3211	826;3543	590; 4050
WiMAX	2555;0	2198;1549	2040;2040	1665;2875	2237;1887	1865;3004	1708;3288	1368;3678
3G,WiMAX	3224;0	2649;1302	2383;1665	1781;2273	2715;1666	2100;2538	1834;2714	1235; 2867
WiFi	2228;0	1985;1700	1887;2237	1666; 3207	0;0	-	-	-
WiFi,3G	3187;0	2719;1429	2512;1865	2046; 2592	-	-	-	-
WiFi,WiMAX	4097;0	3543;1318	3288;1708	2714;2326	-	-	-	-
WiFi,3G,WiMAX	4336;0	3558;1082	3186;1368	2375; 1727	-	-	-	-

Table VI. Benefits matrix for the WiFi-Dominant game.

$c_{1,WiFi} = c_{2,WiFi} = 0 \in$ due to the past presence of both providers on this technology. The results of the technological game are displayed in Table VI. One can see here that, again, two Nash equilibria exist and are the same as those of the previous game in part 6.3, with user and social welfare equal to 2799 and 10028 for the first equilibrium, and 2709 and 9944 for the second one. That is, the existence of the WiFi infrastructure for Provider 2 does not affect the Nash equilibria. Hence, we deduce that the impact of the WiFi infrastructure cost is negligible compared to the 3G and WiMAX licence and infrastructure costs. Similarly, if the monthly cost per site for 3G gets as low as 694 €, then Provider 1 could keep operating WiFi, since the situation ($\{WiFi,3G,WiMAX\},\{3G,WiMAX\}$) would arise as a technological Nash equilibrium. In that case, the monthly licence cost would be equal to 153 € (the total licence price would be equal to 184M€). The social welfare and user welfare values are in this case respectively 11411 and 4087. Hence reducing the 3G costs would be beneficial, since this last Nash equilibrium yields larger user and social welfare values.

6.5. A 3G technology-positioned provider against an omnipresent one

This last scenario corresponds in France to Bouygues Telecom (Provider 1, operator only owning a 3G infrastructure), against SFR (Provider 2, owning a 3G and WiFi infrastructure). The WiMAX infrastructure and licence costs are again assumed to be the same as before for both operators. In addition, the WiFi infrastructure cost is assumed to be equal to 50 €per month. The 3G licence cost is also equal to 541 €, given that both licences are supposed equal to 649 M€and paid over 10 years. We can notice on Table VII that there exist two symmetric Nash equilibria on technologies which are ($\{WiFi, 3G, WiMAX\},\{3G,WiMAX\}$) and ($\{3G, WiMAX\},\{WiFi, 3G,WiMAX\}$). Both providers have an interest to invest in the WiMAX technology and to keep their 3G infrastructure active. This conclusion contrasts with the one opposing a 3G operator to a dominant one depicted in Table VI. Since the found Nash equilibria maximize the user and social welfare values among every technology combinations of operators, no new interesting Nash equilibrium on technologies would come from a 3G licence price variation.

7. CONCLUSION

In this paper, we have presented a three-level competition model on technology investments among wireless telecommunications service providers. Each level corresponds to a different time scale. At a first level, users choose the couple operator-technology offering them the best compromise between congestion and price per flow unit, where total demand is supposed elastic. Some demand can be

1 \ 2	\emptyset	3G	WiMAX	3G,WiMAX	WiFi	WiFi,3G	WiFi,WiMAX	WiFi,3G,WiMAX
\emptyset	0;0	0;1929	0;2555	0;3716	0;2178	0;3629	0;4047	0; 4778
3G	1929;0	1659;1679	1549;2198	1302;3141	1700;1935	1429;3161	1318 ;3493	1082; 4000
WiMAX	2555;0	2198;1549	2040;2040	1665;2875	2237;1837	1865;2954	1708;3238	1368; 3628
3G,WiMAX	3716;0	3141;1302	2875;1665	2273;2273	3207 ;1616	2592 ;2488	2326 ;2664	1727 ; 2817
WiFi	2228;0	1985;1700	1887;2237	1666; 3207	0;-50	-	-	-
WiFi,3G	3679;0	3211;1429	3004;1865	2538; 2592	-	-	-	-
WiFi,WiMAX	4097;0	3543;1318	3288;1708	2714; 2326	-	-	-	-
WiFi,3G,WiMAX	4828 ;0	4050 ;1082	3678 ;1368	2867 ; 1727	-	-	-	-

Table VII. Benefits matrix for the 3G-Dominant game.

possibly be shared among providers under a predefined rule based on their reputation. At a second level, operators choose their price per flow unit maximizing their revenue at the obtained flow distribution, such that no one would have an incentive to change it. At a third level, operators choose the technology combination maximizing their revenue, which is based on the price and flow distribution of the previous levels and the infrastructure and licence costs. We illustrate our model with a simple competition study among french wireless operators, where considered technologies are 3G, WiMAX and WiFi. Hence, given some initial infrastructure or licence price reduction, it has been shown in the four scenarios opposing two operators that it is in their interest to invest in the WiMAX infrastructure. A licence cost reduction can be necessary in some cases, because this reduction can generate a new equilibrium on technologies maximizing the social welfare.

The work presented in the paper can be extended in several ways. A first possibility is to adapt the model to an unshared technology zone covered by several shared technology subzones under similar flow equilibrium constraints. A second way to model this is to modify the model such that some smooth increase of a minimal perceived price does not jeopardize the whole corresponding demand and to analyze the existence of price equilibrium in the case where technologies are shared. Finally, customers may not have multiple interfaces, i.e., they may not be able to choose among all technologies. This heterogeneity could be taken into account.

Acknowledgment

This paper is an extended version of a paper submitted to IEEE Globecom 2010 (reference [1]). The authors would like to acknowledge the support of the ANR Verso Project CAPTURES <http://captures.inria.fr/>

APPENDIX: Proof of Proposition 4.2

Proof: First, if a Nash equilibrium exists, then from Proposition 4.1, there are no shared-spectrum technologies for which more than one provider obtain some demand. But at a Nash equilibrium, any shared-spectrum technology $t \in \mathcal{T}_s$ gets some demand: if this were false, this would mean that all operators of t have a price above \bar{p} and thus get no revenue, while they could obtain a strictly positive revenue by choosing a price strictly below \bar{p} .

Consequently, at a Nash equilibrium each shared-spectrum technology is effectively operated (i.e.,

with strictly positive demand) by exactly one provider, that is the cheapest one among the operators of that technology. As a result, infinitesimal price variations from providers do not change the identity of the cheapest ones on shared-spectrum technologies. A consequence is that in a vicinity of the Nash price profile (p_1^*, \dots, p_N^*) , each shared-spectrum technology behaves exactly as if it were an owned-spectrum technology of its cheapest operator.

We can therefore apply Proposition 2 of [13], where all links (in our context, technologies) are owned by providers, but only the property of Nash equilibria being local maxima of revenues is used. That property can also be used in our model, considering all technologies as (locally) owned-spectrum ones.

The only difference from [13] is that we allow here providers to operate on several technologies. However, since they declare a unique price and demand distributes itself according to a Wardrop equilibrium, then all technologies for a provider have the exact same perceived cost (this should hold only for technologies *with demand*, but here all technologies get demand since $\ell_{i,t}(0) = 0$ for all $t \in \mathcal{S}_i$), and therefore the same congestion cost $\ell_{i,t}(d_{i,t})$. If a provider gets a total demand $d_i = \sum_{t \in \mathcal{S}_i} d_{i,t}$, we refer to that common value of the congestion cost by $\ell_i(d_i)$, which should satisfy:

$$\forall d_i \geq 0, \quad \ell_i(d_i) = \ell_{i,t}(x_t) \quad \forall t \in \mathcal{S}_i \quad (6)$$

$$s.t. \quad x_t \geq 0, \quad \sum_{t \in \mathcal{S}_i} x_t = d_i. \quad (7)$$

Now consider an infinitesimal variation of ε from an initial $d_i > 0$. We denote by ε_t the corresponding variation of x_t , for each $t \in \mathcal{S}_i$. From (6), we have for all $t \in \mathcal{S}_i$, $\ell_i(d_i + \varepsilon) = \ell_{i,t}(x_t + \varepsilon_t)$. Since the functions $\ell_{i,t}$ are strictly increasing and concave, $\varepsilon_t = O(\varepsilon)$, and moreover $x_t > 0$ and $\ell'_{i,t}(x_t) > 0$. Thus,

$$\begin{aligned} \ell_i(d_i + \varepsilon) &= \ell_{i,t}(x_t + \varepsilon_t) = \ell_{i,t}(x_t) + \varepsilon_t \ell'_{i,t}(x_t) + o(\varepsilon_t) \\ &= \ell_i(d_i) + \varepsilon_t \ell'_{i,t}(x_t) + o(\varepsilon), \end{aligned}$$

and therefore $\varepsilon_t = \frac{\ell_i(d_i + \varepsilon) - \ell_i(d_i)}{\ell'_{i,t}(x_t)} + o(\varepsilon)$. Then (7) yields

$$\sum_{t \in \mathcal{S}_i} \varepsilon_t = (\ell_i(d_i + \varepsilon) - \ell_i(d_i)) \sum_{t \in \mathcal{S}_i} \frac{1}{\ell'_{i,t}(x_t)} + o(\varepsilon) = \varepsilon,$$

which implies that ℓ_i is differentiable, with derivative

$$\ell'_i(d_i) = \frac{1}{\sum_{t \in \mathcal{S}_i} \frac{1}{\ell'_{i,t}(x_t)}}. \quad (8)$$

As a result, from the provider point of view, the set of technologies \mathcal{S}_i behaves exactly as a unique technology that would have a congestion cost function ℓ_i . Moreover, since $\ell_{i,t}$ is convex for all t , so is ℓ_i . Proposition 4.2 is then directly obtained by plugging (8) into the Proposition 2 of [13]. ■

REFERENCES

1. P. Maillé, B. Tuffin, and J.-M. Vigne. Economics of technological games among telecommunication service providers. Submitted, available at http://www.rennes.enst-bretagne.fr/~pmaillage/papiers/tech_submit.pdf, 2010.
2. IEEE 802.16-2004. *IEEE Standard for Local and Metropolitan Area Networks - Part 16: Air Interface for Fixed Broadband Wireless Access Systems*, October 2004.

3. J.G. Wardrop. Some theoretical aspects of road traffic research. *Proc. of the Institute of Civil Engineers*, 1:325–378, 1952.
4. M.J. Osborne and A. Rubenstein. *A Course on Game Theory*. MIT Press, 1994.
5. D. Acemoglu and A. Ozdaglar. Competition and efficiency in congested markets. *Mathematics of Operations Research*, 2006.
6. F. Bernstein and A. Federgruen. A general equilibrium model for industries with price and service competition. *Management Science*, 52(6):868–886, 2004.
7. P. Maillé and B. Tuffin. Analysis of price competition in a slotted resource allocation game. In *Proc. of IEEE INFOCOM 2008*, Phoenix, AZ, USA, April 2008.
8. P. Maillé and B. Tuffin. Price war with partial spectrum sharing for competitive wireless service providers. In *Proceedings of IEEE Globecom*, December 2009.
9. M.H. Manshaei, J. Freudiger, M. Felegyhazi, P. Marbach, and J.P. Hubaux. On wireless social community networks. In *Proc. of IEEE INFOCOM*, Phoenix, AZ, USA, Apr 2008.
10. Y. Xing, R. Chandramouli, and C. Cordeiro. Price dynamics in competitive agile spectrum access markets. *Selected Areas in Communications, IEEE Journal on*, 25(3):613–621, April 2007.
11. P. Njoroge, A. Ozdaglar, N. Stier-Moses, and G. Weintraub. Competition, market coverage, and quality choice in interconnected platforms. In *Proceedings of the Workshop on the Economics of Networked Systems (NetEcon)*, 2009.
12. R. Johari, G. Y. Weintraub, and B. Van Roy. Investment and market structure in industries with congestion. *Operations Research*, submitted.
13. A. Ozdaglar. Price competition with elastic traffic. *Networks*, 52(3):141–155, 2008.
14. M. Beckmann, C. B. McGuire, and C. B. Winsten. *Studies in the economics of transportation*. Yale University Press, New Haven, Connecticut, 1956.
15. H.Z. Aashtiani and T.L. Magnanti. Equilibria on a congested transportation network. *SIAM Journal of Algebraic and Discrete Methods*, 2:213–226, 1981.
16. T. Roughgarden. *Selfish Routing and the Price of Anarchy*. MIT Press, 2005.
17. A. Hayrapetyan, É. Tardos, and T. Wexler. A network pricing game for selfish traffic. In *Proc. of ACM SIGACT-SIGOPS Symposium on Principles of Distributed Computing (PODC)*, 2005.
18. 01NetPro. Le débit sur le réseau 3G d'Orange passe à 14 mb/s, January 2010. <http://pro.01net.com/editorial/511553/le-debit-sur-le-reseau-3g-dorange-passe-a-14-mbits-s/>.
19. Wireless design mag. WiMAX 101 : the Evolution of WiMAX, 2010. <http://www.wirelessdesignmag.com>.
20. Wikipédia. Wi-Fi. <http://fr.wikipedia.org/wiki/Wi-Fi>.
21. Mobinaute. 4ème licence 3G : Orange fait pression auprès du gouvernement pour ralentir son attribution, October 2009. <http://www.mobinaute.com/304016-4eme-licence-3g-orange-pression-aupres-gouvernement-ralentir-attribution.html>.
22. Capital. Mobiles : comment Free compte faire baisser les prix, May 2009. <http://www.capital.fr/enquetes/strategie/mobiles-comment-free-compte-faire-baisser-les-prix-380294>.
23. Wimaxday. Worldmax launches mobile WiMAX network, June 2008. <http://www.wimaxday.net/site/2008/06/17/worldmax-launches-mobile-wimax-network>.
24. 01NetPro. Le Wi-Fi devient officiellement une activité commerciale, May 2007. <http://pro.01net.com/editorial/348616/le-wi-fi-devient-officiellement-une-activite-commerciale/>.
25. Arcep. Décision no. 2007-0408 de l'autorité de régulation des communications électroniques et des postes en date du 26 avril 2007 mettant fin au régime d'expérimentation de réseaux ouverts au public utilisant la technologie rlan, April 2007.

Appendix C

Main publications evoked in Chapter 4

Influence of search neutrality on the economics of advertisement-financed content

PIERRE COUCHENEY, Université de Versailles / PRiSM
 GIUSEPPE D'ACQUISTO, Università di Roma Tor Vergata
 PATRICK MAILLÉ, Institut Mines-Telecom / Telecom Bretagne
 MAURIZIO NALDI, Università di Roma Tor Vergata
 BRUNO TUFFIN, Inria

The search neutrality debate questions the ranking methods by search engines. We analyze the issue when content providers offer content for free, but get revenues from advertising. We investigate the non-cooperative game among competing content providers under different ranking policies. When the search engine is not involved with high-quality content providers, it should adopt neutral ranking, maximizing also user quality-of-experience. If the search engine controls high-quality content, favoring its ranking and adding advertisement yield a larger revenue. Though user perceived quality may not be impaired, the advertising revenues of the other content providers decrease drastically.

Categories and Subject Descriptors: K.6.0 [Management of computing and information systems]: Economics

General Terms: Economics, Management, Performance

Additional Key Words and Phrases: Computational advertising, Content providers, Game theory, Search engines, Search neutrality

ACM Reference Format:

Pierre Coucheney, Giuseppe D'Acquisto, Patrick Maillé, Maurizio Naldi, and Bruno Tuffin. *ACM Trans. Internet Technol.* V, N, Article A (January YYYY), 20 pages.
 DOI: <http://dx.doi.org/10.1145/0000000.0000000>

1. INTRODUCTION

Search engines are the dominant way to access content on the web, exerting a strong influence on what people see and read on the Internet. Early on, the possibility that, either by accident or by deliberate choice, such influence turns into bias has been noted and whistleblowed against [Introna and Nissenbaum 2000]. Actually, the structure of the market of search engines raised the issue of a strong concentration as early as in 2001, when the Hirschman-Herfindahl concentration index (HHI) was 0.116 [Sheu and Carley 2001]. The low barriers for new entrants observed in [Gandal 2001] have turned into the present extremely strong dominance of Google [Telang et al. 2004], with an HHI well above 0.6.

Author's addresses: Pierre Coucheney, Université de Versailles / PRiSM, France, E-mail: pierre.coucheney@uvsq.fr; Giuseppe D'Acquisto, Università di Roma Tor Vergata, Italy, E-mail: dacquisto@ing.uniroma2.it; Patrick Maillé, Institut Mines-Telecom / Telecom Bretagne, France, E-mail: patrick.maille@telecom-bretagne.eu; Maurizio Naldi, Università di Roma Tor Vergata, Italy, E-mail: naldi@disp.uniroma2.it; Bruno Tuffin, Inria, France, E-mail: bruno.tuffin@inria.fr.

Acknowledgments: This work was supported in part by the Italian Ministry of Education, University, and Research (MIUR) under PRIN 2012C4E3KT national research project *AMANDA* (Algorithmics for MAssive and Networked DATA).

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or permissions@acm.org.

© YYYY ACM 1533-5399/YYYY/01-ARTA \$15.00

DOI: <http://dx.doi.org/10.1145/0000000.0000000>

The present near-monopoly structure of that market further strengthens concerns about the capability to influence operators located elsewhere in the value chain.

Search engines' business model is based on advertising [Levene 2011], with both organic and sponsored links. Slots are assigned to sponsored links through auctions [Maillé et al. 2012] and to organic ones through a proprietary ranking mechanism, based on the relevance of the linked webpage for the user's query. However, when ranking organic links, the search engine may use metrics related to its own interest, inducing an unfair result for some links, ranking them below where they should be and making them rarely reached by users.

Ranking criteria are crucial for the business model of all the stakeholders: search engines, content providers, and advertisers. Content providers wish to improve their ranking among organic links, e.g. through search engine optimization techniques without affecting the quality of the displayed content [Berman and Katona 2013]. Quality is a main driver for users' choices, in the competition between organic and sponsored links [White 2013], but among sponsored links as well [Chen and He 2011; Athey and Ellison 2011], though being influenced by the ranking strategy of the search engine. Search engines may use that influence to their own advantage, distorting both users' choices and the market structure. Acting as intermediaries between buyers and sellers, they can divert their users from their preferred websites to websites for which the search engine has a vested interest [Hagiu and Jullien 2011]: a *non-neutral* behaviour, which raises the issue of *search neutrality*.

A parallel can be made with the issue of *net neutrality* [D'Acquisto et al. 2012], where network providers may unduly discriminate among the service/content providers that use their network. Search neutrality is considered as the next frontier even if net neutrality should prevail [Odlyzko 2009]. Critics of the search neutrality approach have cast doubts about the capability of measures to lead to neutral search results and really protect users against the abuse perpetrated by websites [Grimmelmann 2010], and have even promoted search bias as the product of the competitive process, and the presence of vertical integration (the search engine favouring its own content) as a generally efficient and pro-competitive practice [Manne and Wright 2011]. But the search engine's ranking strategy may adversely affect the market position of content providers, and the debate on search neutrality should consider the impact of a non neutral behavior on the content providers' revenues.

In this paper, we define a model to analyze the influence of search neutrality on the distribution of free content (funded by advertising revenues). A *neutral* search engine should rank the results of a search based on the relevance of contents for users, i.e., their *quality of experience* (QoE), which is negatively affected by advertising. On the other hand, a non-neutral SE would display results based on the potential revenue from CPs (such a model encompasses both cases where the search engine favors its own content and where it receives payment from CPs to favor their content). Our objective is to investigate the consequences (as to user QoE, content provider and search engine revenue) of the presence of a regulation imposing search neutrality. We focus on the competition among content providers, though we consider the presence of competition among search engines through the possibility for users to reduce their queries on a search engine (hence, leaving for competing search engines). For that purpose, we define a noncooperative game among the content providers, which use their level of advertising as a strategic leverage to maximize their revenues. Their behavior is strongly influenced by the ranking rules applied by the search engine. The application of rules favoring some CPs at the expense of others may strongly distort the market, especially with a monopolistic search engine. The onset of this situation should call for a regulatory intervention. Here we provide grounds for such a decision, examining the case where the SE decides its own ranking criterion. We consider both the case of a monopolistic search engine (the rate of requests to the SE being assumed independent of the QoE) and that where that rate depends on the quality of the content providers it addresses users to. The ranking policies we examine are based on either quality alone (the neutral approach) or the revenues the content provider transfers to the search engine, possibly weighted through

the quality of experience perceived by the user – representing two different non-neutral approaches. We also consider the case of vertical integration between the search engine and one of the content providers, which may then be favoured by its parent company. This case has become particularly interesting: the acquisition of YouTube by Google is a major example of vertical integration, with a search engine owning a video content provider.

To the best of our knowledge, this paper is one of the very first to try to model mathematically the impact of a non-neutral search engine behavior on the Internet actors, and to analyze it thanks to game theory. The other noticeable reference is [Coucheney et al. 2012], but it is not focusing on the impact of advertisement as we are doing here.

The paper is organized as follows. Section 2 introduces the mathematical models for the QoE resulting from content with advertising, the influence of ranking on the visit rate of content providers, and the strategies of content providers. In Section 3, we analyze the case where content providers neglect their individual influence on the success of the search engine (the number of requests per time unit), through a non-cooperative game. We prove the existence of a Nash equilibrium, whatever the ranking policy. Under a neutral policy, we find that content providers with the same intrinsic quality (that in the absence of advertising) are led to reduce their advertising level as the number of competitors grows. Content providers with higher intrinsic quality are instead led to advertise more, getting larger revenues. If the search engine adopts a revenue-based ranking policy, we prove that content providers are led to set their advertising level to the maximum possible value, zeroing the quality perceived by users, regardless of their relationship to the search engine. If the revenue-based policy is mitigated by considering the QoE as well, vertically integrated content providers are favored, yielding more advertising and larger revenues. Section 4 treats the case of content providers including their effect on the request arrival rate in their strategic decision, and compares the performance of neutral and non-neutral rankings. In that case, we prove that the Nash equilibrium can be found as the solution of a system of polynomial equations. Under the neutral policy, content providers with the same intrinsic quality set their advertising level so as to halve their QoE. Under a non-neutral policy, content providers are instead led to increase their advertising as the number of competitors grows, though their revenues decline until being lower than in the neutral case. The conclusions are given in Section 5.

2. MODELS FOR THE BEHAVIOR OF THE STAKEHOLDERS

We consider a single search engine that has to rank different webpages hosting content when a user performs a search. We focus on a single search engine, but we do account for the possibility that the user selects an alternative search engine, abandoning the strictly monopolistic market structure for SEs. Contents are controlled by different CPs, which can play with the amount of advertisement on their webpage. In this paper, we consider a piece of information that is searched for by a user and proposed by a set \mathcal{I} of CPs. In this section, we provide models for the behavior of all the stakeholders: users, the search engine, and content providers. A general discussion on modeling issues and interactions between stakeholders in the field of network economics can be found in [Maillé and Tuffin 2014].

2.1. User's quality of experience with content providers

The content delivered in response to a query may be, e.g., a video sequence, a movie, a TV show, a document. The quality of the content delivered by CP i is Q_i and captures both the case where the same content is delivered by different CPs and the case where the contents offered in response to a query by different CPs are different. Different contents have naturally a different quality. Even if the content is the same, the quality of CPs may be influenced by several factors: the graphical design of the user interface, the format of the content (e.g., a document delivered as either a pdf or a Word file), the number of clicks needed to reach the content, the information to provide to get the content (e.g., a registration phase), the time elapsed before accessing the content.

Whatever the quality of the content delivered, the user experience is also affected by the advertising included by the CP, e.g., a sequence of a few seconds before a video can be watched. Advertisements are perceived as a nuisance by users, lowering their quality of experience, and have therefore a two-fold effect: positive on the content providers' advertising revenues, but negative on the QoE perceived by users. The total amount A_i of advertising introduced by CP i can be interpreted in two ways, depending on the point of view.

- For the CP: A_i is proportional to the advertising revenue that the CP gets each time a user clicks on the link.
- For users: A_i corresponds to a nuisance, and advertisement is supposed to yield a loss of QoE proportional to A_i .

Through some inessential changes of unit, we consider that CP i earns A_i each time its content is accessed, and that the quality of experience V_i for the user with CP $i \in \mathcal{I}$ is

$$V_i = Q_i - A_i,$$

where $Q_i > 0$ is the intrinsic quality of the content of CP i (that experienced by the user if there were no advertisement). We assume that V_i is an intrinsic characteristic of the *content and advertisement* bundle, depending neither on the decisions taken by the search engine (e.g., its ranking criterion), nor on the user's tastes. We also consider the *relative* amount $a_i := A_i/Q_i$ of advertisement introduced by CP i ; since we limit the advertising level to values that give a non-negative QoE, we have $A_i \in [0, Q_i]$, or equivalently $a_i \in [0, 1]$.

2.2. User's choice of a search engine

We focus only on one SE in this paper, but the user may not choose that SE, e.g., if the quality of the results provided by the SE is bad. We summarize that effect through the average request arrival rate β (for the considered content) that the SE receives per time unit. We assume that it depends on the expected QoE of the user with that SE. We use β as a proxy for the probability that the user chooses that SE among all the possible choices.

2.3. SE ranking policies and click-through-rates

While the user's QoE is determined by both the webpage's intrinsic quality and the amount of advertisement, the ranking criterion influences the behavior of users, and may reduce the visibility of CPs with good QoE. In general, most users click on one of the links in the highest slots. Here we assume that the link clicked after a search depends only on the ranking chosen by the SE based on the *scores* $s = (s_i)_{i \in \mathcal{I}}$ attributed by the SE to each CP. The SE can adopt one of several score functions, depending on its aim. In this section, we examine the most relevant, falling into the two classes of neutral and non neutral behavior.

Rather than addressing separately ranking and the subsequent user clicking behavior, here we aggregate them by considering that the SE allocates the slots (possibly introducing randomness) based on the scores s , so that the proportion of clicks on CP i is

$$C_i = \frac{s_i}{\sum_{j \in \mathcal{I}} s_j}. \quad (1)$$

The average total number of clicks per search may be different from 1 (some users may not click any link while others may try several ones), but we can however consider it equal to 1 without loss of generality, since it can be included in the value of β (to be interpreted as the number of links visited via the SE per time unit), while C_i is the probability (conditional on a link being clicked) that CP i is accessed and describes the overall behavior of the user.

2.3.1. Neutral ranking behavior. Search neutrality should correspond to a situation where QoE is the only thing that matters when ranking, without any consideration for profit. The

resulting score, equalling the QoE, and the proportion C_i of clicks on CP i from Eq. (1) are

$$s_i = V_i, \quad C_i = \frac{V_i}{\sum_{j \in \mathcal{I}} V_j} = \frac{Q_i(1 - a_i)}{\sum_{j \in \mathcal{I}} Q_j(1 - a_j)}. \quad \forall i \in \mathcal{I}. \quad (2)$$

2.3.2. Non-neutral ranking strategies. We say that the SE is non-neutral when the scores are not the QoE values $(V_i)_{i \in \mathcal{I}}$. In particular, we investigate the case when the ranking criterion considers that the SE gets a share b_i of the advertising revenue of each CP $i \in \mathcal{I}$. Both parties (the search engine and the content provider) know where the traffic comes from: the SE knows how many times a specific ad was clicked; the CP knows the referral traffic, i.e. which search engine the traffic came from (through the HTTP referer field of the HTTP header). This is an incentive for the SE to favor CPs with a large b_i and large advertising levels. In particular, we consider two possible non-neutral ranking strategies:

- *Revenue-based ranking*, where the SE ranks CPs on the basis of the revenue it can collect from them rather than the quality experienced by the users. The ranking adopted by the SE impacts on the click-through-rate, and therefore on the revenues of CPs. If the search engine receives money by content providers, it has a real interest in favoring those that may generate more revenues. The ranking scores and the resulting click-through rate are

$$s_i = b_i A_i = b_i Q_i a_i, \quad C_i = \frac{b_i Q_i a_i}{\sum_{j \in \mathcal{I}} b_j Q_j a_j}, \quad \forall i \in \mathcal{I}. \quad (3)$$

- *Weighted-QoE ranking*, where the SE modifies the neutral ranking rule, introducing a bias to favor the CPs that provide larger revenues. The bias is modelled through the corrective factor b_i in the neutral rule (2), which may be interpreted as the share of the CP controlled by the SE: in the search neutrality debate search engines may be accused of favoring the contents they (partially) own. The scores and the click-through rate are

$$s_i = b_i V_i = b_i(Q_i - A_i) = b_i Q_i(1 - a_i), \quad C_i = \frac{b_i Q_i(1 - a_i)}{\sum_{j \in \mathcal{I}} b_j Q_j(1 - a_j)} \quad \forall i \in \mathcal{I}. \quad (4)$$

2.4. Content providers: revenues and strategies

The expected revenues (per time unit) of a content provider $i \in \mathcal{I}$ are denoted by R_i . Since we are considering free content in this work, CPs' revenues only come from advertising and are proportional to the amount of advertisement added to their content, but also to the number of clicks they receive per time unit, depending on the SE ranking through (1).

After deducting the fee paid to the SE, the average revenues per time unit of any CP are

$$R_i = \beta C_i A_i(1 - b_i) = \beta C_i Q_i a_i(1 - b_i). \quad (5)$$

The revenue of a CP i depends on its strategic choice a_i , but also on the amount of advertising $(a_j)_{j \in \mathcal{I}, j \neq i}$ set by the other CPs, through the proportion of clicks C_i to that CP defined in (3). We assume that CPs are able to evaluate the strategies of their opponents by visiting their pages, measuring the advertising level, and adapting their strategies. Here we model those interactions among CPs as a noncooperative game [Osborne and Rubinstein 1994], where each CP chooses its advertising load a_i to maximize its revenues, and distinguish two sub-cases, where content providers act as either *price takers* or *price setters*.

2.4.1. Price-taking content providers. The search engine market is currently dominated by Google: the statistics for the period October 2011 - March 2012 give Google a share of 80.39% (see http://www.statowl.com/search_engine_market_share.php), with a normalized Hirschmann-Herfindahl index of 0.66, which indicates strong dominance. In the short-term, we can expect that this quasi-monopoly situation would remain even if the quality of the results displayed were affected by a change in the ranking policy (from neutral to non-neutral). In our model, this can be interpreted as the rate β of requests not being affected

by the quality of the results (in fact, under a monopoly you have no choice but submit your requests to the only search engine, whatever its quality). Actually, that rate may still vary with the average QoE, but CPs do not consider that effect when deciding their advertising policy. That bias can, e.g., stem from a large number of CPs (hence the individual effect of each CP on β is small, and neglected). In that sense, CPs are price takers: they do not consider the effect of their own actions on the global *price* (here, the search rate β).

2.4.2. Price-setting content providers. In this model, content providers do anticipate the effect of advertising strategies on the global success of the SE, embodied by $\beta = \beta(a)$, where $a = (a_i)_{i \in \mathcal{I}}$ is the advertising profile of CPs. Each CP i chooses then its advertising level a_i so as to maximize

$$R_i = \beta(a_i, a_{-i})C_i(a_i, a_{-i})a_i(1 - b_i), \quad (6)$$

where a_{-i} is the profile of advertising strategies of all CPs but i , i.e., $a = (a_i, a_{-i})$.

In such a case, CPs are said to be price setters, since they are aware of their contribution to the search rate β , which may decrease as the amount of advertisement increases. Acting as price setters corresponds to making strategic moves with an eye on the long term.

3. EQUILIBRIUM ADVERTISING STRATEGIES OF PRICE-TAKING CONTENT PROVIDERS

In this section, we investigate the case when CPs are price takers, i.e., they treat the total request rate β as a constant when determining their advertising strategy. We study the response of CPs to the ranking strategy of SEs as a non cooperative game. Content providers act as the players using the level of advertising as their strategic leverage to maximize their revenues. The game is solved by searching for a Nash equilibrium, after identifying the best response function of each player. For each ranking policy, we will provide results for the general case, before treating two specific situations:

- the symmetric case, where all CPs are identical;
- the duopoly case where only two CPs compete.

3.1. Neutral ranking

Since the multiplicative factor $(1 - b_i)$ is constant, and β is considered as constant by CPs in this section, the quantity that the i -th content provider intends to maximize, under neutral ranking with the scoring function defined in (2), is the utility proportional to

$$U_i := \frac{R_i}{\beta(1 - b_i)} = a_i Q_i C_i = a_i Q_i \frac{V_i}{\sum_{j \in \mathcal{I}} V_j} = a_i Q_i \frac{Q_i(1 - a_i)}{\sum_{j \in \mathcal{I}} Q_j(1 - a_j)}. \quad (7)$$

We can establish here the existence of a (non-trivial) Nash equilibrium for the noncooperative game played among CPs.

PROPOSITION 3.1. *When the search engine adopts a neutral ranking (i.e., based on relevance), there exists at least one Nash equilibrium $a^{NE} \in (0, 1)^{|\mathcal{I}|}$ in the noncooperative game played by CPs setting their advertising level, and any Nash equilibrium is such that*

$$1/2 < 1 + \phi_i - \sqrt{\phi_i^2 + \phi_i} \leq a_i^{NE} < 1, \quad \phi_i := \sum_{j \in \mathcal{I} \setminus \{i\}} \frac{Q_j}{Q_i}. \quad (8)$$

PROOF. When all the other CPs set their advertising quantities, the i -th content provider seeks the quantity a_i maximizing U_i , that is, its best-response to the others' strategic choices.

The case when all competitors of CP i set their advertising level to the maximum value $a_j = 1$ is degenerate. In that case, $C_i = 1$ when $a_i < 1$, and $U_i = a_i Q_i$ is strictly increasing in a_i . But C_i is not defined for $a_i = 1$, hence no exact best-response exists. However, that case is not a problem, since the strategy $a_i = 1$ is dominated for each CP i , and strictly dominated when at least one opponent j sets $a_j < 1$. It is therefore an unlikely situation.

We now consider the case when at least one CP $j \neq i$ sets $a_j < 1$. In that case, U_i is a continuous function of a_i , as can be easily seen in (7).

We remark that $U_i = 0$ when $a_i = 0$ and when $a_i = 1$, and that $\frac{\partial^2 U_i}{\partial a_i^2}$ has the same sign as $a_i Q_i - (Q_i + \sum_{j \in \mathcal{I} \setminus \{i\}} Q_j (1 - a_j))$, that is strictly negative. As a result, U_i is a strictly concave function of a_i on the interval $[0, 1]$, and has a unique maximum that is in $(0, 1)$.

Therefore, the best response of CP i is the only solution in $(0, 1)$ of the equation

$$\frac{\partial U_i}{\partial a_i} = \frac{\partial}{\partial a_i} \left[a_i Q_i \frac{Q_i (1 - a_i)}{\sum_{j \in \mathcal{I}} Q_j (1 - a_j)} \right] = 0, \quad (9)$$

which brings a quadratic equation in the advertising quantity

$$a_i^2 - 2 \left[1 + \sum_{j \in \mathcal{I} \setminus \{i\}} \frac{Q_j}{Q_i} (1 - a_j) \right] a_i + 1 + \sum_{j \in \mathcal{I} \setminus \{i\}} \frac{Q_j}{Q_i} (1 - a_j) = 0. \quad (10)$$

The larger solution of the quadratic equation is to be discarded, since it would give $a_i > 1$. The smaller one is then in $(0, 1)$, and gives us the best response function for CP i :

$$a_i^{\text{BR}} = 1 + \psi_i - \sqrt{(1 + \psi_i)^2 - (1 + \psi_i)} = 1 + \psi_i - \sqrt{\psi_i (1 + \psi_i)}, \quad (11)$$

$$\text{with} \quad \psi_i := \sum_{j \in \mathcal{I} \setminus \{i\}} \frac{Q_j}{Q_i} (1 - a_j). \quad (12)$$

For convenience we define $f(x) := 1 + x - \sqrt{x(1+x)}$, so that $a_i^{\text{BR}} = f(\psi_i)$.

Differentiating f , we get for $x > 0$

$$\frac{df(x)}{dx} = 1 - \frac{2x+1}{2\sqrt{x(1+x)}} = 1 - \frac{x+1/2}{\sqrt{(x+1/2)^2 - 1/4}} < 0, \quad (13)$$

therefore f is strictly decreasing on \mathbb{R}^+ , and thus

$$\lim_{x \rightarrow \infty} f(x) = 1/2 < f(\phi_i) < a_i^{\text{BR}} = f(\psi_i) < f(0) = 1, \quad \phi_i := \sum_{j \in \mathcal{I} \setminus \{i\}} \frac{Q_j}{Q_i}. \quad (14)$$

Now let us consider a small $\epsilon \in (0, 1)$, and consider any strategy vector a in the compact set $[0, 1 - \epsilon]^{|Z|}$. From (12), $\psi_i \geq (n-1) \frac{Q_{\min}}{Q_{\max}} \epsilon$, where $Q_{\min} := \min_{j \in \mathcal{I}} Q_j$ and $Q_{\max} := \max_{j \in \mathcal{I}} Q_j$. From (11), the best-response of each CP i to that strategy vector equals

$$a_i^{\text{BR}} = f(\psi_i) \leq f\left((n-1) \frac{Q_{\min}}{Q_{\max}} \epsilon\right).$$

Remark from (13) that the derivative of f is continuous on $(0, +\infty)$ and tends to $-\infty$ at 0, therefore we can find ϵ small enough so that $f\left((n-1) \frac{Q_{\min}}{Q_{\max}} \epsilon\right) \leq f(0) - \epsilon$, using the continuity of f . Since $f(0) = 1$, we obtain that the best-response correspondence

$$G : \begin{cases} [0, 1 - \epsilon]^{|Z|} \mapsto [0, 1]^{|Z|} \\ a \mapsto (a_i^{\text{BR}})_{i \in \mathcal{I}} \end{cases}$$

is such that $f([0, 1 - \epsilon]^{|Z|}) \subset [0, 1 - \epsilon]^{|Z|}$. Since G is continuous and $[0, 1 - \epsilon]^{|Z|}$ is a compact convex subset of $\mathbb{R}^{|Z|}$, from Brouwer's fixed point theorem, it has a fixed point that constitutes a Nash equilibrium with strategies $a_i \in [0, 1)$.

The lower bound in equilibrium comes from the lower bound of Equation (14). \square

Though not solvable analytically, the collection of best response functions (11) builds a system of nonlinear equations in the a_i 's, to be solved numerically.

3.1.1. Symmetric case. We can consider the symmetric case where all CPs have the same intrinsic quality, so that $Q_i = Q$ in Equation (10) and the utility for the i -th provider becomes as follows, allowing for the game to be fully solved:

$$U_i = a_i \frac{1 - a_i}{\sum_{j \in \mathcal{I}} (1 - a_j)}.$$

PROPOSITION 3.2. *When the search engine adopts a neutral ranking, the noncooperative game played by CPs setting their advertising level has the unique Nash equilibrium*

$$a_i^{NE} = \frac{n}{2n - 1} \quad \forall i \in \mathcal{I}, n := |\mathcal{I}|. \quad (15)$$

PROOF. The revenue optimization procedure leads to the equation

$$(1 - 2a_i) \sum_{j \in \mathcal{I}} (1 - a_j) + a_i(1 - a_i) = 0. \quad (16)$$

This implies that for any $i, k \in \mathcal{I}$, we have

$$(1 - 2a_i) \sum_{j \in \mathcal{I}} (1 - a_j) + a_i(1 - a_i) = (1 - 2a_k) \sum_{j \in \mathcal{I}} (1 - a_j) + a_k(1 - a_k),$$

which yields $(a_k - a_i) \left(a_k + a_i - 1 + 2 \sum_{j \in \mathcal{I}} (1 - a_j) \right) = 0$.

But from Proposition 3.1, we know that at a Nash equilibrium $a_j > 1/2$ for all $j \in \mathcal{I}$, hence the right factor is strictly positive, and $a_i = a_k$: Nash equilibria are necessarily symmetric, of the form $a_i = a$ for all $i \in \mathcal{I}$. Plugging that condition into (16), we obtain a unique equilibrium, where the optimal advertising quantity a for any provider is as in (15). \square

The Nash equilibrium advertising level of (15) is a decreasing function of the number of content providers: each content provider is led to stuff less advertisements as the competition level (number of content providers) grows. In the limit, when the number of competitors becomes very large, we have the optimal advertising quantity that cuts by half the QoE with respect to the upper bound represented by intrinsic quality

$$a_{\lim} = \lim_{n \rightarrow \infty} a = \frac{1}{2}.$$

In the symmetric case, the utility for each content provider is

$$U_i = a \frac{1 - a}{n(1 - a)} = \frac{1}{2n - 1},$$

while the cumulated utility of the bunch of content providers (recall that the revenue of each provider is its utility multiplied by $\beta(1 - b)$) is

$$U = \sum_{i \in \mathcal{I}} U_i = \frac{n}{2n - 1}.$$

Since utilities are proportional to revenues as in Eq. (7) and can be taken as a proxy for them, that last expression shows that the aggregated revenue shrinks as the number of players grows: the overall utility reduces by $1/3$ when there are just two providers, but by $1/2$ when the number of players gets very large.

3.1.2. Duopoly case. Another special case of interest is duopoly, where just two content providers (with different quality) are present. In fact, the presence of high fixed costs reducing profit margins may raise barriers to the entrance of new players and favour a monopoly or duopoly situation (a case of high fixed costs and low marginal costs is presented in [Naldi and D’Acquisto 2008]). In this duopoly case, the best response functions (11) become

$$a_1 = 1 - \frac{Q_2}{Q_1}(1 - a_2) \left[\sqrt{1 + \frac{Q_1}{Q_2} \frac{1}{1 - a_2}} - 1 \right]$$

$$a_2 = 1 - \frac{Q_1}{Q_2}(1 - a_1) \left[\sqrt{1 + \frac{Q_2}{Q_1} \frac{1}{1 - a_1}} - 1 \right].$$

Again, we can solve that system of nonlinear equations numerically, finding the Nash equilibrium as the intersection (if any) between the best responses, as in Figure 1 (left). The best-response functions for the duopoly are shown when the providers have equal intrinsic quality and have the same shape in all the other cases we have examined. Though the curves cross in two points, the solution leading to $a_1 = a_2 = 1$ is to be discarded, since, as in the proof of Proposition 3.1, there is no real best-response for CP i when the opponent $j \neq i$ sets $a_j = 1$, with the strategy $a_i = 1$ being dominated, and strictly dominated if $a_j < 1$.

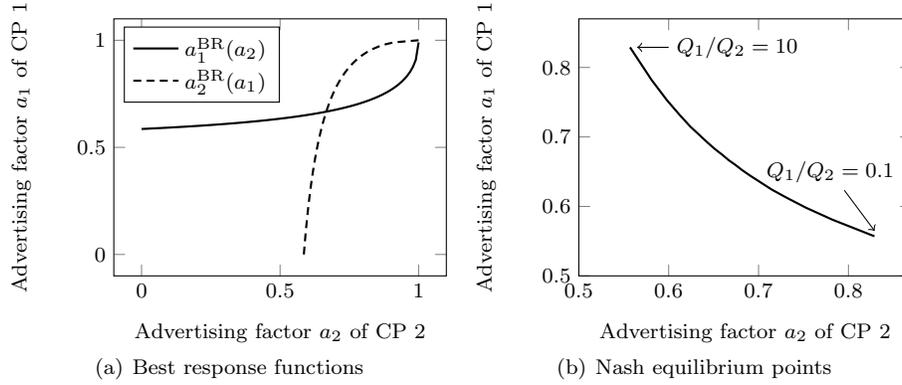


Fig. 1. Best response functions in the symmetric duopoly case (left) and locus of Nash equilibrium points in a duopoly, both under neutral behavior

In Figure 1 (right), we see how the Nash equilibrium point moves as the differences in quality between the two providers change. A content provider with higher intrinsic quality can increase its advertising load. In the symmetric case ($Q_1 = Q_2$), we obtain the Nash equilibrium point $a_1 = a_2 = 2/3$ from Equation (15) with $n = 2$.

3.1.3. Numerical study. We now go back to the general asymmetric case of n content providers and assess the presence and characteristics of Nash equilibrium in typical scenarios. For that purpose, we consider two types of repartition of the intrinsic quality among CPs: linear and geometric. Without loss of generality, we sort the CPs in decreasing order of quality: CP 1 exhibits the largest intrinsic quality, and CP n has the lowest one. We can define the quality of the generic i -th content provider as a function of the two bounds Q_1 and Q_n . In the linear model, the intrinsic quality of the i -th content provider is

$$Q_i = Q_1 - \frac{i-1}{n-1}(Q_1 - Q_n). \quad (17)$$

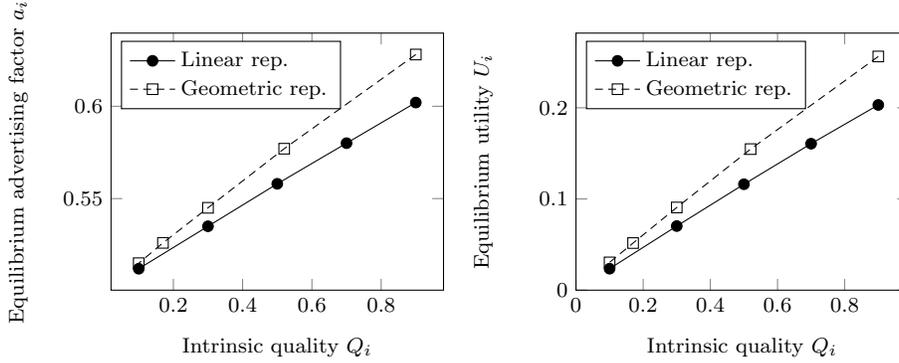


Fig. 2. Impact of the intrinsic quality on the game outcome (left) and utility (right), for price-taking CPs and a neutral ranking

In the geometric model, we have instead

$$Q_i = Q_1 \exp\left(-\frac{i-1}{n-1} \ln(Q_1/Q_n)\right).$$

Note that we consider here that the share of benefits taken by the search engine is the same for all CPs, i.e. $b_i = b$ for all $i \in \mathcal{I}$.

We report here the case of 5 content providers, with $Q_1 = 0.9$ and $Q_5 = 0.1$. We use the solution (11) to see how the intrinsic quality influences the optimal amount of advertisement. Both in the linear and the geometric case, we find a single Nash equilibrium. The relation between the intrinsic quality and the amount of advertising a_i is shown in Figure 2: content providers with larger intrinsic quality put more advertising, though the sensitivity is quite small: with a ninefold increase in the intrinsic quality the advertising factor increases by just 17.58% in the linear model and by 21.94% in the geometric one.

Utility (and therefore, the revenue of the CP) is also affected by the intrinsic quality, as shown in Figure 2 (right). Though both trends are approximately linear, the more uneven repartition of qualities in the geometric case leads to wider imbalances in the repartition of utilities. While the ratio Q_{\max}/Q_{\min} of extreme intrinsic qualities is 9 in both cases, the range of utility $U_{\max} - U_{\min}$ is larger for the geometric repartition (though the high-to-low ratio for utility is 8.63 for the linear case and just 8.41 for the geometric one)

After examining the impact of the intrinsic quality on the individual strategic decisions, we now consider the impact on the quality perceived by users. For CP i , the introduction of advertisement lowers its quality from the intrinsic value Q_i to the QoE $V_i = Q_i(1 - a_i)$. In the above example, we considered a wide range for intrinsic quality values to examine if the introduction of advertisement (which is the leverage through which content providers seek their maximal profits) magnifies those differences in quality or levels them out.

In order to assess that impact, we employ the Herfindahl-Hirschman Index, which measures the concentration of a market among a number of competitors (i.e., the level of competition) and is the most sensitive among such indicators [Naldi 2003]. Given a set of market shares $\{m_1, m_2, \dots, m_n\}$, which satisfy the constraint $\sum_{i \in \mathcal{I}} m_i = 1$, the HHI is

$$\text{HHI} = \sum_{i \in \mathcal{I}} m_i^2, \quad (18)$$

and lies in the interval $[0, 1]$. Higher values of the HHI indicate a larger degree of concentration (hence a lower level of competition). Here we do not have the market shares (expressed as fractions of the overall revenues), but we consider utility values as their proxy. In fact,

Table I. Attraction power concentration (price-taking CPs)

	Linear repartition	Geometric repartition
HHI _Q	0.264	0.305
HHI _V	0.253	0.279
HHI _U	0.261	0.298

utilities are proportional to revenues, the proportionality constant being equal for all competitors. After normalizing the utility values to their sum, we get figures equal to the market shares. In order to distinguish the HHI computed by using utilities from that computed using the market shares (though they lead to the same numerical result), we use the definition

$$\text{HHI}_U = \sum_{i \in \mathcal{I}} \left(\frac{U_i}{\sum_{j \in \mathcal{I}} U_j} \right)^2 = \frac{\sum_{i \in \mathcal{I}} U_i^2}{(\sum_{i \in \mathcal{I}} U_i)^2}.$$

In addition to measuring the concentration of the market, we can use the HHI to measure the attraction power due to quality. Though used typically for market shares, the HHI is basically a concentration metric, not unlike the classical Gini index (which is used, e.g., to measure the distribution of wealth). For example, the HHI has been used to measure the distribution of patents [Pilkington and Liston-Heyes 2004]. We can draw a similarity between market shares and quality values. Let's consider two cases at either end of a market structure: quasi monopoly and perfect competition. In a quasi monopoly a single company has nearly 100% of the market and its competitors have negligible market shares. The analogous case for quality is where a SE has a quality much higher than the other SEs. Instead, in a perfect competition, the market is divided equally among all the companies. The analogous case for quality is where the SEs have all the same quality. In this context, the HHI can be used to measure the degree of concentration of quality among the content providers. If we replace the market shares in the definition (18) with quality values, we obtain two quality-based HHI employing respectively the intrinsic and the perceived quality.

$$\text{HHI}_Q = \sum_{i \in \mathcal{I}} \left(\frac{Q_i}{\sum_{j \in \mathcal{I}} Q_j} \right)^2 = \frac{\sum_{i \in \mathcal{I}} Q_i^2}{(\sum_{i \in \mathcal{I}} Q_i)^2},$$

$$\text{HHI}_V = \sum_{i \in \mathcal{I}} \left(\frac{V_i}{\sum_{j \in \mathcal{I}} V_j} \right)^2 = \frac{\sum_{i \in \mathcal{I}} V_i^2}{(\sum_{i \in \mathcal{I}} V_i)^2},$$

Larger values of HHI_Q mean higher imbalances in quality. By comparing HHI_Q (no advertising) with HHI_V (including advertising), we can assess the effect of advertising choices of all content providers on quality distribution. We report the results in Table I. Under both the linear and the geometric model, the HHI index is larger for the intrinsic quality than for the QoE: the introduction of advertisement brings along a slight levelling of the quality perceived by the user. In Table I, we also report the HHI (denoted by HHI_U) pertaining to market concentration. We see that the concentration is somewhat intermediate between that of the intrinsic quality and that of the perceived quality: the repartition of utility is less affected by the introduction of advertisement than the quality perceived by users.

3.2. Non-neutral behavior: revenue-based ranking

When the ranking is only based on the potential revenue for the SE, and scores are taken from (3), with the position $X_i := b_i Q_i$, the utility of the content provider is given by:

$$U_i = \beta A_i C_i (1 - b_i) = \beta \frac{1 - b_i}{b_i} \frac{a_i^2 (b_i Q_i)^2}{\sum_{j \in \mathcal{I}} a_j Q_j b_j} = \frac{1 - b_i}{b_i} \frac{a_i^2 X_i^2}{\sum_{j \in \mathcal{I}} a_j X_j}.$$

The following proposition shows that the non-neutral behavior incentivizes CPs to increase their advertising level with respect to the neutral case. That incentive is indeed twofold, since a larger a_i yields more revenue per click, but also attracts more clicks because of the non-neutral ranking (where s_i increases with a_i).

PROPOSITION 3.3. *If the search engine adopts a revenue-based ranking, the noncooperative game played by price-taking CPs setting their advertising level has a unique Nash equilibrium where each CP sets its advertising level to the maximum possible value, $a_i = 1, \forall i \in \mathcal{I}$.*

PROOF. We simply see that the revenue R_i of a CP $i \in \mathcal{I}$ is strictly increasing in a_i : it is indeed null for $a_i = 0$, and for $a_i > 0$ it is strictly positive with

$$\frac{\partial U_i}{\partial a_i} = \beta \frac{1 - b_i}{b_i} X_i^2 \frac{2a_i \sum_{j \neq i} a_j X_j + a_i^2 X_i}{(\sum_{j \in \mathcal{I}} a_j X_j)^2} > 0. \quad \square$$

We immediately remark that, in that case, the resulting QoE is $V_i = 0$ for each CP i . The revenue of each CP $i \in \mathcal{I}$ becomes $R_i = \beta \frac{1 - b_i}{b_i} \frac{X_i^2}{\sum_{j \in \mathcal{I}} X_j}$.

If all CPs transfer to the search engine the same share of their revenue, i.e. $b_i = b$ for all i , then we obtain $R_i = \beta(1 - b) \frac{Q_i^2}{\sum_{j \in \mathcal{I}} Q_j}$. Additionally, if as in Section 2.3.1 we consider the symmetric case, where the n content providers have the same intrinsic quality, i.e. $Q_i = Q$, then each CP gets a revenue $R_i = \beta(1 - b) \frac{Q}{n}$.

Note that when β varies (increases) with the average QoE for users, that revenue can be much smaller than initially expected by CPs, i.e., as price takers made their strategic choices considering β to be a constant. For example, if as in Section 4 we take β proportional to the expected user QoE, then the final outcome is a situation where users prefer not to use the search engine (since $\beta = 0$), and the CPs make no revenue.

3.3. Non-neutral behavior: weighted-QoE ranking

We investigate here the situation where the ranking scores are taken from (4), i.e., the search engine considers the QoE as in the neutral case, but introduces some weights among them so as to favor the CPs it has most interest in.

The resulting proportion of clicks per search on CP i is then

$$C_i = \frac{Q_i(1 - a_i)b_i}{\sum_{j=1}^n Q_j(1 - a_j)b_j}.$$

Accordingly, the utility of the content provider is given by the difference between what it receives through advertising and what it passes to the search engine:

$$U_i = a_i C_i (1 - b_i) = \frac{Q_i a_i b_i (1 - a_i) (1 - b_i)}{\sum_{j=1}^n Q_j (1 - a_j) b_j}. \quad (19)$$

Again, each content provider seeks to maximize its revenue by setting the quantity of advertising. The analysis follows that carried out in Section 3.1, where for each CP i the parameter Q_i is replaced by $b_i Q_i$. In particular, the best response function of CP i is then

$$a_i = 1 + \sum_{\substack{j=1 \\ j \neq i}}^n \frac{Q_j b_j}{Q_i b_i} (1 - a_j) - \sqrt{1 + \sum_{\substack{j=1 \\ j \neq i}}^n \frac{Q_j b_j}{Q_i b_i} (1 - a_j)} \sqrt{\sum_{\substack{j=1 \\ j \neq i}}^n \frac{Q_j b_j}{Q_i b_i} (1 - a_j)}.$$

We then have the counterpart of Proposition 3.1.

Table II. Intrinsic quality and revenue transfer to the search engine in the non-neutral case study

Intrinsic quality Q_i		Revenue transfer b_i	
Linear	Geometric	Scenario A	Scenario B
0.1	0.1	0.1	1
0.3	0.17	0.1	0.1
0.5	0.3	0.1	0.1
0.7	0.52	0.1	0.1
0.9	0.9	1	0.1

PROPOSITION 3.4. *When the search engine performs a weighted-QoE ranking with weights $(b_i)_{i \in \mathcal{I}}$, the noncooperative game played by price-taking CPs fixing their advertising level has at least one Nash equilibrium $a^{NE} \in (0, 1)^{|\mathcal{I}|}$. More precisely, there exists a Nash equilibrium, and any Nash equilibrium is such that*

$$1/2 < 1 + \bar{\phi}_i - \sqrt{\bar{\phi}_i^2 + \bar{\phi}_i} \leq a_i^{NE} < 1, \quad \bar{\phi}_i := \sum_{j \in \mathcal{I} \setminus \{i\}} \frac{b_j Q_j}{b_i Q_i}.$$

Remark that the symmetric case, where the n content providers have the same intrinsic quality and transfer to the search engine the same share of their utility (i.e., $Q_i = Q$, $b_i = b$) gives the exact same case as the one analyzed in Section 3.1.1 for the neutral ranking.

3.3.1. Numerical study. We examine now the asymmetric case, solving the game numerically. We use the case of Section 3.1.3, with 5 content providers, one of which is owned by the search engine (it transfers all its utility to it) and all the others pay the same share $b_i = 0.1$. Their intrinsic quality follows either a linear or a geometric trend, with the content provider owned by the search engine exhibiting either the highest intrinsic quality (Scenario A) or the lowest one (Scenario B), as reported in Table II, i.e., the two extreme cases. The two most important questions here concern the impact of vertical integration on the CP's results. If the CP has the worst intrinsic quality, how much does it gain from being the search engine's favorite one? And if the CP is instead already the best in the group, does it gain or does it lose from being owned by the SE? In this section, we examine the advertising choices made by the content providers and their impact on utility.

We report the results of the game, namely the resulting advertising factor a_i , in Figure 3 for the two scenarios. In both cases, the content providers not owned by the search engine use an advertising factor that increases slightly with the intrinsic quality through a roughly linear trend. We expect that in the absence of vertical integration the CP would follow the trend depicted for the other CPs. Instead, we see that with vertical integration the behavior of the vertically integrated CP differs markedly from that trend: the content provider owned by the search engine increases substantially its advertising load, especially when its intrinsic quality is large: the boost is much higher in Scenario A (where the vertically integrated content provider has an intrinsic quality of 0.9) than in Scenario B (where that intrinsic quality is just 0.1). For example, let's consider the linear case in Figure 3a with Scenario A, with the behavior of the CPs represented by black dots. We see that there is a linear trend: the advertising factor is expected to grow linearly with the intrinsic quality. On this basis, we would expect the vertically integrated CP, whose intrinsic quality is 0.9 (as shown in Table II), to follow that linear trend (if it were not vertically integrated) and end up with an advertising factor around 0.55. Instead we see its black dot in the upper right corner just below of 0.8. That means that the vertically integrated CP is led to put much more advertising in its content because of vertical integration. The same can be said for Scenario B, where the vertically integrated CP is instead that with the lowest intrinsic quality.

If we consider the net utility (that remaining after transferring a share to the search engine), the vertically integrated content provider has of course zero utility. But we get a

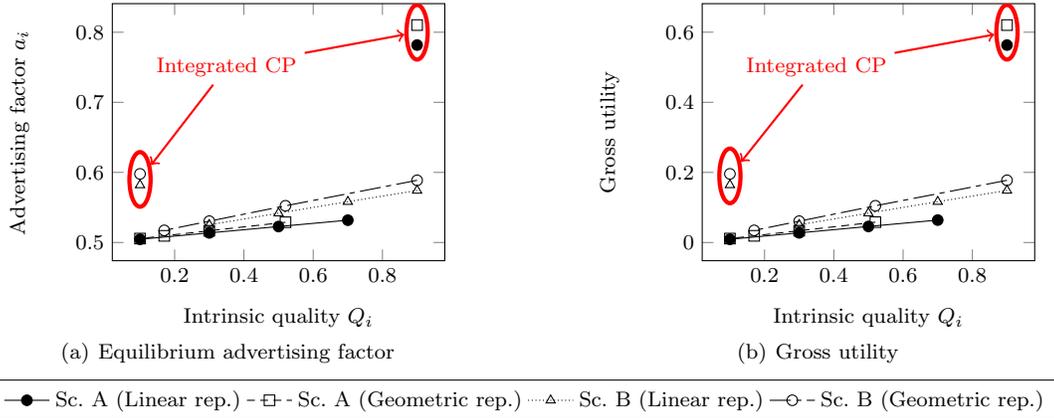


Fig. 3. Impact of the intrinsic quality on the advertising factor (left) and gross utility (right), with price-taking CPs, weighted-QoE ranking.

Table III. Concentration of attraction power with price-taking CPs and weighted-QoE based ranking

Scenarios	A linear	B linear	A geometric	B geometric
HHI_Q	0.264	0.264	0.305	0.305
HHI_V	0.247	0.259	0.249	0.292
$\text{HHI}_{\hat{U}}$	0.644	0.227	0.708	0.261

better view of the competition between content providers by considering the gross utility (that obtained prior to paying the tax to the search engine), defined after Equation (19).

In Figure 3 (right), we see that for all providers but that owned by the search engine the utility grows roughly linearly with the intrinsic quality, though at a faster rate in the geometric case. The utility of the content provider owned by the search engine is instead boosted by the larger cash flow it transfers to the search engine, which in turn raises its score in the non-neutral case. The boost is again larger in Scenario A, i.e., when the vertically integrated content provider has the largest intrinsic quality. However, even in Scenario B the boost is enough to bring the vertically integrated content provider to include as much advertisement as the content provider with the highest intrinsic quality.

We can now perform the same concentration analysis as in the neutral case. For the gross utility defined in (19), we can similarly define the HHI

$$\text{HHI}_{\hat{U}} = \frac{\sum_{i \in \mathcal{I}} \hat{U}_i^2}{\left(\sum_{i \in \mathcal{I}} \hat{U}_i\right)^2}.$$

The values obtained for HHIs are shown in Table III. The comparison of HHI_Q and HHI_V (before and after introducing advertisements) tells us that the introduction of advertisement levels the quality perceived by users, since $\text{HHI}_V < \text{HHI}_Q$. Instead, the market concentration is altered, depending on which content provider is owned by the search engine. If the search engine owns the content provider with the highest intrinsic quality (Scenario A), that further boosts its score and its utility, leading to a quite stronger concentration: $\text{HHI}_{\hat{U}}$ is more than twice as large as HHI_Q . When the search engine owns the content provider with the lowest intrinsic quality (Scenario B), the utility of the latter is likewise pushed up, but this leads to a more balanced repartition and a diminishing HHI.

4. EQUILIBRIUM ADVERTISING STRATEGIES OF PRICE-SETTING CONTENT PROVIDERS

After examining the situation of a quasi-monopolistic search engine market in Section 3, in this section we consider that users may select a different search engine. We define a game model and find the best response functions, for the cases of neutral and non neutral behavior, examining the resulting advertising strategies in some scenarios.

The propensity of a user to use the search engine under consideration is still represented by β , assumed here to be proportional to the average QoE of content providers accessed through that search engine (the expected user's QoE when clicking on a link, C_i being the probability that the user ends up visiting CP i):

$$\beta = \sum_i C_i V_i.$$

We recall the general expression (6) for CP i 's revenues

$$R_i = \beta(1 - b_i)C_i a_i, \quad (20)$$

where $C_i = s_i / \sum_{j \in \mathcal{I}} s_j$, and s_i is the score credited to CP i by the SE.

4.1. Neutral behavior

As in Section 3.1, the neutral ranking is based on V_i , hence the proportion of clicks on CP i

$$C_i = \frac{V_i}{\sum_{j \in \mathcal{I}} V_j}.$$

By neglecting the revenue transfer to the search engine in the general expression (20) for the revenues, we can use the utility $U_i = \beta C_i a_i$.

Under that ranking behavior, the revenue of the i -th content providers is proportional to

$$U_i = \frac{\sum_{j \in \mathcal{I}} V_j^2}{\sum_{j \in \mathcal{I}} V_j} \frac{V_i}{\sum_{j \in \mathcal{I}} V_j} a_i, \quad (21)$$

which, by setting $X = \sum_{j \in \mathcal{I}} V_j$ and $Y = \sum_{j \in \mathcal{I}} V_j^2$, can be simplified to $U_i = V_i \frac{Y}{X^2} a_i$.

The i -th content provider optimizes its advertising behavior by maximizing its utility through the appropriate amount of advertising. Remark that $U_i = 0$ if $a_i = 0$ or $a_i = 1$, and $U_i > 0$ when $a_i \in (0, 1)$, therefore if the equation $\partial U_i / \partial a_i = 0$ has a unique solution, that solution would give the best response function $a_i^* = f(a_1, a_2, \dots, a_{i-1}, a_{i+1}, \dots, a_n)$. However, since $\partial U_i / \partial a_i = U_i / a_i + \partial U_i / \partial V_i \cdot \partial V_i / \partial a_i$, and $\partial V_i / \partial a_i = -Q_i$, the optimization equation becomes simply $U_i / (Q_i - V_i) - \partial U_i / \partial V_i = 0$. By expanding this equation, we obtain the following simplified form of the optimization equation for CP i

$$2V_i(V_i X - Y) \left(1 - \frac{V_i}{Q_i}\right) + XY \left(1 - 2\frac{V_i}{Q_i}\right) = 0. \quad (22)$$

If we now replace the full expression for X and Y , and rearrange terms, we obtain a fourth-degree polynomial equation in the QoE

$$\begin{aligned} V_i^4 + \left(2 \sum_{j \in \mathcal{I} \setminus \{i\}} V_j - \frac{Q_i}{2}\right) V_i^3 - \left(\frac{3}{2} Q_i \sum_{j \in \mathcal{I} \setminus \{i\}} V_j\right) V_i^2 + \\ \sum_{j \in \mathcal{I} \setminus \{i\}} V_j^2 \left(\sum_{j \in \mathcal{I} \setminus \{i\}} V_j + \frac{Q_i}{2}\right) V_i - \frac{Q_i}{2} \sum_{j \in \mathcal{I} \setminus \{i\}} V_j \sum_{j \in \mathcal{I} \setminus \{i\}} V_j^2 = 0. \end{aligned}$$

By collecting the n similar expressions for the best response functions of all the content providers, we end up with a system of n polynomial equations, which has to be solved to find Nash equilibria. The system of equations can be solved numerically.

We can consider the special symmetric case by setting $Q_i = Q$, $b_i = b$, and $a_i = a$ in the general Equation (21), since nothing else depends on the specific content provider:

$$U = \frac{nV^2}{nV} \frac{V}{nV} a = \frac{Q}{n} a(1 - a).$$

The first order optimality condition is then $\frac{\partial U}{\partial a} = \frac{Q}{n}(1 - 2a) = 0$, whose solution is $a = \frac{1}{2}$.

4.2. Non-neutral behavior: revenue-based ranking

In that case, the utility function of the search engine becomes

$$U_i = C_i a_i \sum_{j \in \mathcal{I}} C_j V_j = b_i a_i^2 \frac{\sum_{j \in \mathcal{I}} b_j Q_j a_j (1 - a_j)}{(\sum_{k \in \mathcal{I}} b_k a_k)^2}. \quad (23)$$

From the utility maximization condition, we get

$$\frac{\partial U_i}{\partial a_i} = 2b_i a_i \frac{\sum_{j \in \mathcal{I}} b_j Q_j a_j (1 - a_j)}{(\sum_{k=1}^n b_k a_k)^2} + b_i^2 a_i^2 \frac{Q_i (1 - 2a_i) (\sum_{k=1}^n b_k a_k) - 2 \sum_{j \in \mathcal{I}} b_j Q_j a_j (1 - a_j)}{(\sum_{k=1}^n b_k a_k)^3} = 0,$$

which leads to the following optimization equation whose solution should provide the best response function for the i -th content provider

$$2 \sum_{j \in \mathcal{I}} b_j Q_j a_j (1 - a_j) \left(1 - \frac{b_i a_i}{\sum_{k=1}^n b_k a_k} \right) + b_i a_i Q_i (1 - 2a_i) = 0,$$

giving the following third degree polynomial equation in a_i :

$$-2b_i a_i^3 + \left(b_i - 4 \sum_{k \neq i} b_k a_k \right) a_i^2 - 3 \sum_{k \neq i} b_k a_k a_i + 2 \frac{\sum_{k \neq i} b_k a_k \sum_{j \neq i} b_j Q_j a_j (1 - a_j)}{b_i Q_i} = 0.$$

Instead of the numerical approach required in the general case, we can find a simple form of the best response function in the symmetric case. If all the content providers transfer the same percentage of their utility to the search engine and exhibit the same intrinsic quality, we expect their best response function to be the same. After setting $b_i = b$, $Q_i = Q$, and $a_i = a$, and some manipulation, we obtain the symmetric equilibrium advertising level

$$a = 1 - \frac{1}{2n}. \quad (24)$$

When the number of providers grows, the relative amount of advertising tends towards the saturation value 1. If we insert the solution (24) in the general expression of utility (23), we get the utility for the symmetric case

$$U = b a^2 \frac{n b Q a (1 - a)}{n^2 b^2 a^2} = Q \frac{2n - 1}{4n^3}.$$

4.3. Non-neutral behavior: weighted-QoE ranking

With the scores taken from (4), the utility function of CP i becomes

$$U_i = b_i Q_i a_i (1 - a_i) \frac{\sum_{j \in \mathcal{I}} b_j Q_j^2 (1 - a_j)^2}{\sum_{j \in \mathcal{I}} b_j Q_j (1 - a_j)}.$$

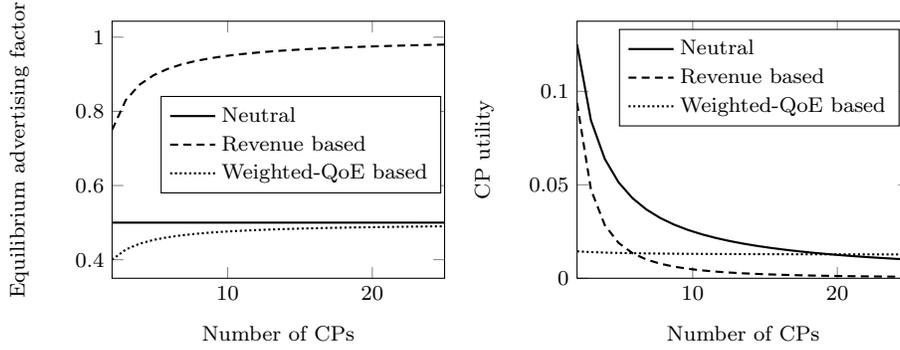


Fig. 4. Nash equilibrium advertising factors (left) and CP utility (right) in the neutral and two non-neutral scenarios, for the symmetric case and price-setting CPs, with $b = 0.1$ and $Q = 1$

If we set, for sake of simplicity, $X = \sum_{j \in \mathcal{I}} b_j Q_j^2 (1 - a_j)^2$ and $Y = \sum_{j \in \mathcal{I}} b_j Q_j (1 - a_j)$, the optimization equation can be written as a fourth-degree polynomial equation in a_i

$$(1 - 2a_i)XY - 2b_i Q_i^2 a_i (1 - a_i)^2 Y + b_i Q_i a_i (1 - a_i)X = 0. \quad (25)$$

If we consider the symmetric case, we have $X = nbQ^2(1 - a)^2$ and $Y = nbQ(1 - a)$, which, when replaced in (25), give the symmetric equilibrium advertising factor

$$a = \frac{n}{2n + 1}, \quad (26)$$

which tends to the limit $a = 1/2$ when the number of providers grows.

In the symmetric case, the utility of each CP is therefore

$$U = bQa(1 - a) \frac{nbQ^2(1 - a)^2}{nbQ(1 - a)} = bQ^2 \frac{n(n + 1)^2}{(2n + 1)^3}.$$

4.4. Neutral vs Non-neutral Ranking

We now compare the ranking strategies when the CPs have identical intrinsic qualities. We plot in Figure 4 the equilibrium advertising levels in the symmetric case vs the number of content providers. We call the equilibria in (24) and (26) respectively as revenue-based and weighted-QoE scoring. Both functions grow with the number of providers, but achieve different values. When the behavior of the SE is purely greedy, the advertising factor starts at 0.75 with two CPs and tends to 1 when the number of CPs grows (under tough competition the QoE gets very close to zero). Instead, if the scoring function includes the QoE, the optimal advertising factor starts at 0.4 under duopoly, grows, but is upper bounded by the value 0.5: users will get a QoE never lower than in the neutral case (half the intrinsic quality value). In the same figure we also plot the gross utility of each CP. A neutral ranking favors CPs, unless the number of competing CPs becomes large (above 20 when $b = 0.1$). In that case, the weighted-QoE based ranking would yield a larger utility to CPs.

We now assess the impact of the ranking policy in a non-symmetric setting, with a linear repartition of quality (17), when the number of CPs changes (the geometric distribution yields similar results). We consider two scenarios: the CP owned by the SE is that with the highest (Figure 5) or the lowest intrinsic quality (Figure 6). We solve the game iteratively. In order to evaluate the impact on users, we define *user welfare* (or, with a slight abuse of vocabulary, the revenue of users) as being equal to β , i.e. the propensity of users to use the search engine. We examine the revenue of the SE (which includes that of its CP), the gross revenue (the aggregated revenue of all CPs but the owned one prior to paying the SE, the

revenue of the SE hence including the integrated CP revenue), and the global revenue, i.e. the sum of both. Notice that, even in the neutral scenario, the equilibrium depends on the CP vertically integrated with the SE, since their revenue is shared between both entities.

When the smallest CP is integrated with the SE, we observe on Figure 6 that all revenues decrease with the number of CPs. In addition, when there are more than three CPs, the ranking policies are Pareto ordered: the revenue-based ranking is worst for every stakeholder, and the neutral one is always preferred. This suggests that there is no need to enforce search neutrality: even for the search engine it is preferable to implement a neutral ranking.

In the case of the biggest CP integrated, we see on Figure 5 that the comparison is not as clear. The revenues are decreasing w.r.t. the number of players for every stakeholder, except the gross revenue, which increases at first: as the number of CPs grows, they get a bigger and bigger proportion of the global revenue at the expense of the SE. The weighted-QoE based ranking provides larger revenues both for the SE and on the overall, probably because owning the greatest quality CP (rather than the smallest quality one) gives the SE a stronger position in the game so as to adopt a non-neutral ranking. Also, users get a larger welfare for any ranking policy compared to Figure 6 and prefer weighted-QoE based ranking. Since the user's welfare corresponds to the propensity of using the SE, the average relevance is better with the non-neutral ranking. This observation, although not intuitive, shows that the use of non-neutral ranking may result in a Nash equilibrium where less advertisement is set compared to the neutral equilibrium. But this holds here because the SE owns the most relevant CP, and therefore should not be taken as a valid argument against neutrality. Indeed, one of the objectives of search neutrality is to enable innovation, by making (relevant) new entrants reachable through the SE: in our case a weighted-QoE based ranking goes against that objective, since the average perceived relevance is lower than in the neutral case in Figure 6 (when the best CPs are not owned by the SE).

When there is no vertical integration, but the CPs do not offer the same intrinsic quality, the results we obtain are very close to the case where the worst CP is vertically integrated to the SE: the integration effects are boosted by the intrinsic quality of the CP.

Finally, note that in both scenarios the revenue of the non-integrated CPs is larger with a neutral ranking, which suggests enforcing neutrality. Indeed, even if a non-neutral stance increases the user's QoE when the SE integrates quality content, it harms the other CPs revenue, possibly preventing them from innovating and improving their quality.

5. CONCLUSIONS

This paper provides a mathematical model for the analysis of different ranking policies by search engines, in a context when content providers have to compete for users, and get revenue through advertising. Depending on the ranking adopted, content providers can choose their advertising level, balancing larger advertising revenues against lower quality-of-experience and less users. We have analyzed the noncooperative game played among content providers in different settings, and studied the corresponding equilibria.

Our results indicate that the neutral ranking provides users with the largest perceived quality-of-experience, which is not surprising. But we also observe that such a ranking policy can be preferred as well by a search engine willing to maximize revenue, a less intuitive outcome: this is true if the search engine does not control the best-performing content providers. However, if the search engine integrates quality content, then it can increase its revenue by switching to a non-neutral ranking; this may even benefit users who will perceive a better quality-of-experience, but would be at the expense of the other content providers, and can then be seen as an impediment to innovation for new entrants.

REFERENCES

- S. Athey and G. Ellison. 2011. Position Auctions with Consumer Search. *The Quarterly Journal of Economics* 126, 3 (2011), 1213–1270.

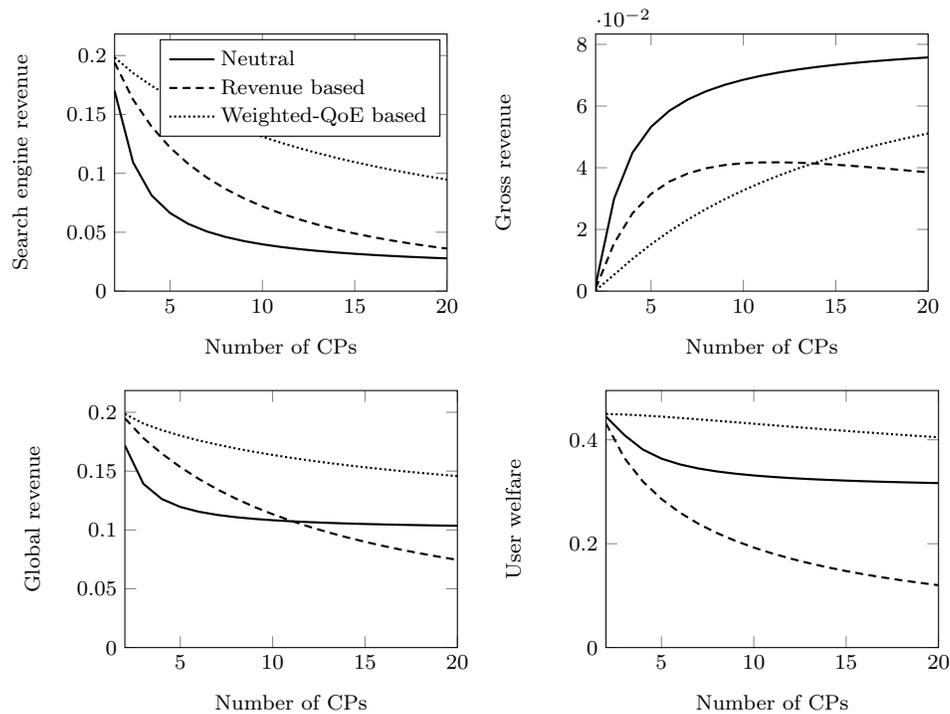


Fig. 5. Utility indices when the highest quality CP is vertically integrated with the SE.

- R. Berman and Z. Katona. 2013. The Role of Search Engine Optimization in Search Marketing. *Marketing Science* 32, 4 (2013), 644–651.
- Y. Chen and C. He. 2011. Paid Placement: Advertising and Search on the Internet*. *The Economic Journal* 121, 556 (2011), F309–F328.
- P. Coucheny, P. Maillé, and B. Tuffin. 2012. Comparison of search engines non-neutral and neutral behaviors. *SIGMETRICS Perform. Eval. Rev.* 40, 2 (Oct. 2012), 14–17.
- G. D’Acquisto, P. Maillé, M. Naldi, and B. Tuffin. 2012. Impact on retail prices of non-neutral wholesale prices for content providers. In *Proceedings of the 8th Euro-NF Conference on Next Generation Internet NGI 2012*. IEEE, Karlskrona, Sweden, 111–117.
- N. Gandal. 2001. The dynamics of competition in the Internet search engine market. *International Journal of Industrial Organization* 19, 7 (2001), 1103 – 1117.
- J. Grimmelmann. 2010. *The next digital decade: Essays on the future of the Internet*. TechFreedom, Washington, DC, Chapter Some skepticism about search neutrality, 435–460.
- A. Hagiu and B. Jullien. 2011. Why do intermediaries divert search? *The RAND Journal of Economics* 42, 2 (2011), 337–362.
- L. D. Intronza and H. Nissenbaum. 2000. Shaping the Web: Why the Politics of Search Engines Matters. *Information Society* 16, 3 (2000), 169–185.
- M. Levene. 2011. *An introduction to search engines and web navigation*. John Wiley, New Jersey.
- P. Maillé, E. Markakis, M. Naldi, G. D. Stamoulis, and B. Tuffin. 2012. Sponsored search auctions: an overview of research with emphasis on game theoretic aspects. *Electronic Commerce Research* 12 (2012), 265–300. Issue 3.
- P. Maillé and B. Tuffin. 2014. *Telecommunication Network Economics: From Theory to Applications*. Cambridge University Press.
- G. A. Manne and J. D. Wright. 2011. *If search neutrality is the answer, what’s the question?* Law and Economics Research Paper Series 11-37. George Mason University.
- M. Naldi. 2003. Concentration indices and Zipf’s law. *Economics letters* 78, 3 (2003), 329 – 334.

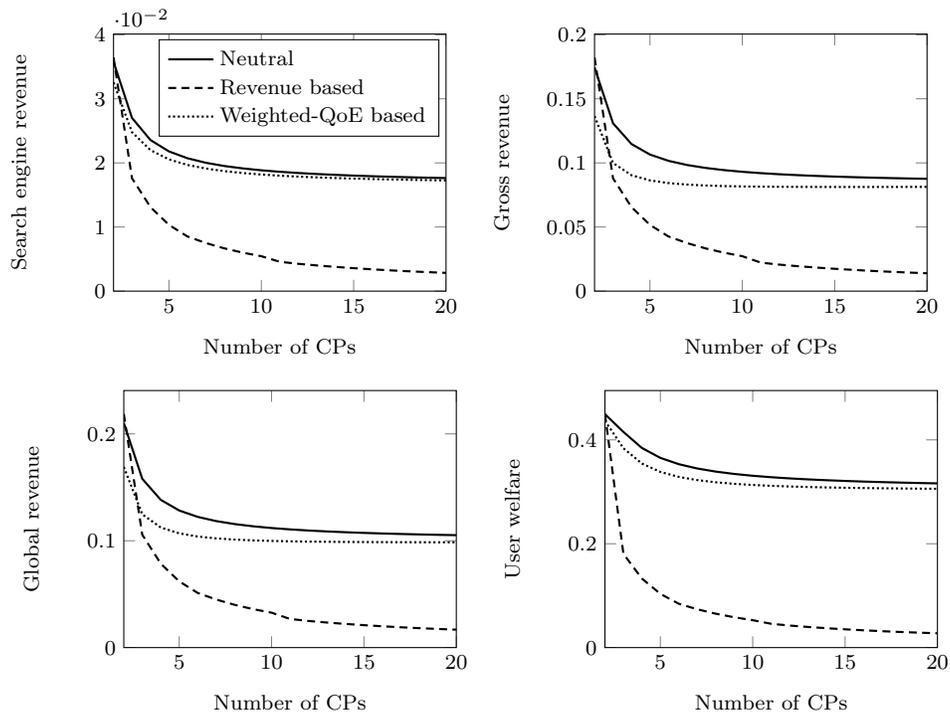


Fig. 6. Utility indices when the smallest quality CP is vertically integrated with the SE.

- M. Naldi and G. D'Acquisto. 2008. Performance of the Vickrey auction for digital goods under various bid distributions. *Performance Evaluation* 65, 1 (2008), 10–31.
- A. Odlyzko. 2009. Network Neutrality, Search Neutrality, and the Never-ending Conflict between Efficiency and Fairness in Markets. *Review of Network Economics* 8, 1 (2009), 40–60.
- M. J. Osborne and A. Rubinstein. 1994. *A Course in Game Theory*. MIT Press, Cambridge, Mass.
- A. Pilkington and C. Liston-Heyes. 2004. Inventive Concentration: An Analysis of Fuel Cell Patents. *Science and Public Policy* 31, 1 (2004), 14–25.
- T.-R. Sheu and K. Carley. 2001. Monopoly Power on the Web - A Preliminary Investigation of Search Engines. In *29th TPRC Conference*. Alexandria, USA.
- R. Telang, U. Rajan, and T. Mukhopadhyay. 2004. The Market Structure for Internet Search Engines. *J. of Management Information Systems* 21, 2 (2004), 137–160.
- A. White. 2013. Search engines: Left side quality versus right side profits. *International Journal of Industrial Organization* (2013), 1–12.

Appendix D

Main publications evoked in Chapter 5

Impact of Competition Between ISPs on the Net Neutrality Debate

Pierre Coucheney, Patrick Maillé, Bruno Tuffin

Abstract—Network neutrality is the topic of a vivid and very sensitive debate, in both the telecommunication and political worlds, because of its potential impact in everyday life. That debate has been raised by Internet Service Providers (ISPs), complaining that content providers (CPs) congest the network with insufficient monetary compensation, and threatening to impose side payments to CPs in order to support their infrastructure costs. While there have been many studies discussing the advantages and drawbacks of neutrality, there is no game-theoretical work dealing with the observable situation of competitive ISPs in front of a (quasi-)monopolistic CP. Though, this is a typical situation that is condemned by ISPs, and, according to them, another reason of the non-neutrality need.

We develop and analyze here a model describing the relations between two competitive ISPs and a single CP, played as a three-level game corresponding to three different time scales. At the largest time scale, side payments (if any) are determined. At a smaller time scale, ISPs decide their (flat-rate) subscription fee (toward users), then the CP chooses the (flat-rate) price to charge users. Users finally select their ISP (if any) using a price-based discrete choice model, and decide whether to also subscribe to the CP service. The game is analyzed by backward induction. As a conclusion, we obtain among other things that non-neutrality may be beneficial to the CP, and not necessarily to ISPs, unless the side payments are decided by ISPs.

Index Terms—Network neutrality, Game theory, Pricing

I. INTRODUCTION

There has recently been a strong debate around the so-called *network neutrality*. The debate has been ignited by the increasing traffic asymmetry between Internet Service Providers (ISPs), mainly due to some prominent and resource consuming content providers (CPs) which are usually connected to a single ISP. The typical example is YouTube (owned by Google), accessed by all users while hosted by a single Tier 1 ISP, and whose traffic now constitutes a non-negligible part of the whole Internet traffic. Another example is the subscription-based video service Netflix, that is in the US the most bandwidth-consuming source of traffic, representing 23.3% of the total Internet traffic in late 2011 [1], while having commercial relationships with only one ISP. For those reasons, there has been a surge of protest among ISPs, complaining that the current Internet business model where ISPs charge both end-users and content providers directly connected to them, and have public peering or transit agreements with other ISPs, is not relevant anymore. The main solution proposed is that ISPs should also charge content providers that are associated

with other ISPs [2], as first advocated by Ed Whitacre (CEO of AT&T) at the end of 2005 [3].

The underlying concern is that investment is made by ISPs but content providers get an important part of the dividends. The revenue arising from on-line advertising (meaning showing graphical ads on regular web pages) is estimated at approximately a \$24 billion in 2009 [4], while textual ads on search pages has led to a combined revenue of \$8.5 billion in 2007 [5], those figures increasing every year. Meanwhile, transit prices - which constitute the main source of revenues for transit ISPs - are decreasing. ISPs argue that there is no sufficient incentive for them to continue to invest on the network infrastructures if most benefits go to content providers. The threat is to lower the quality of service of CPs that do not pay any fee to them, or even to block their traffic. This possibility has led to protests from CPs and user associations, complaining that this might impact the network development and is an impingement of freedom of speech [3]. The debate was thus launched, essentially at the law and policy makers level, to decide whether the Internet should be *neutral*, i.e., all packets should receive equal treatments in terms of price and service. In the US, the Federal Trade Commission (FTC) released in 2007 a report not supporting neutrality constraints, increasing the debate at the political level. This debate is also active in Europe and in France, as illustrated by the open consultation on network neutrality launched in 2010. For instance, the French regulation authority, ARCEP, has published in its response a proposal intending to define how net neutrality could be implemented [6], [7].

There has been an increasing attention in the literature on providing a mathematical analysis of the advantages and drawbacks of network neutrality. The idea is to investigate the output of the interactions between selfish actors that are end users, CPs and ISPs, using the framework of non-cooperative game theory [8], [9]. Let us briefly describe here, non exhaustively, some important existing works in that direction. In [10], [11], the authors propose to share the revenue among providers using the Shapley value, the only mechanism that satisfies a set of axioms representing a sense of fairness; in this case CPs participate to the network access cost. The work in [12] analyzes how neutrality or non-neutrality affects provider investment incentives, network quality, and user prices. A similar comparison is made in [13] between a two-sided pricing scheme where ISPs are allowed to charge CPs, and one-sided pricing where such side-payments are not allowed. In each case, at the equilibrium of the game, the levels of investment in content and architecture are determined. The paper gives conditions on the ratio between parameters characterizing

Pierre Coucheney and Bruno Tuffin are with INRIA Rennes Bretagne-Atlantique, Campus de Beaulieu, 35042 RENNES Cedex - France

Patrick Maillé is with Télécom Bretagne, 2, rue de la Chataignerai, 35576 Cesson Sévigné Cedex - France

advertising rates and end-user price sensitivity, under which a non-neutral network outperforms a neutral one in terms of social welfare. On the other hand, [14] investigates the case where ISPs negotiate joint investment contracts with a CP in order to enhance the quality of service and increase industry profits. It is found that an unregulated regime leads to higher quality investments, but that ISPs have an incentive to degrade content quality. The paper [15] studies the implications of non-neutral behaviors, taking into account advertising revenues and considering both cooperative and non-cooperative scenarios. In [16], we analyze and compare thanks to game-theoretic tools three different situations of interactions *between* ISPs: the case of peering between the ISPs, the case where ISPs do not share their traffic (exclusivity arrangements), and the case where they fix a transfer price per unit of volume. The paper supports the transit price scenario and suggests a limited regulation (enforcing global connectivity) to prevent incumbent ISPs from having a dominant position in the bargaining. Finally, in [17], a game-theoretic model is considered with a single CP, a single ISP, a (consumers') demand function that depends on price and quality of service, and involving advertisement and network investment components.

In those works, there is in general a single ISP, and one or several CPs. Though, in practice, we often have ISPs in competition for customers, while for many services, the CPs are in a quasi monopoly, a characteristic ISPs complain about. (Typical examples are YouTube for non-copyrighted videos, and Netflix for movies and TV shows in the US.) We propose to specifically address this issue in this paper. Remark that in addition to [13], considering competitive ISPs has been proposed in [18], but with competition over consumers, quality and prices for heterogeneous CPs: none of those works consider a monopolistic CP as can be encountered for some applications. We already addressed this type of problem in [19], whose model was inspired by [15], users were assumed to always go with the cheapest provider. As a consequence, we ended up with a price war (a classical Bertrand competition) such that ISPs always decrease their subscription price in order to attract all demand.

We consider here a more realistic user association model such that users make their choice still based on the price of ISPs, but also on other unknown considerations, hence the use of a classical discrete choice model as in [16]. This requires a derivation of results totally different from [19]. To analyze that situation, we propose a multi-level game where decisions are taken at different time scales. The solutions of the games at the largest time scales, played first, are determined using *backward induction*, meaning that players *anticipate* the impact on, and the resulting solution of, the games played later on at smaller time scales.

The paper is organized as follows. Section II presents the basic assumptions of the model we are going to consider, the different levels of game, and the mathematical description of the investigated comparison between the neutral and the non-neutral regimes. It also describes how users select their ISP (if any), and how the aggregated demand at the CP is determined. The next sections present the various game levels for providers' decisions: we describe in Section III how the

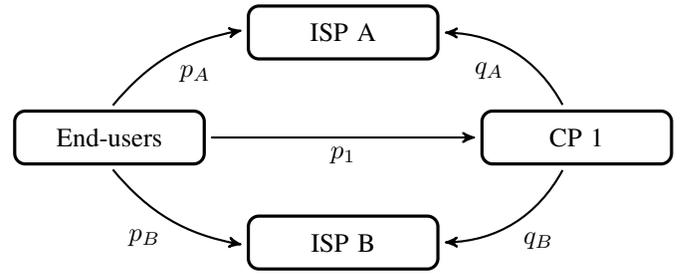


Fig. 1. Charging interactions between stakeholders. Prices p_1 , p_A and p_B are positive flat rates, whereas q_A and q_B are positive per volume unit prices.

CP, anticipating the decisions of end users, chooses the content price. At a higher level, by backward induction, ISPs play a game on the access charge for end users; this competition is described in Section IV. We then describe the game at the highest level, on the economic relationships between the ISPs and the CP, by determining the side payments of the CP to the ISPs in Section V. We address the case when those prices are fixed by ISPs, based on a game; we also look at the case when they are decided by the CP, or by a regulator (maximizing the supply chain value for instance). Section VI concludes by discussing the impact and relevance of side payments on the providers' revenues, highlighting that it is not always in the interest of the ISPs (but could be), while an appropriate choice of side payments may increase the CP revenue. We also give in that section directions for future research.

II. MODEL

A. Topology and Pricing Structure

We consider a single CP, whose parameters will be indexed by 1, and two ISPs, named (and indexed by) A and B . The access prices charged to users are *flat rate* subscription fees, denoted by p_1 , p_A and p_B for respectively the CP, ISP A and ISP B . In order to study non-neutrality, we also introduce side payments q_A and q_B representing the per unit of volume prices that the CP has to pay to A and B , respectively. All prices are assumed to be positive. Finally, the set of end users is considered continuous and (without loss of generality) of mass one, so that we will indifferently refer to "mass" and "proportion" of users. The charges imposed by actors to other players are summarized in Figure 1, the arrows indicating the cash flows.

B. User Demand

Users have to pay both the ISP and the CP to access the content. Users first choose their ISP, to which they pay a flat-rate subscription fee, and then subscribe to the CP too if its (flat) fee is not too high. Since users need an Internet access, not only to reach the content of the CP, but also for other purposes (e-mail, web browsing, ...), we decouple the ISP choice from the (individual) decision to subscribe to the CP, hence the independence from p_1 . Among those access subscribers, the proportion who also subscribe to the CP depends on p_1 on the other hand, but not on the subscription price to the ISPs as a first-order model, assuming

as a consequence that users do not relate their consumption of content to the previously paid ISP subscription fee.

Let us focus first on the ISP selection by users. In a previous work [19], we have considered users simply selecting the cheapest ISP (or choosing it randomly if price equality holds). However, this does not take into account the phenomenon of *stickiness* or *loyalty* of the users, highlighted in [20]. In the model considered here, user choices are influenced by the ISP subscription prices, but also by other considerations (reputation or preferences) that can be modeled as an additive noise to the main criteria determining the choice. We consider here a *discrete choice model*, a standard paradigm in economy, transportation, etc., and surprisingly not often applied in telecommunications. It is a common way to model how decision makers choose among a set of alternatives. Mathematically, we assume that a user has a valuation of the form $v_i = \beta \log(1/p_i) + \kappa_i$ for ISP i , and selects the highest-valued option (see [16] for details). The term κ_i is an individual-specific random term, taking into account unknown aspects and assumed to follow a Gumbel distribution as in standard discrete choice models [21]. We additionally assume that there is a fictitious price p_0 , assumed to be strictly positive, and representing the cost of the outside option, i.e., the perceived cost of not having access to the Internet. Thus if the (random) valuation associated with that outside option is larger than the ones associated with each ISP, the user prefers not to join the network. The parameter $\beta > 0$ represents the *user sensitivity* to the subscription prices: values of β close to zero lead to an uniform choice over the three alternatives (connecting to one of the two ISPs, or not having access to the Internet) regardless of the prices set, whereas large values of β make the users choose the least expensive option. The term $\log(1/p_i)$ expresses the dissatisfaction for higher prices, the logarithm being used to represent the fact that the same variation of price is felt smaller at high prices than at low prices: users are sensitive to relative price variations rather than absolute ones. Finally, the case $p_i = 0$ leads to the maximal possible valuation independently of the random terms, meaning that ISP i attracts all users, or (say, from symmetry) half of them if the other ISP chooses a null price as well.

At a macroscopic level, by discrete choice analysis, the proportion (or equivalently, the mass) σ_i of users selecting ISP i (with also $j \in \{A, B\}$; $j \neq i$), can be shown to equal from our expression of v_i (see [21] for a general derivation)

$$\sigma_i = \begin{cases} \frac{p_i^{-\beta}}{p_A^{-\beta} + p_B^{-\beta} + p_0^{-\beta}} & \text{if } p_A > 0 \text{ and } p_B > 0 \\ 1 & \text{if } p_i = 0 \text{ and } p_j > 0 \\ 1/2 & \text{if } p_A = 0 \text{ and } p_B = 0 \\ 0 & \text{if } p_i > 0 \text{ and } p_j = 0. \end{cases} \quad (1)$$

This repartition function expresses the fact that all users select an ISP if at least one of the subscription prices is null ($p_A = 0$ or $p_B = 0$). Of course, the higher the price, the fewer subscribers; this effect increasing with the user sensitivity to price β .

In this paper, we propose a new aggregated user demand in terms of data volume on ISP i , where users having selected an

ISP then decide whether to use the content offered by the CP, depending on the *flat-rate* price p_1 . We consider that users' willingness-to-pay for the CP service (i.e., the access to all the content offered by the CP) follows an exponential distribution with mean value $1/\alpha > 0$ over the population, independently of the ISP choice. Therefore, a proportion $e^{-\alpha p_1}$ of each ISP's subscribers also subscribes to the CP, hence a number $\sigma_i e^{-\alpha p_1}$ subscribing to both the CP and ISP i . Note that the consequence of our assumption is an exponentially decreasing demand function (in terms of the CP price), a usual setting in economic theory.

We assume that there is an average volume D_0 of data that a user downloads from the CP if subscribing to it. The value will be helpful to compute the *volume-based* transit costs for the CP to the ISPs. This volume is assumed to be independent of the ISP choice. Instead of considering an average volume, we could say that it is the same value for all users without changing the expressions; an average value just allows to take into account the potential variations between users that are averaged when summing over all subscribers of an ISP. D_0 is additionally assumed independent of p_1 , meaning that users just get what they "need" if subscribing. As a result, the data volume for users subscribing simultaneously to the CP and to ISP i is given by:

$$D_i = D_0 \sigma_i e^{-\alpha p_1}. \quad (2)$$

The parameter $\alpha > 0$ can be interpreted as the sensitivity of users to the CP price: the global demand (sum of demands on all ISPs) is a decreasing function of α .

Notice that demand does not directly depend on the side payments q_A and q_B . But the introduction of side payments will induce a reaction on the prices p_A , p_B and p_1 set by ISPs and the CP at equilibrium, which, in turn, indirectly affects demand. Finally, the global volume demand for CP data $D_A + D_B$ equals $(\sigma_A + \sigma_B) D_0 e^{-\alpha p_1}$.

C. Utility and revenue functions

Among the proportion $\sigma_A + \sigma_B$ of users having accepted to pay for an access to the network, and then paying a flat-rate price p_1 to the CP, some would have accepted to pay more to benefit from the content of the CP. The surplus of users that would have accepted to pay p is $p - p_1$, while the proportion of users willing to pay more than p is $e^{-\alpha p_1}$.

We can then compute the *user welfare* associated with the existence of the CP, as the sum over all users of the benefit they make accessing the content of the CP. Note that this does not include the benefit that users make by selecting an ISP, which is associated with other (free) on-line services. From our assumption of an exponential distribution of users' willingness-to-pay for the content (yielding a density $\alpha e^{-\alpha x}$), the user welfare due to the CP can be computed as:

$$\begin{aligned} \text{UW}_{\text{CP}} &= (\sigma_A + \sigma_B) \int_{p_1}^{\infty} \alpha e^{-\alpha x} (x - p_1) dx \\ &= (\sigma_A + \sigma_B) \frac{e^{-\alpha p_1}}{\alpha} \\ &= \frac{D_A + D_B}{\alpha D_0}. \end{aligned} \quad (3)$$

The utilities (revenues) of the ISPs come from the end users subscription fee, and from the CP through the possible side payment. The first one depends on the mass of users with the ISP, and the second one on the total amount of volume downloaded by users. Hence, for ISP i ($i \in \{A, B\}$), the revenue is

$$U_i = p_i \sigma_i + q_i D_i. \quad (4)$$

We can remark here that the revenue is always positive since we do not consider the cost of the network.

The utility of the CP in this model is the sum of revenues gained by users subscribing through A and through B . Those gains come from the flat-rate subscriptions by users through each ISP i , $p_i \sigma_i e^{-\alpha p_i}$, but the volume-based side payments $q_i D_i$ paid to i also have to be taken into account. The CP net benefit (utility) is thus given by

$$\begin{aligned} U_1 &= (p_1 \sigma_A e^{-\alpha p_1} - q_A D_A) + (p_1 \sigma_B e^{-\alpha p_1} - q_B D_B) \\ &= (p_1/D_0 - q_A) D_A + (p_1/D_0 - q_B) D_B. \end{aligned} \quad (5)$$

Since p_1 is decided after q_A and q_B , the CP can also ensure a positive revenue by setting $p_1/D_0 \geq \max(q_A, q_B)$.

D. Multi-stage Decision Problem

The decision variables are the prices p_1, p_A, p_B, q_A, q_B , impacting end users (demand), as well as revenues of providers. Those variables are decided at different time scales or levels, that can be described as follows.

- 1) At the largest time scale, the side payments q_A and q_B are decided. In the neutral case, they are either fixed to 0, or determined as a common value. They can be different in the non-neutral case, and can be determined either by the ISPs (in a game), the CP, or a regulator. All those options will be investigated. Those determinations will be obtained anticipating the solution of the games below whatever the values of q_A and q_B (the so-called *backward induction*).
- 2) At a smaller time scale, for fixed values of q_A and q_B , the ISPs fix their prices p_A and p_B during a non-cooperative game to attract customers and maximize their revenues. Here again, the decisions are made anticipating the solutions at lower levels.
- 3) At an even smaller time scale, the CP sets the price p_1 .

Finally, for those fixed values of p_1, p_A, p_B, q_A, q_B , users choose their ISP (if not too expensive), and decide whether to use the service offered by the CP, as described by formulas (1) and (2).

All those interacting levels are now solved by backward induction, from the smallest time scale to the largest one.

It is possible to perform the same analysis with a different hierarchy in the three levels of game, assuming for example that subscription prices are decided before the side payments. Though, we think it is reasonable to consider that the decisions on side-payments to take place on a large time scale, given the difficulty, span, and costs of such decision processes. Similarly, regarding the time scale difference between prices set by the CP and by the ISPs, our approach is consistent with the (commonly spread) vision that CPs can adapt faster than ISPs,

in part because of the different contract durations binding users to providers (larger durations with ISPs than with CPs), the various difficulties to switch ISPs, etc.; see also [22].

III. CONTENT PROVIDER PRICE DETERMINATION

The CP aims at maximizing his revenue U_1 , for fixed values of p_A, p_B, q_A, q_B , making use of what the total user demand $D_A + D_B$, with D_i given by (2), will be. For convenience, we define $P_i := p_i^\beta$.

Proposition 1. *Given the side payments q_A and q_B and the prices p_A and p_B decided by the ISPs, the price of the CP maximizing its revenue (5) is*

$$p_1^* = \begin{cases} \frac{P_A}{P_A + P_B} (D_0 q_B + \frac{1}{\alpha}) + \frac{P_B}{P_A + P_B} (D_0 q_A + \frac{1}{\alpha}) & \text{if } p_A > 0 \text{ or } p_B > 0, \\ D_0 \frac{q_A + q_B}{2} + \frac{1}{\alpha} & \text{if } p_A = 0 \text{ and } p_B = 0 \end{cases} \quad (6)$$

Proof: We first consider the case $p_A > 0$ and $p_B > 0$. The derivative $\frac{\partial U_1}{\partial p_1}$ of the CP revenue is then equal to

$$P_0 e^{-\alpha p_1} \frac{P_A (\alpha q_B + (1 - \alpha p_1)/D_0) + P_B (\alpha q_A + (1 - \alpha p_1)/D_0)}{P_0 P_A + P_0 P_B + P_A P_B}$$

which is strictly positive until p_1 achieves the value given in the first equation of (6), and strictly negative after. Hence the result.

If $p_A = 0$ (and then $P_A = 0$) and $p_B > 0$ (the opposite case is symmetric and then omitted), the CP revenue is $e^{-\alpha p_1} (p_1 - D_0 q_A)$, whose derivative is $e^{-\alpha p_1} (1 - \alpha (p_1 - D_0 q_A))$.

Finally, if $p_A = p_B = 0$, then the CP revenue is $\frac{1}{2} e^{-\alpha p_1} (2p_1 - D_0 (q_A + q_B))$, and its derivative is $e^{-\alpha p_1} (1 - \alpha p_1 + \frac{\alpha}{2} D_0 (q_A + q_B))$. ■

Notice that the optimal price does not depend on the outside option valuation p_0 . One can also check that it increases with the price p_i of ISP i that has the biggest side payment q_i , and decreases with the other price. In the limit (that can be interpreted as *neutral*) case $q_A = q_B = q$, the optimal pricing for the CP is $q D_0 + 1/\alpha$ whatever the value of p_A and p_B . In that case, as an important remark, the CP's revenue is $\frac{D_A + D_B}{D_0 \alpha}$, which corresponds to the CP-related user welfare: the interest of users and that of the CP coincide here. Finally, we can remark that the optimal price for the CP is always greater than $D_0 \min(q_A, q_B) + 1/\alpha$, because it is a convex combination of $D_0 q_A + 1/\alpha$ and $D_0 q_B + 1/\alpha$. In particular, it is greater than the inverse of the user price sensitivity α .

IV. PRICING GAME BETWEEN ISPs

Before the users decide which ISP to join and the CP chooses p_1 , the ISPs play a pricing game, making use of what the CP and users decisions would be. In this section, we determine the Nash equilibrium solutions of this pricing game in an analytical way when there are no side payments, and numerically (because intractable) in the general case. Recall (see [9]) that a Nash equilibrium would be a price profile (p_A, p_B) such that no ISP can improve (strictly) his utility by unilaterally changing his price. The best-response curves are

defined as (by expliciting the dependence of U_A and U_B on p_A, p_B)

$$\begin{aligned} \text{BR}_A(p_B) &= \arg \max_{p_A \geq 0} U_A(p_A, p_B) \text{ and} \\ \text{BR}_B(p_A) &= \arg \max_{p_B \geq 0} U_B(p_A, p_B). \end{aligned}$$

With those notations, a Nash equilibrium is a point $(p_A^{\text{NE}}, p_B^{\text{NE}})$ for which $\text{BR}_A(p_B^{\text{NE}}) = p_A^{\text{NE}}$ and $\text{BR}_B(p_A^{\text{NE}}) = p_B^{\text{NE}}$. Graphically, if we draw the two best-response curves on the same figure, the set of Nash equilibria is then the (possibly empty) set of intersection points of those curves.

A. No side payments

In the case where no side payments are established, $q_A = q_B = 0$, we get a simple formulation for the revenue of ISPs. From the previous section, the optimal CP pricing is $1/\alpha$. Using the notation $P_i := p_i^\beta$, the revenue of ISP A is then (the revenue of ISP B being symmetrical)

$$U_A = \begin{cases} \frac{P_0 P_B p_A}{P_0 P_A + P_0 P_B + P_A P_B} & \text{if } p_A > 0 \text{ and } p_B > 0 \\ 0 & \text{if } p_A = 0 \text{ or } p_B = 0 \end{cases} \quad (7)$$

We first stress that $p_A = p_B = 0$ is a Nash equilibrium since no player can strictly increase his revenue by unilaterally changing his action: the revenue always remains equal to zero. But setting one's price to zero is a dominated strategy, that is strictly dominated as soon as the adversary price is not zero: it always yields no revenue whereas a strictly positive revenue can be guaranteed with any other choice. Therefore it is not likely to be chosen by ISPs if another equilibrium exists.

Proposition 2. *Assuming that there are no side payments, i.e. $q_A = q_B = 0$, then*

- if $\beta \leq 1$, there is no Nash equilibrium different from $(0, 0)$ with finite prices: the only other alternative is both ISPs setting infinitely large prices ($P_A^{\text{NE}} = P_B^{\text{NE}} = \infty$),
- if $1 < \beta < 2$, there is a unique Nash equilibrium different from $(0, 0)$ with $P_A^{\text{NE}} = P_B^{\text{NE}} = \frac{2 - \beta}{\beta - 1} P_0$,
- if $\beta \geq 2$, $(0, 0)$ is the unique Nash equilibrium, yielding no revenue for the ISPs.

The proof relies on the following general result about symmetric games.

Lemma 1. *If the best response functions BR_A and BR_B are*

- equal: $\text{BR}_A = \text{BR}_B = \text{BR}$,
- single-valued,
- strictly increasing,

then (p_A, p_B) is a Nash equilibrium if and only if $p_A = p_B = p$ with p a fixed point of the best-response function: $p = \text{BR}(p)$.

Proof: (Lemma) The pair of prices (p_A, p_B) is a Nash equilibrium if and only if $p_A = \text{BR}(p_B)$ and $p_B = \text{BR}(p_A)$. Let us suppose that $p_A \neq p_B$, for instance $p_A > p_B$ (should the indexes be permuted). Then:

$$p_B = \text{BR}(p_A) > \text{BR}(p_B) = p_A,$$

where the inequality comes from the strict increasingness of BR , hence a contradiction. At Nash equilibrium, $p_A = p_B$ is then a necessary condition, and from the definition of such an equilibrium, it is necessary and sufficient to have $p = \text{BR}(p)$. ■

Proof: (Proposition) Assuming that $p_A > 0$ and $p_B > 0$, the derivative of ISP A revenue (7) is

$$\frac{P_B P_0}{(P_A P_B + P_B P_0 + P_A P_0)^2} (P_A(1 - \beta)(P_B + P_0) + P_B P_0).$$

Hence the derivative has the same sign as the affine function of P_A : $P_A(1 - \beta)(P_B + P_0) + P_B P_0$. When $\beta \leq 1$, that derivative is always strictly positive for $P_A \geq 0$, thus (by symmetry) each ISP should set infinitely large prices.

When $\beta > 1$ the derivative of ISP A revenue is strictly positive while P_A is smaller than the unique root of the affine function above, and negative afterwards. Given $p_B > 0$, the best-response of ISP A is then

$$\text{BR}_A(P_B) = \frac{P_B P_0}{(\beta - 1)(P_B + P_0)}.$$

For the case $\beta > 1$, notice that the best response is the same function for ISP B due to symmetry, that equals $\text{BR}(P) = \frac{1}{(\beta - 1)(1/P_0 + 1/P)}$, and is a strictly increasing function of P . Hence it follows from the previous lemma that every Nash equilibrium is symmetric, which results here in the necessary and sufficient condition at Nash equilibrium $P_B^{\text{NE}} = P_A^{\text{NE}} = \text{BR}_A(P_A^{\text{NE}})$. The last equation has a unique strictly positive solution in the case $1 < \beta < 2$, the one given in the proposition, and no solution otherwise. ■

This proposition shows in particular that, when the price sensitivity of users is high ($\beta \geq 2$), we are led to the same price war as in the model of Bertrand competition studied in [19]. But for smaller levels of price sensitivity, this does no longer happen: the price set by ISPs at equilibrium is strictly positive, hence providing some revenue from users to both ISPs.

B. Positive side payments

In the general case, the computation of the Nash equilibrium or even the best response function is not analytically tractable. We are then led to study numerically the price competition between ISPs. From here, we take $\alpha = 1$, $p_0 = 1$, $D_0 = 1$, and $\beta = 1.5$. We remark that the choices of p_0 and D_0 are made without loss of generality (they correspond to unit changes of ISP prices and data volumes). The choice of α also corresponds to a unit change in the CP prices, and is therefore linked to that of p_0 (i.e. for a given user sensitivity to the CP price, the value of α is a consequence of the monetary unit choice, and thus is determined by the choice for p_0). Therefore our choice of α and of β are subject to discussion. We however stress here that we have also run experiments with other values (in the range $(1, 2)$ for parameter β) and observed similar results.

Let us first remark that if the price set by an ISP (say ISP A) is equal to zero, then the other ISP (ISP B), at a Nash equilibrium, sets his price to zero as well. This is because ISP B does not attract any users if his price is not

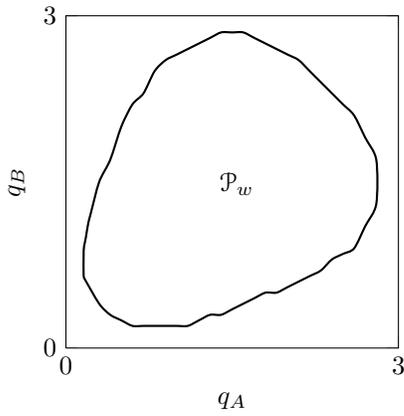


Fig. 2. Set \mathcal{P}_w of side payments (q_A, q_B) for which the Nash equilibrium is such that $p_A^{\text{NE}} = p_B^{\text{NE}} = 0$, i.e. price war holds.

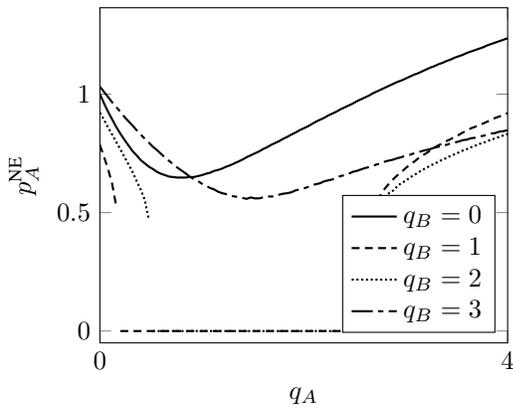


Fig. 3. ISP A user price p_A^{NE} at equilibrium as a function of the side payments q_A with $q_B \in \{0, 1, 2, 3\}$.

zero, and then his revenue is null, whereas the revenue is $\frac{1}{2}q_B D_0 e^{-\frac{\alpha}{2} D_0 (q_A + q_B) - 1}$ otherwise, hence strictly positive.

Numerical computations show that there is a set of side payments q_A and q_B for which the price war phenomenon between ISP happens. That set is shown in Figure 2. In the following, we will denote by \mathcal{P}_w the set of side payments for which a price war between ISPs takes place, leading to null subscription prices.

Figures 3 to 6 show the prices and revenues at Nash equilibrium (recall that this is for q_A and q_B fixed). Numerical computations point out the fact that the revenue of the CP, and the user welfare he creates, are always equal at equilibrium, the reason why we do not plot user welfare here. While this equality is clear when side payments are the same, our numerical results suggest that it remains true in the general case.

Figure 3 represents the price p_A^{NE} set by ISP A at Nash equilibrium, when the side payment q_A varies, and for several ISP B side payments q_B . This reveals that the price at equilibrium first decreases with the side payment set by the ISP, and then increases. For some value of the opponent ISP side payment, it goes to zero when a threshold is reached. That threshold corresponds to the case where the side payment revenue that ISPs get by setting their prices to zero, and then

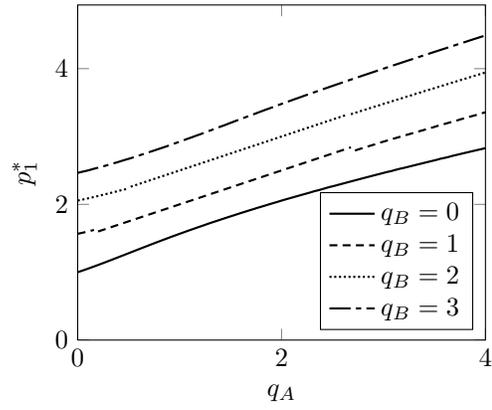


Fig. 4. CP optimal price at equilibrium as a function of the side payments q_A with $q_B \in \{0, 1, 2, 3\}$.

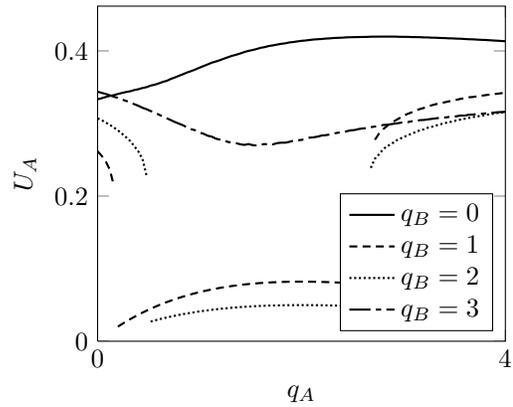


Fig. 5. ISP A revenue at equilibrium as a function of the side payment q_A with $q_B \in \{0, 1, 2, 3\}$.

attracting the whole set of users, becomes larger than the one they get on a limited market share with both the subscription fees and the side payments. This is the price war situation: equilibrium subscription prices of both ISPs fall down to 0 as the side payments enter the “price war zone” \mathcal{P}_w highlighted in Figure 2, and each ISP finally only attracts half of the user set, hence the discontinuity. Finally, we observe that there is no monotonicity in the opponent side payment.

Figure 4 displays the optimal price p_1^* of the CP in terms of q_A for different values of q_B . We can notice the discontinuity due to the price war thresholds (for the cases $q_B = 1$ and $q_B = 2$, since there are no such thresholds for the two other cases from Figure 2). However, this discontinuity is barely visible, which is quite surprising when compared to that observed at the ISP level. It can also be remarked that the optimal price increases with q_A (and q_B) both before and after the thresholds, but not in general. One can also check here the general property that $p_1^* \geq D_0 \min(q_A, q_B) + 1/\alpha$, which, in particular, ensures a strictly positive revenue to the CP.

Figure 5 shows that the revenue of ISPs is not monotonic with the side payment. Again, some discontinuities occur when side payments enter and leave the price war zone \mathcal{P}_w . Moreover, depending on the ISP B side payment, the maximal

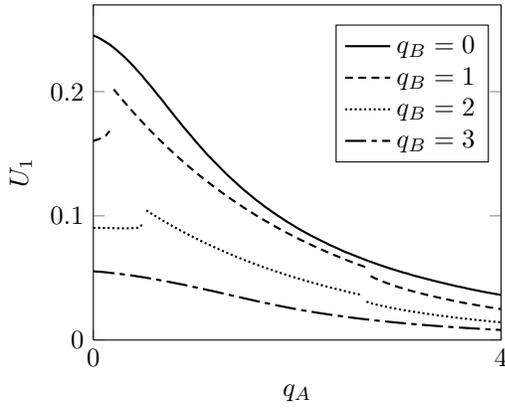


Fig. 6. CP revenue at equilibrium as a function of the side payment q_A , with $q_B \in \{0, 1, 2, 3\}$.

revenue of ISP A is reached either for a null side payment (e.g., when $q_B = 3.0$) or for a strictly positive value (e.g., when $q_B = 0.0$). This illustrates that predicting the effect of the various parameters in the output of the game(s) is a difficult task and game-theoretic tools are helpful here to get a result: a side payment increase can indeed lead to a content price increase by the CP to compensate for the loss, and a reaction of the other ISP; this also has an effect on demand and depending on the variations, this may or may not result in an ISP revenue increase.

On the other hand, the CP revenue, plotted in Figure 6, has a tendency to decrease with side payments, even if it is not strictly the case, as can be seen for $q_B = 1$ and small values of q_A . When a discontinuity occurs, the CP revenue is maximized for the smallest side payment leading to the price war. The non-monotonicity is due to the threshold for the price war between ISPs, inducing a discontinuity in the ISP prices. Because of those prices becoming null, the revenue jumps up.

Finally, over the price war set \mathcal{P}_w , the CP subscription price is $p_1 = D_0 \frac{q_A + q_B}{2} + \frac{1}{\alpha}$ and the revenues are

$$U_i = \frac{1}{2} q_i D_0 e^{-\frac{\alpha D_0}{2}(q_A + q_B) - 1} \quad (8)$$

$$U_1 = UW_{CP} = \frac{1}{\alpha} e^{-\frac{\alpha D_0}{2}(q_A + q_B) - 1}. \quad (9)$$

V. SIDE PAYMENTS DETERMINATION

We consider at the highest level three possibilities for the choice of the side payments q_A and q_B . We first look at the case when they are determined by the CP (even if unlikely in practice), then the case when they result from a game played between ISPs, and finally the case when they are determined by a regulator (e.g., to maximize social welfare).

Since we don't have the analytical expression for the ISPs price at Nash equilibrium, we provide numerical results, where we take $\alpha = 1$, $\beta = 1.5$, $p_0 = 1$ and $D_0 = 1$.

A. Determined by the CP

The revenue of the CP is maximized when the side payments are $q_A = q_B = 0.3$ as illustrated in Figure 7 (instead of plotting a hard-to-read 3D-curve of CP revenue in terms of q_A

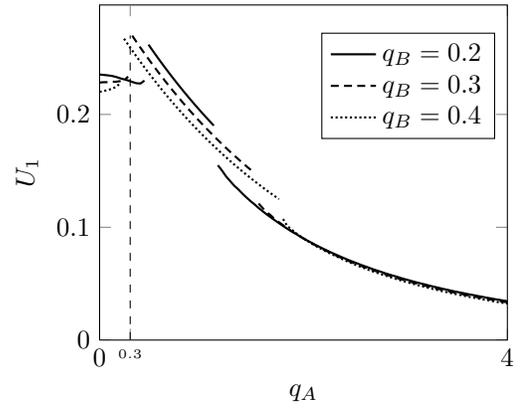


Fig. 7. CP revenue at equilibrium as a function of the side payment q_A , with $q_B \in \{0.2, 0.3, 0.4\}$. The maximal revenue is reached when $q_A = q_B = 0.3$.

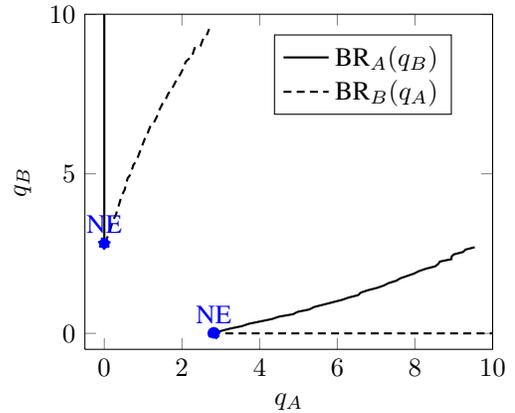


Fig. 8. The optimal side payment of each ISP as a function of the opponent ISP side payment. There is a threshold $q_i \approx 2.80$ beyond which the best response falls to zero. There are two Nash equilibria (NE) where one ISP sets its side payment to zero.

and q_B , we have preferred to draw 2D-curves in terms of one of the parameters for various values of the second parameter close to optimal). It is interesting to notice that, for such a set of side payments, there is a price war on the user prices, i.e., $p_A^{NE} = p_B^{NE} = 0$ here. In fact it corresponds to the symmetric ($q_A = q_B$) point of the price war set \mathcal{P}_w described in Figure 2 for which the sum of side payments is minimized. Indeed, if $p_A^{NE} = p_B^{NE} = 0$, then the revenue of the CP can be rewritten as $U_1 = \frac{1}{\alpha} e^{-\frac{\alpha D_0}{2}(q_A + q_B) - 1}$, and then is maximized when $q_A + q_B$ is minimal.

At this point the revenues of the stakeholders are $U_A = U_B \approx 0.04$, and $UW_{CP} = U_1 \approx 0.27$. Hence the revenues of ISPs are much smaller than the one of the CP. Note that the situation is quite counter-intuitive, since the CP gains to introduce side payments. This is because those payments exacerbate the competition between ISPs, which is beneficial to end users, and finally to the CP who can reach more customers.

B. Determined by the ISPs, through a game

If we instead assume that the side payments are non-cooperatively determined by the ISPs, we are led to study

the best response of each ISP to the other ISP side payment. As shown in Figure 8, the best response is first increasing in the other ISP side payments, and then falls to zero above a threshold, which is approximately 2.80. Since the best-response to a null price is 2.80, it follows that $(0, 2.80)$ and the symmetric point $(2.80, 0)$ are Nash equilibria, and they are the only ones. It is interesting to notice that the resulting side payments are not symmetric at Nash equilibrium, so are the revenues equalling 0.42 for the ISP with side payment 2.80 and 0.34 for the other ISP.

Now let us compare that outcome to the case without any side payments. From Subsection IV-A, with $\beta = 1.5$, the revenue of ISPs is 0.33. Hence the ISPs global revenue increases by about 15%, which goes in the direction of ISPs arguments about side payments improving their revenue. On the other side, the CP revenue decreases from 0.25 to 0.06, hence losing nearly 75% of its value. The benefit of ISPs is then at the expense of the CP and consequently of the user welfare.

C. Determined by a regulator

A regulator can either decide to maximize the revenue of the supply chain (sum of utilities of the ISPs plus the CP), the user welfare (end-users surplus), or the social welfare (including user welfare and all providers utilities).

The total value of the *supply chain* is the total revenue got from the users, i.e., $U_1 + U_A + U_B$.

User welfare can be decomposed into two components: the user welfare due to the existence of the CP -that is computed in (3)-, and the user welfare due to the presence of the ISPs. Let us focus on the latter part: we have assumed that users not connected to the Internet perceive a cost p_0 (thus p_0 can be seen as the value of the connectivity service). When a user decides to subscribe to ISP i and pays the corresponding price p_i , its benefit is then $p_0 - p_i$ with respect to the no-ISP situation: the user does not bear anymore the cost p_0 of not having Internet access, and instead perceives the monetary cost p_i . Aggregating over the whole population, the user welfare that is due to the presence of the ISPs (with their prices p_A and p_B) equals

$$UW_{\text{ISPs}} = \sigma_A(p_0 - p_A) + \sigma_B(p_0 - p_B).$$

The global user welfare generated by the system (ISPs and CP) is therefore

$$UW = UW_{\text{CP}} + UW_{\text{ISPs}} \quad (10)$$

Finally, *social welfare* is defined as the overall value of the service for the society. It therefore includes the surpluses of all actors, and equals $SW = U_1 + U_A + U_B + UW$. Note that Social Welfare also corresponds to the total value that the service has for subscribers, without considering any monetary exchanges because they stay within the society. We indeed obtain, simplifying the sum of the terms in SW:

$$SW = (\sigma_A + \sigma_B) \left(p_0 + \left(p_1 + \frac{1}{\alpha} \right) e^{-\alpha p_1} \right),$$

where the term $(\sigma_A + \sigma_B)p_0$ is the value of the connectivity service for ISPs' subscribers, and the other term is the value

of the CP service for CP subscribers, computed as

$$(\sigma_A + \sigma_B) \int_{p_1}^{\infty} \alpha e^{-\alpha x} x dx.$$

1) *Side payments to maximize User Welfare*: Since CP revenue and user welfare are equal at Nash equilibrium, it follows that user welfare is maximized when the CP revenue is maximized. This case has already been treated in Subsection V-A.

2) *Side payments to maximize Social Welfare*: We have obtained numerically that social welfare is maximized for the same side payments as the ones maximizing the CP revenue and the user welfare.

3) *Side payments to maximize the supply chain value*: The supply chain value is maximized when the side payments are both null, which has been studied in Subsection IV-A. In this neutral case the revenue of ISPs is approximately 0.33 whereas that of the CP (and the induced user welfare) is 0.25. We can remark that among the three alternatives considered in this subsection, this one leads to the fairest revenue sharing between stakeholders.

VI. DISCUSSION, CONCLUSIONS AND FUTURE WORK

We have provided in this paper a model describing the interactions between two ISPs in competition, a CP, and end users connecting to the network. With respect to the literature, we believe that considering competitive ISPs and a single CP is a more realistic representation of the current network where we have a quasi-monopoly for some applications (for instance YouTube or Netflix), while several ISPs are in competition (an argument of ISPs). The goal is to study the impact of side payments on providers' revenues, and conclude whether they can help reduce the unfairness of the current revenue sharing among all actors, as claimed by ISPs in the current network neutrality debate. We restricted the analysis to two ISPs because in many cases, competition is limited to a very small number of ISPs, and very often only two ISPs are available to a given user [22]; but a numerical analysis can be extended to more than two.

In this paper, we have presented a three-level game where (from the largest to the shortest time scale) the side payments are first determined, then a pricing game is played between ISPs, followed by the content provider price, and finally, knowing all those prices, end users choose their ISP (or none if too expensive) and possibly decide to subscribe also to the CP service. All those levels are played by backward induction, meaning that players anticipate the solutions of the later games when choosing their strategies.

Our results have highlighted the fact that side payments, *unless decided by ISPs*, have little chance to address the concern from ISPs regarding the fairness of the revenue sharing associated with users accessing content through their infrastructures. This is due, to a great extent, to the competition played among ISPs on the access prices that drives their revenues to low values. On the other hand, the CP being in a monopolist situation, always obtains significant revenues. An interesting paradox we have highlighted is that side payments may be beneficial to the CP. Nevertheless, when side payments are decided by the ISPs (non-cooperatively), it can

be beneficial to them, but at the expense of both the CP and the users. Remarkably, the side payments maximizing social and user welfare are the same than those maximizing the CP revenue. But looking at the whole supply chain, in order to avoid too big disparities between revenues of providers, the neutral case is the most suitable. If the side payments are decided non-cooperatively by ISPs, in our experiment, one (only) is a big winner, while the other ISP gains a bit more than in the neutral case. This asymmetry may be a problem and can create complicated tensions and negotiations.

As future research, we would like to go into several directions: first to include several CPs with different contents, but such that some end users are targeting only a subset of them, for all possible subsets. ISPs may also charge each other to let the CPs not connected to them reach their own customers (transit pricing). Other extensions to our model could include architecture investment and content innovation characteristics, for the ISPs and the CP respectively.

ACKNOWLEDGMENTS

This work has been supported by INRIA's ARC project MENEUR and EuroNF NoE's project INNIS on network neutrality, and by the French ANR VERSO project CAPTURES on competition between providers.

REFERENCES

- [1] Sandvine, "Global Internet phenomena report: Fall 2011," available at www.sandvine.com, 2011.
- [2] T. Lenard and R. E. May, *Net Neutrality or Net Neutering: Should Broadband Internet Services be Regulated*. Springer, 2006.
- [3] A. Odlyzko, "Network neutrality, search neutrality, and the never-ending conflict between efficiency and fairness in markets," *Review of Network Economics*, vol. 8, no. 1, pp. 40–60, 2009.
- [4] A. Ghosh and A. Sayedi, "Expressive auctions for externalities in online advertising," in *Proceedings of the 19th World Wide Web conference (WWW)*, Raleigh, NC, USA, April 2010.
- [5] D. Liu, J. Chen, and A. Whinston, "Competing keyword auctions," in *Proceedings of 4th Workshop on Ad Auctions*, Chicago, IL, USA, July 2008.
- [6] ARCEP, "Neutralité de l'internet et des réseaux : propositions et orientations," 2010, http://www.arcep.fr/uploads/tx_gspublication/net-neutralite-orientations-sept2010.pdf.
- [7] S. Wong, R.-M. J., and E. Altman, "Public consultations on Net Neutrality 2010: USA, EU and France," SSRN, Tech. Rep., 2010, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1684086.
- [8] D. Fudenberg and J. Tirole, *Game Theory*. MIT Press, Cambridge, Massachusetts, 1991.
- [9] M. Osborne and A. Rubinstein, *A Course in Game theory*. MIT Press, 1994.
- [10] R. Ma, D.-M. Chiu, J. Lui, V. Misra, and D. Rubenstein, "Interconnecting eyeballs to content: A Shapley value perspective on ISP peering and settlement," in *Proc. of International Workshop on Economics of Networked Systems (NetEcon)*, 2008, pp. 61–66.
- [11] —, "On cooperative settlement between content, transit and eyeball internet service providers," in *Proc. of ACM International Conference on Emerging Networking EXperiments and Technologies (CoNEXT)*, 2008.
- [12] J. Musacchio, J. Walrand, and G. Schwartz, "Network neutrality and provider investment incentives," in *Proc. of 41st Asilomar Conference on Signals, Systems and Computers (ACSSC)*, Nov 2007, pp. 1437 – 1444.
- [13] J. Musacchio, G. Schwartz, and J. Walrand, "A two-sided market analysis of provider investment incentives with an application to the net-neutrality issue," *Review of Network Economics*, vol. 8, no. 1, 2009.
- [14] C. Saavedra, "Bargaining power and the net neutrality debate," 2009, sites.google.com/site/clauidiasaavedra/attachments/bargaining_power.pdf.

- [15] E. Altman, P. Bernhard, S. Caron, G. Kesidis, J. Rojas-Mora, and S. Wong, "A study of non-neutral networks with usage-based prices," in *Proc. of 3rd ETM Workshop*, Amsterdam, The Netherlands, Sept 2010.
- [16] P. Coucheney, P. Maillé, and B. Tuffin, "Network Neutrality Debate and ISP Inter-Relations: Traffic Exchange, Revenue Sharing, and Disconnection Threat," Inria, Rapport de recherche, Jan. 2012. [Online]. Available: <http://hal.inria.fr/hal-00659278>
- [17] E. Altman, A. Legout, and Y. Xu, "Network non-neutrality debate: An economic analysis," in *Proceedings of Networking*. Barcelona, Spain: LNCS 6641, Springer Verlag, May 2011, pp. 68–81.
- [18] P. Njoroge, A. Ozdaglar, N. Stier-Moses, and G. Weintraub, "Investment in two sided markets and the net neutrality debate," Columbia University, Decision, Risk and Operations Working Papers Series, Tech. Rep. DRO-2010-05, 2010.
- [19] F. Boussion, P. Maillé, and B. Tuffin, "Net neutrality debate: Impact of competition among ISPs," in *Proceedings of the Fourth International Conference on COMMunication Systems and NETWORKS (COMSNETS)*, Bangalore, India, 2012.
- [20] S. Caron, G. Kesidis, and E. Altman, "Application neutrality and a paradox of side payments," in *Proceedings of the Re-Architecting the Internet Workshop*. ACM, 2010.
- [21] M. Ben-Akiva and S. Lerman, *Discrete choice analysis*. The MIT Press, Cambridge, Massachusetts, 1985.
- [22] N. Economides, "The economics of the internet backbone," in *Handbook of Telecommunications*, S. Majumdar et al., Ed. Amsterdam: Elsevier Publishers, 2005, vol. 2, pp. 373–412.



Pierre Coucheney received his Master degree in applied mathematics from universit  Paris 6, and his Ph.D. in computer science from university of Grenoble. He is currently in a post-doctoral position at Inria-Rennes, working on the network neutrality topic. His research interests include game theory and its applications to the optimization and economic modelling of networks.



Patrick Maill  graduated from Ecole polytechnique and Telecom ParisTech, France, in 2000 and 2002, respectively. He has been an assistant professor at the Networks, Security, Multimedia department of Telecom Bretagne since 2002, where he obtained his PhD in applied mathematics in 2005, followed by a 6-month visit to Columbia University in 2006. His research interests are on game theory and economic concepts applied to telecommunication ecosystems: resource pricing, routing, consequences of user selfishness on network performance.



Bruno Tuffin received his PhD degree in applied mathematics from the University of Rennes 1 (France) in 1997. Since then, he has been with INRIA in Rennes. He spent eight months as a postdoc at Duke University in 1999. His research interests include developing Monte Carlo and quasi-Monte Carlo simulation techniques for the performance evaluation of telecommunication systems, and developing new Internet-pricing schemes and telecommunication-related economic models. He has published more than one hundred papers on those issues. He has also led or participated into several French and European projects, and co-organized several conferences. He is currently Associate Editor for *INFORMS Journal on Computing*, *ACM Transactions on Modeling and Computer Simulation* and *Mathematical Methods of Operations Research*. He has written or co-written two books devoted to simulation: "Rare event simulation using Monte Carlo methods" published by John Wiley & Sons in 2009, and "La simulation de Monte Carlo" (in French), published by Hermes Editions in 2010.

Submitted to *Operations Research*
manuscript (Please, provide the manuscript number!)

Authors are encouraged to submit new papers to INFORMS journals by means of a style file template, which includes the journal title. However, use of a template does not certify that the paper has been accepted for publication in the named journal. INFORMS journal templates are for the exclusive purpose of submitting to an INFORMS journal and should not be used to distribute the papers in print or online or to submit the papers to another publication.

Revenue-Maximizing Rankings for Online Platforms with Quality-Sensitive Consumers

Pierre L'Ecuyer

Département d'Informatique et de Recherche Opérationnelle, Université de Montréal, C.P. 6128, Succ. Centre-Ville, Montréal, H3C 3J7, Canada; and Inria Rennes – Bretagne Atlantique, lecuyer@iro.umontreal.ca

Patrick Maillé

Telecom Bretagne, 2 rue de la Châtaigneraie CS 17607, 35576 Cesson Sévigné Cedex, France,
patrick.maille@telecom-bretagne.eu

Nicolás Stier-Moses

Universidad Torcuato Di Tella Business School, Saenz Valiente 1010, Buenos Aires, Argentina; and CONICET Argentina,
nstier@utdt.edu

Bruno Tuffin

Inria Rennes – Bretagne Atlantique, Campus Universitaire de Beaulieu, 35042 Rennes Cedex, France, bruno.tuffin@inria.fr

When a keyword-based search query is received by a search engine (SE), a classified ads website, or an online retailer site, the platform has exponentially many choices in how to sort the output to the query. Two extreme rules are (a) to return a ranking based on relevance only, which attracts more requests (customers) in the long run because of perceived quality, and (b) to return a ranking based only on the expected revenue to be generated by immediate conversions, which maximizes short-term revenue. Typically, these two objectives (and the corresponding rankings) differ. A key question then is what middle ground between them should be chosen. We introduce stochastic models and propose effective solution methods to compute a ranking strategy that optimizes long-term revenues. A key feature of our model is that customers are quality-sensitive and are attracted to the platform or driven away depending on the average relevance of the output. The proposed methods are of crucial importance in e-business and encompass: (i) SEs that can reorder their organic output and place their own content in more prominent positions than that provided by third-parties, to attract more traffic to their content and increase their expected earnings as a result; (ii) classified ad websites which can favor paid ads by ranking them higher; and (iii) online retailers which can rank products they sell according to buyers' interests and also the margins these products have. This goes in detriment of just offering rankings based on relevance only and is directly linked to the current *search neutrality debate*.

1. Introduction

The Internet occupies an increasingly important position in our daily lives. Electronic commerce, in particular, has enabled marketplaces in which participants can buy, sell, or rent a huge variety of objects and services in a very convenient way. Because the current Internet is a complex ecosystem of companies, there are various business models that have proved profitable. Among them, we will specially focus on three specific classes: search engines (SE) such as Google, that allow users to find content of their interest on the web, and use these transactions as a chance to sell advertisement; online retailers such as Amazon.com that act as intermediaries between producers and consumers; and classified ad websites such as eBay that allow sellers or service-providers, and buyers or service-consumers, respectively, to meet and conduct transactions. To be profitable, those marketplaces typically rely on one or more of the following revenue streams. In some cases, they charge a commission equal to a percentage of the agreed price-tag (e.g., eBay or Airbnb). Some marketplaces provide a basic service for free but charge sellers to display their classified ads in premium locations or for additional time (e.g., leboncoin.fr in France, or Mercado Libre in Latin America). In addition, they offer additional services such as insurance or delivery for a fee. Finally, another common revenue source comes from third-party advertisers that display text or banners within the pages of the marketplace in exchange for payment.

The common feature in all those platforms is that when a user connects to them and enters a query or category, the site provides a list of relevant items that match what the user wants. To provide value to users, it is crucial to present the relevant items in the platform in the correct order so the user can find the most appropriate ones. Indeed, by presenting certain items first, the site can boost users' interest by increasing relevance. For example, eBay provides relevance-based ranking, among other possibilities such as time until the end of the auction, distance, price, etc. The details of how to assign a relevance value to a query vary depending on the intrinsic details of the platform. For example, eBay may use the distance between the query string and item description as well as the rating of the seller, Amazon may use the number of conversions for a product and its quality, and Google may use PageRanks as inputs (Google 2011). How to define and compute these indices has been the topic of several studies, especially in the case of SEs. Examples include Avrachenkov and Litvak (2004), Austin (2006), Auction Insights (2008), Williams (2010). Since our focus is finding the correct ranking and not how to compute those relevance indices, we assume that they are given as part of the input.

Instead of sorting items by relevance, a platform could take a myopic approach to increase short-term revenue by placing highly-profitable items in prominent positions. The purpose of our work is to study the compromise that can be made by the platform when choosing how to order the items

corresponding to a user query. The tradeoff is between maximizing the expected revenue that could be obtained directly and indirectly from this request, and the long-term impact on the future arrival rate of requests (which impacts future expected revenues). Our aim is to find an optimal ordering in the long run, taking both effects into account. That is, to achieve a balance between both goals when customers are quality-sensitive and their likelihood of visiting the platform is a function of the perceived relevance. Our results provide optimal ranking policies with respect to long-term revenue maximization. Also, we compare the optimal policy to other possible rankings—such as those based on relevance only or those based on short-term revenue only—in terms of expected revenue for the platform, expected revenue for the various content providers, and consumer welfare (captured by the expected quality).

Note that most platforms display paid ads (usually referred to as *sponsored search*), in addition to the regular output (usually referred to as *organic search*). The most common arrangement is that advertisers pay the platform whenever users click on an ad. The payment amount is automatically decided using a bidding process between the SE and all advertisers interested in that keyword. We want to stress that our discussion applies to organic content, since sponsored search is handled using an ad-hoc and well-studied procedure. Our choice relies in that sponsored links and their ordering are much less likely to impact the future arrival rate of requests than the ranking of organic links. We think it is quite reasonable to assume that the user satisfaction (and likelihood to use the platform again) depends mostly on whether the user is pleased with what she finds among the proposed organic links, and not on whether the ads that were displayed are relevant. Actually, a big percentage of users is by now trained to not look at the portion of the screen that displays the ads. Henceforth, our model assumes that the average arrival rate of search requests is influenced by the average relevance of organic links, and not by the sponsored results given in the ads section of the page. In the case of sponsored search results, the ordering of items is typically determined by a generalized second price (GSP) auction. This mechanism is also used to fix the price that advertisers pay to have their ads displayed. The mechanism orders ads from higher to lower expected revenue (notice that the click-through-rate takes into account the relevance of the ad). Precisely, that is the outcome at equilibrium of a GSP auction. For details of these mechanisms, we refer the reader to Varian (2007), Edelman et al. (2007), Lahaie et al. (2007), Athey and Ellison (2011), Maillé et al. (2012) and the references therein. Since it is not our focus, we abstract away the sponsored search mechanism, and represent the total expected ad revenue (per user per visit) by a fixed coefficient β . While there is an extensive literature on sponsored search, the impact of using alternative rankings to classify organic links has not yet received a similar level of attention.

Most platforms employ unpublished and secret algorithms to rank relevant results, so it is anybody's guess what they do exactly. In the last couple of years, some SEs have been under scrutiny by

individuals and organizations that oversee the Internet as well as by regulators in various countries because some believe that the organic search ranking is not only done with respect to an objective measure of relevance but that some revenue-making ingredients also play a role (Crowcroft 2007). For example, it has been said that Google may favor YouTube and other of its own content because of the extra revenue it generates. This has been discussed even by the Federal Trade Commission in the US (Brill 2013) and in a Senate hearing (Rushe 2012). It has also been amply documented that search bias occurs in experiments (Edelman and Lockwood 2011, Wright 2012, Maillé and Tuffin 2014). A search for a video in Google is likely to generate some organic links to YouTube pages, which contain ads that directly benefit Google. Since videos in competitors' platforms do not generate additional revenue, Google has a financial interest in the user to click on YouTube content. Similarly, the expected revenue may increase if a link to a Google map is included in the output instead of a link to MapQuest, Yahoo Maps, etc. There are many other similar situations like this, including weather reports, movies and showtimes, product search, news articles, and so forth. Heterogenous characteristics, including different ownership, can be captured by explicitly associating each link within the organic output to the expected revenue attained when somebody clicks on it, and using those expected revenues as input to find the optimal ordering. The debate about whether SEs should or should not enter into these considerations when ranking links is usually referred to as *search neutrality* and has ignited public interest (Crowcroft 2007, Inria 2012). It relates to other policy debates regarding whether or how to regulate the Internet; the most prominent example being *network neutrality* (see, e.g., Odlyzko (2009) for a discussion about both issues). A neutral SE should only use relevance to construct its rankings, and ignore the revenue parameter mentioned earlier. This would allow new entrants that perform well (i.e., that are commonly clicked) to be listed near the top of the list of organic search results. The risk of a non-neutral ranking is that it may slow down innovation by favoring the incumbents that are known to generate profits, thereby preventing new applications/content from being shown, and hence to become known and successful.

Motivated by the fact that many of these platforms are public companies that strive to maximize returns to stakeholders, and also considering the search neutrality debate, we study the impact of ranking policies on platforms' revenue, as well as on social welfare. The policies may range from the extreme of being neutral (and hence only considering relevance) to being just profit-driven (and hence giving prominence to links that generate the most profit). The main goal of this paper is to develop a modeling framework that permits us to design tractable algorithms for computing optimal ranking policies for the platform, assuming that customers are quality-sensitive and may defect to competing platforms if they do not find what they are looking for. Surprisingly, as far as we know, we are the first to provide an economic analysis of ranking policies, and to show how

to design optimal policies from the perspective of the platform. The tools we develop can prove useful to websites that want to fine-tune ranking policies to achieve long-term profitability. Since SEs play an important role in our connected society by allowing end-users to access content and applications without necessarily knowing of them, such revenue-maximizing strategies are directly linked to the search neutrality debate. Hence, the framework we introduce can also be of high interest to regulators who study the impact that search neutrality has on users and on overall social welfare. This may allow regulators to determine if intervention is warranted, and to study the consequences of doing so. In particular, our study may prove useful to provide arguments for or against non-neutral SEs.

To capture that users are more or less likely to visit a platform depending on the long-term reputation, in our model requests arrive at a rate that depends on the average relevance of displayed links; the more relevant the expected results, the larger the number of visits. The expected revenue is the rate of visits to each page, multiplied by the expected revenue per visit for that page, summed over all possible pages. These quantities depend on the *ranking policy*, defined as a rule that assigns a permutation of matching pages to each possible keyword. A ranking based only on immediate revenue is generally suboptimal; it must also take into account the impact on relevance because that affects the rate of visits since customers are quality-sensitive. For the purpose of this study, we consider that the distributions of relevance and expected revenues for each page are known in advance, so we consider them as inputs. (These distributions can be estimated empirically using data available in the SEs' and marketplaces' servers. Although exploring this data is an interesting direction of research, we leave this to follow-up work.)

Our main contribution is the characterization of the optimal stationary policy for the ranking problem. We propose an algorithm that exploits the characterization of the optimal policy that allows the SE to assign a scalar number to each matching item to then find the optimal ranking by simple sorting. Our model is an infinite-horizon time-stationary sequential decision process that fits the general framework of stochastic dynamic programming (DP), so one could think of using DP methodology to characterize and compute an optimal ranking policy. However, a ranking policy in general is a mapping which to each state of the system, assigns a permutation of all relevant links. Since the state must contain (at least) the current request and the current arrival rate, the number of states is huge, so computing and storing so many permutations appears impractical. Moreover, this is not an ordinary discrete-time DP model because the arrival rate depends on the policy that is used: the objective function is not additive and classical DP tools do not readily apply. Fortunately, our characterization of the optimal policy, which is our main technical contribution, resolves this issue. We draw some inspiration from the derivation of optimal ranking conditions

proved via an interchange argument, in a DP setting, in Bertsekas (2005, Section 4.5). However, our solution is more involved because of the impact of the policy on future arrivals.

While there are many other ‘simple’ heuristics that platforms may have selected to factor in profitability in their algorithms, we show that the particular one that turns out to be optimal is clear and simple. We think this is a nice insight that can inform platforms about how to better position their results to tradeoff relevance with profits. Of course, the time-stationary modeling assumption simplifies reality. But reputation is built over a large time-horizon and one can argue that a stationary model is a reasonable way to capture that market dynamic. A model whose parameters depend on time would give rise to complexities that go beyond the scope of this paper.

The rest of the article is organized as follows. Section 2 presents our modeling framework while Section 3 explains how the ranking problem can be simplified so one does not need to consider the exponentially-many possible orderings. Using the conditions presented there, we show how it suffices for a SE to sort the pages with respect to a scalar number, coming from a linear combination of relevance and revenue. Having characterized optimal rankings, in Section 4 we present an algorithm that computes the correct linear combination of relevance and revenue, which enables the SE to execute the sorting procedure. Section 5 discusses the impact that arises from an SE that takes a middle ground between offering a search-neutral output and a myopic one considering only short-term revenue. Finally, we offer conclusions in Section 6.

2. Model Formulation

In this section we provide the definition of the model we consider. For the presentation, we use the context of a SE that receives keyword-based queries and generates a list of organic results using links to relevant and/or profitable web pages. The model could be easily adapted to be used by other marketplaces such as electronic retailers and classified-ad websites.

For each arriving request (i.e., a query sent to the SE by a user), different content providers (CPs) host pages that are relevant. Out of a universe of m_0 pages available online, we denote by $M \leq m_0$ the number of pages that match the arriving request. Each page $i = 1, \dots, M$ has a relevance value $R_i \in [0, 1]$, and an expected revenue per click $G_i \in [0, K]$ for the CP (here, K is a positive constant) of which the SE receives a fraction $\alpha_i \in [0, 1]$. Consequently, the SE’s expected revenue per click from page i is $\alpha_i G_i$. The SE might sometimes also be the CP for a subset of the pages matching the request; in those cases $\alpha_i = 1$ because it receives all the revenue. Putting this all together, the instance of the ranking problem corresponding to a given request is encoded by a vector $Y = (M, R_1, G_1, \alpha_1, \dots, R_M, G_M, \alpha_M)$ that we assume to belong to a universe of admissible requests. After getting the request, the SE must select a permutation $\pi = (\pi(1), \dots, \pi(M))$ of the M pages and use it to display links to those pages in order. A stationary *ranking policy* μ is a function

that assigns a permutation $\pi = \mu(Y)$ to each possible realization of Y . Except when otherwise indicated, we shall only consider *deterministic* stationary policies, as opposed to *randomized* ones, which map each Y to a probability distribution over the set of permutations of M elements.

The *click-through-rate* (CTR) of a link that points to a page is defined as the probability that the user clicks on that link (Hanson and Kalyanam 2007, Chapter 8). This probability depends on the relevance of the content but also on the position number where the link is displayed. We assume that the CTR of the link to page i placed at position $\pi(i)$ can be expressed as the (separable) product of a position effect and a relevance effect. That is, CTR is given by

$$\text{CTR}(i) = \theta_{\pi(i)} \psi(R_i),$$

where $1 \geq \theta_1 \geq \theta_2 \geq \dots \geq \theta_{m_0} > 0$ is a non-increasing sequence of fixed positive constants that describe the importance of each position in the ranking. The non-decreasing function $\psi : [0, 1] \rightarrow [0, 1]$ maps the relevance to the (position-independent) probability of the page. The assumption that the CTR is separable is pervasive in the e-Commerce literature (Varian 2007, Maillé et al. 2012). We will rely on it to derive simple optimality conditions. According to this assumption, to increase the CTR, we can either choose a more relevant page or we can choose a position closer to the top of the list.

Fixing a request Y and a permutation π , we now define the various objective functions we shall consider. The *local relevance* captures the attractiveness of the ordering from the consumer's perspective. It is computed by

$$r(\pi, Y) := \sum_{i=1}^M \text{CTR}(i) R_i = \sum_{i=1}^M \theta_{\pi(i)} \psi(R_i) R_i = \sum_{i=1}^M \theta_{\pi(i)} \tilde{R}_i, \quad (1)$$

where $\tilde{R}_i := \psi(R_i) R_i$. The expected total revenue arising from the request equals

$$g_0(\pi, Y) := \sum_{i=1}^M \text{CTR}(i) G_i = \sum_{i=1}^M \theta_{\pi(i)} \psi(R_i) G_i, \quad (2)$$

out of which the SE receives

$$g(\pi, Y) := \sum_{i=1}^M \text{CTR}(i) \alpha_i G_i = \sum_{i=1}^M \theta_{\pi(i)} \psi(R_i) \alpha_i G_i = \sum_{i=1}^M \theta_{\pi(i)} \tilde{G}_i, \quad (3)$$

where $\tilde{G}_i := \alpha_i \psi(R_i) G_i$.

To obtain an optimal ranking policy, we must consider that since customers are quality-sensitive, the choice of policy $\mu(\cdot)$ influences the future arrivals of customers. This has deep implications because a myopic policy for the SE (i.e., choosing $\mu(Y) \in \arg \max_{\pi} g(\pi, Y)$ for each Y) does not suffice to achieve optimality. To capture the dependence on future end-users that arrive to the SE,

we consider the multivariate distribution of the input requests Y . Each request is then interpreted as a realization of Y according to that distribution.

We estimate the long-term value induced by a stationary ranking policy μ by taking expectations of the objectives presented earlier with respect to the distribution of input requests. Therefore, the expected relevance per request is

$$r := r(\mu) = \mathbb{E}[r(\mu(Y), Y)], \quad (4)$$

the expected total revenue per request is

$$g_0 := g_0(\mu) = \mathbb{E}[g_0(\mu(Y), Y)], \quad (5)$$

and the expected SE revenue per request is

$$g := g(\mu) = \mathbb{E}[g(\mu(Y), Y)]. \quad (6)$$

In the three previous definitions, the expectation is taken with respect to the random variable Y .

As discussed in the introduction, a non-myopic SE would be interested in the expected long-run revenue. This must depend on both the expected relevance per request r and on the expected SE revenue per request g . We capture the two dependencies through the general function

$$U_{\text{SE}} = \varphi(r, g), \quad (7)$$

where φ is an increasing function of r and g with bounded second derivatives over $[0, 1] \times [0, K]$. An *optimal policy* from the perspective of the SE is a stationary ranking policy μ that maximizes U_{SE} .

We are going to pay special attention to the class of ranking policies that sort the M pages by decreasing order of their value of $\tilde{R}_i + \rho\tilde{G}_i$, for a given constant $\rho > 0$. We refer to such a policy as a *linear ordering (LO) policy with ratio ρ* (or LO- ρ policy, for short). In fact, if $\theta_k = \theta_{k+1}$, the ordering at positions k and $k + 1$ does not matter, and we still say that we have an LO- ρ policy regardless of the order at these positions. When $\rho = 0$, the ordering is based only on \tilde{R}_i , whereas in the limit as $\rho \rightarrow \infty$, the ordering is based only on \tilde{G}_i . We show below that under mild conditions on the distribution of queries Y , an optimal policy μ^* coincides with an LO- ρ policy for a specific value of ρ that we will characterize. We also highlight that specifying ρ is not enough to uniquely characterize an optimal policy in the case when several $\tilde{R}_i + \rho\tilde{G}_i$ may be equal with positive probability.

The objective function (7) is very general, and the assumptions written after its definition are enough to develop our theoretical analysis. In practice $\varphi(r, g)$ usually takes the form of an (average)

arrival rate of requests multiplied by an expected revenue per request. To develop intuition, the examples we provide below have more structure, as we now describe. We assume that search requests arrive according to a (Poisson) process of (constant) rate $\lambda(r)$, where $\lambda: [0, 1] \rightarrow [0, \infty)$ is an increasing, positive, smooth (continuously differentiable), and bounded function. Its argument r is the average relevance corresponding to the policy in use, as defined in (4). Each time the SE receives a search request, it gets a revenue β in expectation from the third-party advertisement displayed in the page. Hence, the expected SE advertisement revenue per time unit is $\beta\lambda(r)$, which depends on the ranking policy only via r . The main assumption here is that average relevance in organic research drives reputation. Paid search is not going to drive users (significantly) to the website in the long term so the two mechanisms (organic and sponsored) co-exist without much interference between them. On top of this, and as discussed earlier, we assume that the SE receives a proportion of the CP revenue, totalling $\lambda(r)g$. Putting it all together, the total expected SE revenue per unit time in our examples is

$$U_{SE} = \lambda(r)(\beta + g). \quad (8)$$

This expression is increasing in r and in g . When $\alpha_i = 0$ for all i , $g = 0$ and the SE's best interest is being neutral to just maximize r , i.e., to rank based on relevance. Otherwise, if $g > 0$, the SE may be interested in selecting permutations that increase g even if this decreases r a bit. The larger the α_i 's, the stronger the incentive of the SE to consider non-neutral ranking policies. In the next sections, we characterize optimal strategies and develop algorithms to compute or approximate them.

3. Optimality Conditions for Ranking Policies

In this section we derive optimality conditions on the permutation $\pi = \mu(Y)$ associated with any given request vector Y . Later, these conditions will be used to develop computational algorithms that can provide a ranking for each Y . We first develop *approximate* necessary optimality conditions under the assumption that Y has a discrete distribution. Then we show that these conditions must hold in the limit if we assume that each Y has a negligible probability, that is, if Y is a continuous random variable with a density. Under further assumptions, these necessary conditions determine the optimal policy uniquely, up to a set of realizations Y of measure 0. This provides simple, approximate optimality conditions for the situation where the current request Y has a small probability, small enough that changing π in the local relevance $r(\pi, Y)$ can only bring a small change to r and g .

3.1. Necessary Optimality Conditions Under a Discrete Distribution for Y

Let us first suppose that Y has a discrete distribution $p(y) = \mathbb{P}[Y = y]$. Assume that μ is an optimal policy, with r and g the associated objectives. Since μ is optimal, for any y , permuting two successive elements in $\pi = \mu(y)$, say at positions k and $k + 1$, must not increase the expected long-term revenue. Let $\delta = \pi^{-1}$, the inverse permutation to π . Then the numbers of the pages at positions k and $k + 1$ are $\delta(k)$ and $\delta(k + 1)$. Let $\Delta\theta = \theta_k - \theta_{k+1} \geq 0$. Switching the two pages at positions k and $k + 1$ will permute the vectors $(\tilde{R}_{\delta(k)}, \tilde{G}_{\delta(k)})$ and $(\tilde{R}_{\delta(k+1)}, \tilde{G}_{\delta(k+1)})$ in (1) and (3). The changes on r and g resulting from this switch would be

$$\Delta r = (\tilde{R}_{\delta(k+1)} - \tilde{R}_{\delta(k)})\Delta\theta p(y)$$

and

$$\Delta g = (\tilde{G}_{\delta(k+1)} - \tilde{G}_{\delta(k)})\Delta\theta p(y).$$

The corresponding change in U_{SE} is

$$\Delta U_{SE} = \varphi(r + \Delta r, g + \Delta g) - \varphi(r, g) = \varphi_r(r, g)\Delta r + \varphi_g(r, g)\Delta g - \mathcal{O}((|\Delta r| + |\Delta g|)^2),$$

where φ_r and φ_g are the partial derivatives of φ with respect to r and g , respectively. The optimality of π (or equivalently of δ) implies that this change on U_{SE} cannot be positive, so we must have

$$\varphi_r(r, g)\Delta r + \varphi_g(r, g)\Delta g \leq \mathcal{O}((|\Delta r| + |\Delta g|)^2),$$

yielding, whenever $\Delta\theta p(y) > 0$,

$$\varphi_r(r, g)(\tilde{R}_{\delta(k+1)} - \tilde{R}_{\delta(k)}) + \varphi_g(r, g)(\tilde{G}_{\delta(k+1)} - \tilde{G}_{\delta(k)}) \leq \mathcal{O}(\Delta\theta p(y))$$

since \tilde{R} and \tilde{G} are bounded. This can be rewritten as

$$\varphi_r(r, g)\tilde{R}_{\delta(k+1)} + \varphi_g(r, g)\tilde{G}_{\delta(k+1)} \leq \varphi_r(r, g)\tilde{R}_{\delta(k)} + \varphi_g(r, g)\tilde{G}_{\delta(k)} + \mathcal{O}(\Delta\theta p(y)), \quad (9)$$

which must hold for all y and all k for which $\Delta\theta p(y) > 0$. If $\Delta\theta p(y) = 0$, there is no change on Δr or Δg , so the order at positions k and $k + 1$ does not matter.

For every pair (r, g) , we set

$$h(r, g) := \frac{\varphi_g(r, g)}{\varphi_r(r, g)}, \quad (10)$$

for which we assume that $\varphi_r(r, g) > 0$. Using this notation, if $p(y) \ll 1$ and we decide to neglect the $\mathcal{O}(\Delta\theta p(y))$ term in (9), we obtain the following (approximate) *necessary* optimality conditions:

When $\theta_k > \theta_{k+1}$, we must have

$$\tilde{R}_{\delta(k+1)} + h(r, g)\tilde{G}_{\delta(k+1)} \leq \tilde{R}_{\delta(k)} + h(r, g)\tilde{G}_{\delta(k)}. \quad (11)$$

These necessary conditions tell us that if μ is an optimal policy and if the $\mathcal{O}(\Delta\theta p(y))$ terms can be neglected, then μ must be an LO- ρ policy with ratio $\rho = h(r, g)$.

For the special case of the running examples introduced in the previous section, we have $\varphi(r, g) = \lambda(r)(\beta + g)$, which implies that $\varphi_r(r, g) = \lambda'(r)(\beta + g)$, $\varphi_g(r, g) = \lambda(r)$, and

$$h(r, g) = \frac{\lambda(r)}{\lambda'(r)(\beta + g)}. \quad (12)$$

The conditions in (11) suggest that in a search for a (near-)optimal policy, we may restrict ourselves to LO- ρ policies and try to optimize the value of ρ . This may appear simple at first sight, but there are potential difficulties with this plan. First, the $\mathcal{O}(\Delta\theta p(y))$ term may be non-negligible, when $p(y)$ is not very small. Second, finding the optimal ρ is not necessarily obvious or easy. Third, fixing ρ does not necessarily determine a unique policy, because there might be equalities in (11) and then the selected order might still matter. Fourth, when such equalities happen, there are situations where the optimal policy must be randomized (e.g., select one order with some probability p and the other with probability $1 - p$; see below). Fifth, the conditions (11) are necessary for an optimal policy, but perhaps not sufficient. The following example illustrates those difficulties.

EXAMPLE 1. We consider an instance with two pages and a unique request. The input data consists of $Y = (M, R_1, G_1, \alpha_1, R_2, G_2, \alpha_2) = (2, 1, 0, 0, 1/5, 8, 1/4)$ with probability 1, $\psi(R) = 1$ for all R , $\lambda(r) = r$, $\beta = 1$, and $(\theta_1, \theta_2) = (1, 1/2)$. Replacing in the objective, we have $\varphi(r, g) = r(1 + g)$.

At each request, we must select a ranking, either (1, 2) or (2, 1). Suppose that instead of always selecting the same ranking for all requests, we adopt a policy that selects the ranking (1,2) with probability p and (2,1) with probability $1 - p$. We want to find the optimal value of $p \in [0, 1]$. For this randomized policy, we compute

$$\begin{aligned} r &= p(\theta_1 R_1 + \theta_2 R_2) + (1 - p)(\theta_1 R_2 + \theta_2 R_1) = (7 + 4p)/10, \\ g &= p(\theta_1 \alpha_1 G_1 + \theta_2 \alpha_2 G_2) + (1 - p)(\theta_1 \alpha_2 G_2 + \theta_2 \alpha_1 G_1) = 2 - p, \\ \varphi(r, g) &= r(1 + g) = (7 + 4p)(3 - p)/10 = (21 + 5p - 4p^2)/10. \end{aligned}$$

The objective function is quadratic and it attains its maximum at $p^* = 5/8$. Evaluating, $r = 19/20$, $g = 11/8$, and $\varphi(r, g) = 361/160$. Note that by taking $p = 0$ we get $21/10 = 336/160$ and by taking $p = 1$ we get $22/10 = 352/160$. This clearly shows that randomized ranking policies can perform better than deterministic ones.

Here we have $h(r, g) = r/(1 + g) = (7 + 4p)/(10(3 - p))$. With the optimal $p^* = 5/8$, this expression evaluates to $h(r, g) = 2/5$. If we consider the LO- ρ rule with $\rho = \rho^* = 2/5$, we have $\tilde{R}_1 + \rho^* \tilde{G}_1 = \tilde{R}_2 + \rho^* \tilde{G}_2 = 1$. So with the ρ^* that corresponds to the optimal randomized policy, the ordering

conditions are always satisfied, regardless of the order, because the two linear expressions are equal. On the other hand, this ρ^* is not sufficient to determine the optimal policy! Any policy satisfies the ordering conditions with $\rho = \rho^*$, but is not necessarily optimal. Moreover, as mentioned earlier, none of the two possible deterministic policies is optimal.

If we adopt an LO- ρ policy with $\rho \neq \rho^*$, then the choice of ρ defines the policy uniquely, but this policy is not optimal either. Indeed, if $\rho < 2/5$, then we always take the order (1,2), which gives $r = 11/10$, $g = 1$, and $h(r, g) = 11/20 > 2/5$, so the LO rule with $\rho = h(r, g)$ tells us to always select the order (2,1). Reciprocally, if we always select the order (2,1), we obtain $h(r, g) < 2/5$ and the LO rule with $\rho = h(r, g)$ always tells us to select the order (1,2). Thus, we cannot guarantee the existence of optimal rankings with deterministic policies. \square

Perhaps one could argue that the problem of this example comes from the fact that the instance is deterministic so $p(Y) = 1$. In fact, this is not the case. It is always possible to construct request densities that assign small probabilities $p(y)$ to all request realizations y . This can be done by splitting artificially each possible realization of Y into an arbitrary large number of subrealizations, say ℓ , each one having probability $p(y)/\ell$. Conceptually, one would achieve this by adding one artificial component to Y to obtain an extended vector Y' whose added component only identifies what subrealization we have. In this case, the ranking policy could output different rankings according to the subrealization Y' that was drawn. In other words, different permutations π can be selected for the same (original) realization Y . In the limit when $\ell \rightarrow \infty$, this mechanism effectively mimics a randomized ranking policy, where for any given realization Y , each permutation π is selected with a given probability. This randomized ranking policy is effectively specified as a deterministic policy in terms of the extended vector Y' , which in the limit when $\ell \rightarrow \infty$ has a density (the artificial extra component has a continuous distribution).

Inspired by these observations, in the next subsection we study a framework in which Y is assumed to have a continuous distribution. This can be seen as an approximation when the set of possible request inputs is a huge set and each $p(y)$ is increasingly small. We will provide conditions under which there is a non-randomized optimal policy, and then show how it can be computed.

3.2. Approximation by a Continuous Distribution for Y

In this section, we extend the discussion to include continuous distributions for the input requests. We let Y be a continuous random vector, with probability measure ν over the class Ω of Borel subsets of vectors $Y = (M, R_1, G_1, \alpha_1, \dots, R_M, G_M, \alpha_M)$ where $M \in \{1, \dots, m_0\}$ and $(R_i, G_i, \alpha_i) \in [0, 1] \times [0, K] \times [0, 1]$ for each i . We assume that it has a (finite) density function f . That is, for each $D \in \Omega$, $\nu(D) = \int_D f(y) dy$. Then, if $\nu(D) > 0$, we can always select $\tilde{D} \subset D$, $\tilde{D} \in \Omega$, such that $\nu(\tilde{D})$ is positive and arbitrary small.

Suppose that μ is an optimal policy, with its corresponding r and g , and that we change μ into μ' by permuting the two successive elements at positions k and $k+1$ in $\mu(y)$, for all $y \in D$, for some fixed $D \in \Omega$. Note that $\mu(y)$ might not be the same for all $y \in D$. Let $\delta(k)(y)$ and $\delta(k+1)(y)$ be those page numbers at positions k and $k+1$ in $\mu(y)$, for each y . Switching those two pages permutes the vectors $(\tilde{R}_{\delta(k)(y)}, \tilde{G}_{\delta(k)(y)})$ and $(\tilde{R}_{\delta(k+1)(y)}, \tilde{G}_{\delta(k+1)(y)})$, for all $y \in D$. The changes on r and g coming from this switch are

$$\Delta r = \int_D (\tilde{R}_{\delta(k+1)(y)} - \tilde{R}_{\delta(k)(y)}) \Delta \theta f(y) dy$$

and

$$\Delta g = \int_D (\tilde{G}_{\delta(k+1)(y)} - \tilde{G}_{\delta(k)(y)}) \Delta \theta f(y) dy.$$

Since $\Delta \theta \leq 1$, $R_i \in [0, 1]$, $\psi(R_i) \in [0, 1]$, $G_i \in [0, K]$, and $\alpha_i \in [0, 1]$, these changes satisfy $|\Delta r| \leq \nu(D)$ and $|\Delta g| \leq K\nu(D)$. The corresponding change on φ is

$$\Delta U_{SE} = \varphi_r(r, g) \Delta r + \varphi_g(r, g) \Delta g + \mathcal{O}((|\Delta r| + |\Delta g|)^2). \quad (13)$$

We now rigorously define the linear ordering policies with ratio ρ that were mentioned in Section 2, and establish that optimal ranking policies belong to that category for a specific value of ρ .

DEFINITION 1. A policy μ is called an *LO- ρ policy* if for almost all Y (with respect to the measure ν), μ sorts the pages by decreasing order of $\tilde{R}_i + \rho \tilde{G}_i$, except perhaps at positions k and $k+1$ where $\theta_k = \theta_{k+1}$, at which the order can be arbitrary. That is, whenever $\theta_k > \theta_{k+1}$, one must have

$$\tilde{R}_{\delta(k+1)(Y)} + \rho \tilde{G}_{\delta(k+1)(Y)} \leq \tilde{R}_{\delta(k)(Y)} + \rho \tilde{G}_{\delta(k)(Y)}. \quad (14)$$

PROPOSITION 1. If the tuple (r, g) corresponds to an optimal policy, then this policy must be an LO- ρ policy with $\rho = h(r, g)$.

PROOF. The proof is by contradiction. Take a k such that $\Delta \theta := \theta_k - \theta_{k+1} > 0$, and suppose that there exists $\epsilon > 0$ and $D \in \Omega$ such that $\nu(D) > 0$ and

$$\varphi_r(r, g) \tilde{R}_{\delta(k+1)(Y)} + \varphi_g(r, g) \tilde{G}_{\delta(k+1)(Y)} > \varphi_r(r, g) \tilde{R}_{\delta(k)(Y)} + \varphi_g(r, g) \tilde{G}_{\delta(k)(Y)} + \epsilon$$

for all $Y \in D$. Then,

$$\begin{aligned} \epsilon \nu(D) &\leq \int_D \left[\varphi_r(r, g) \tilde{R}_{\delta(k+1)(y)} + \varphi_g(r, g) \tilde{G}_{\delta(k+1)(y)} - \varphi_r(r, g) \tilde{R}_{\delta(k)(y)} - \varphi_g(r, g) \tilde{G}_{\delta(k)(y)} \right] f(y) dy \\ &= (\varphi_r(r, g) \Delta r + \varphi_g(r, g) \Delta g) / \Delta \theta. \end{aligned}$$

The current policy being optimal, performing the permutation of positions k and $k + 1$ over D can only reduce the revenue, i.e., $\Delta U_{SE} \leq 0$. From (13), for $\nu(D)$ sufficiently small there exists a constant C such that

$$\varphi_r(r, g)\Delta r + \varphi_g(r, g)\Delta g \leq C(|\Delta r| + |\Delta g|)^2 \leq C(1 + K)^2\nu^2(D).$$

Therefore, $\nu(D) \geq \epsilon\Delta\theta / (C(1 + K)^2)$. This also holds for D replaced by any $\tilde{D} \subset D$ with $\tilde{D} \in \Omega$. By taking $\tilde{D} \subset D$ small enough so that $\nu(\tilde{D}) < \epsilon\Delta\theta / (C(1 + K)^2)$, we obtain a contradiction. Therefore, (14) must hold with $\rho = h(r, g)$ for all Y except perhaps over a set of measure zero. \square

Proposition 1 tells us that any optimal policy must satisfy the LO- ρ conditions for $\rho = h(r, g)$. But we need further assumptions to make sure that this specifies an optimal policy. In the rest of this section, we assume that the following condition holds.

ASSUMPTION A. For any $\rho \geq 0$, and any $j > i > 0$, $\mathbb{P}[M \geq j \text{ and } \tilde{R}_i + \rho\tilde{G}_i = \tilde{R}_j + \rho\tilde{G}_j] = 0$. \square

One example of a sufficient condition for this assumption to hold is that each pair $(\tilde{R}_i, \tilde{R}_j)$ has a bivariate density with no probability mass concentrated on a set of Lebesgue measure zero in two dimensions (such as a line, for example).

Under Assumption A, for any fixed $\rho \geq 0$, there is an LO- ρ policy $\mu = \mu(\rho)$ that sorts almost any $Y \in \Omega$ (i.e., there is a unique order with probability 1). This ranking policy has corresponding values of $(r, g) = (r(\mu), g(\mu))$ and of $h(r, g)$ that are uniquely defined. To refer to $h(r, g)$ as a function of ρ , we write $\tilde{h}(\rho)$. From Proposition 1, if μ is the optimal policy, then ρ must be a fixed point of \tilde{h} . Indeed, one must have $\rho = h(r(\mu(\rho)), g(\mu(\rho))) = \tilde{h}(\rho)$. The next proposition says that under certain conditions such a fixed point exists and is unique.

PROPOSITION 2. (i) If $h(r, g)$ is bounded over $[0, 1] \times [0, K]$, then the fixed-point equation

$$\tilde{h}(\rho) = \rho \tag{15}$$

has at least one solution in $[0, \infty)$.

(ii) If the derivative $\tilde{h}'(\rho) < 1$ for all $\rho > 0$, then the solution is unique.

PROOF. (i) If $\tilde{h}(0) = 0$, then $\rho = 0$ is already a solution. Otherwise, we have $\tilde{h}(0) > 0$ and $\tilde{h}(\rho) \leq K'$ for all $\rho \geq 0$ for some constant K' . In particular, $\tilde{h}(K') \leq K'$. Since \tilde{h} is a continuous function, there must be at least one point $\rho \in [0, K']$ at which $\tilde{h}(\rho) = \rho$.

(ii) The slope of the function $\tilde{h}(\rho)$ is always less than 1, so it cannot cross the line $f(\rho) = \rho$ more than once. \square

Going back to the special case of our running examples where $\varphi(r, g) = \lambda(r)(\beta + g)$, Proposition 2 becomes:

PROPOSITION 3. Suppose $\varphi(r, g) = \lambda(r)(\beta + g)$.

(i) If $\lambda(r)/\lambda'(r)$ is bounded for $r \in [0, 1]$ and $\beta + g(0) > 0$, then (15) has at least one solution in $[0, \infty)$.

(ii) If $\lambda(r)/\lambda'(r)$ is also non-decreasing in r , then the solution is unique.

PROOF. (i) In this case, we have

$$\tilde{h}(\rho) = \frac{\lambda(r(\mu(\rho)))}{\lambda'(r(\mu(\rho)))(\beta + g(\mu(\rho)))}.$$

Note that $g(\mu(\rho)) \geq g(\mu(0))$ for all $\rho \geq 0$. Therefore, the conditions in (i) imply that $\tilde{h}(\rho)$ is bounded and we can apply Proposition 2 (i).

(ii) If $\lambda(r)/\lambda'(r)$ is non-decreasing in r , then it is non-increasing in ρ since $r(\mu(\rho))$ is non-increasing in ρ . Additionally, since we know that $g(\mu(\rho))$ is non-decreasing in ρ , it follows that $\tilde{h}(\rho)$ is non-increasing in ρ , so $\tilde{h}'(\rho) \leq 0$ and we can apply Proposition 2 (ii). \square

The condition that $\lambda(r)/\lambda'(r)$ is non-decreasing, in (ii), is actually a bit stronger than what we need to satisfy the condition of Proposition 2 (ii). To illustrate when this condition is satisfied, take $\lambda(r) = a_0 + b_0 \ln(c_0 + r)$ for some constants $a_0 \geq 0$, $b_0 > 0$, and $c_0 \geq 1$. Then, $\lambda'(r) = b_0/(c_0 + r)$, and therefore $\lambda(r)/\lambda'(r) = [a_0 + b_0 \ln(c_0 + r)](c_0 + r)/b_0$, which is bounded and increasing in $r \in [0, 1]$. Other simple cases where the condition holds are the monomial forms $\lambda(r) = a_0 r^{b_0}$ for any positive values a_0 and b_0 ; which includes the case $\lambda(r) = r$ considered in several examples in this paper.

When $\alpha = 0$, we immediately find $\rho^* = 0$ and Proposition 3 simplifies to the following intuitive result, which establishes that it is optimal for the SE to rank according to relevance. In this case, the SE has the incentive to conform to search neutrality.

COROLLARY 1. If $\varphi(r, g) = \lambda(r)\beta$, so $g = 0$, an optimal ranking policy must always sort the pages by decreasing value of \tilde{R}_i . \square

The value of r obtained under this ordering, say r_0 , is the maximal possible value, so we always have $r \in [0, r_0]$.

We conclude the section by offering an example that establishes that having a density for Y is not sufficient for the optimal policy to be deterministic and uniquely defined by ρ^* .

EXAMPLE 2. Starting from Example 1, we add a third page with relevance R_3 uniformly distributed over $[0, \epsilon]$ for some small $\epsilon > 0$, and revenue $G_3 = 0$. We assume that $\theta_3 = 1/4$. Since R_3 has a density, $p(y) = 0$ for all $y \in \Omega$. For any $\rho > 0$, if ϵ is small enough, this third page will always be ranked last, and its impact on $h(r, g)$ is very small. Then the problem of ranking the first two pages becomes the same as Example 1, which means that the optimal policy must be randomized.

3.3. Some Illustrative Examples of our Methodology

This section provides two examples that, although simple and stylized, capture some features of the search market in the real world. The first example can model an SE that has content that competes with third-party CPs. Imagine that Google receives a video-search request and there are two pages that match the search; the first is from YouTube, owned by Google, while the second belongs to a competitor such as Dailymotion. Google's revenue generated by the YouTube page is positively correlated with the page relevance because of higher advertisement revenue and higher YouTube perception. Instead, Google's revenue generated by the Dailymotion page is negatively correlated with the page relevance because the more relevant, the more it diverts traffic from YouTube.

To simplify the exposition, we assume in all our illustrative examples that $\psi(R) = 1$, i.e., that the CTR depends only on the position of the page, but this assumption is by no means necessary or realistic, and it does not really make the computations much faster or easier. We also assume that $\varphi(r, g) = \lambda(r)(\beta + g)$.

EXAMPLE 3. Consider an instance with two pages where R_1 and R_2 are independent and uniformly distributed over $[0, 1]$, $G_1 = R_1$, $G_2 = 1 - R_2$, and $\alpha_1 = \alpha_2 = \alpha$. In addition, we let $\lambda(r) = r$, $\theta_1 = 1$, $\theta_2 = 0$, $\psi(R) = 1$, and $\varphi(r, g) = \lambda(r)(\beta + g)$. We also define $\tilde{\rho} := \alpha\rho$.

At the optimal ρ , Page 1 will be ranked before Page 2 if and only if $R_1 + \tilde{\rho}G_1 > R_2 + \tilde{\rho}G_2$; i.e., on the domain

$$D = \left\{ (R_1, R_2) : R_1 > \frac{\tilde{\rho}}{1 + \tilde{\rho}} + R_2 \frac{1 - \tilde{\rho}}{1 + \tilde{\rho}} \right\},$$

We define $\bar{D} := [0, 1]^2 \setminus D$. Then, since

$$0 < \frac{\tilde{\rho}}{1 + \tilde{\rho}} \leq \frac{\tilde{\rho}}{1 + \tilde{\rho}} + R_2 \frac{1 - \tilde{\rho}}{1 + \tilde{\rho}} \leq \frac{1}{1 + \tilde{\rho}} < 1,$$

we find that the LO- ρ policy gives

$$\begin{aligned} r &= \int_D r_1 dr_1 dr_2 + \int_{\bar{D}} r_2 dr_1 dr_2 \\ &= \int_0^1 \int_{\frac{\tilde{\rho} + r_2(1 - \tilde{\rho})}{1 + \tilde{\rho}}}^1 r_1 dr_1 dr_2 + \int_0^1 \int_0^{\frac{\tilde{\rho} + r_2(1 - \tilde{\rho})}{1 + \tilde{\rho}}} r_2 dr_1 dr_2 \\ &= \frac{2}{3} - \frac{\tilde{\rho}^2}{6(1 + \tilde{\rho})^2} \end{aligned}$$

and

$$\begin{aligned} g/\alpha &= \int_D r_1 dr_1 dr_2 + \int_{\bar{D}} (1 - r_2) dr_1 dr_2 \\ &= \frac{1}{3} + \frac{\tilde{\rho}}{6(1 + \tilde{\rho})^2} + \int_{\bar{D}} dr_1 dr_2 - \frac{1}{6} - \frac{1}{6(1 + \tilde{\rho})} \\ &= \frac{2}{3} - \frac{1}{6(1 + \tilde{\rho})^2}. \end{aligned}$$

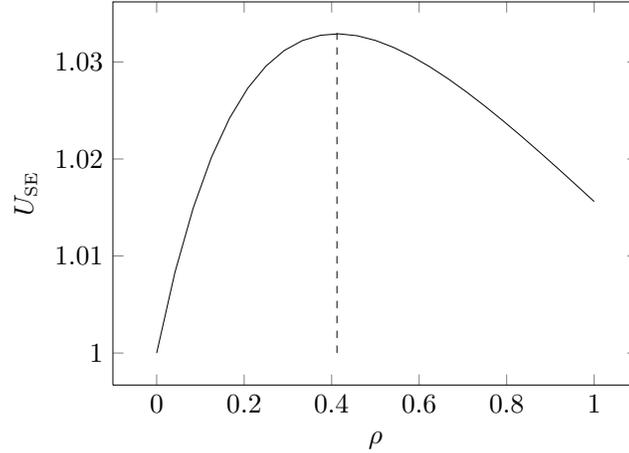


Figure 1 Expected SE Revenue in terms of ρ when $\alpha = \beta = 1$

The SE revenue is thus

$$\begin{aligned} U_{SE} &= r(\beta + g) = \left(\frac{2}{3} - \frac{\tilde{\rho}^2}{6(1 + \tilde{\rho})^2} \right) \left(\beta + \frac{2\alpha}{3} - \frac{\alpha}{6(1 + \tilde{\rho})^2} \right) \\ &= \frac{\alpha\tilde{\rho}^2/36 - (3\beta + 2\alpha)\tilde{\rho}^2(1 + \tilde{\rho})^2/18 - \alpha(1 + \tilde{\rho})^2/9}{(1 + \tilde{\rho})^4}, \end{aligned}$$

which we want to maximize over $\tilde{\rho} \geq 0$. Taking the derivative with respect to $\tilde{\rho}$ and setting it to 0, we get the following equation, whose root divided by α (since $\tilde{\rho} = \alpha\rho$) provides us with ρ^* :

$$(3\beta + 2\alpha)\tilde{\rho}^3 + (6\beta + 5\alpha/2)\tilde{\rho}^2 + (3\beta - 5\alpha/2)\tilde{\rho} - 2\alpha = 0.$$

It follows from Proposition 3 (ii) that this ρ^* is unique. Moreover, Assumption A is satisfied, so ρ^* defines the order uniquely with probability 1.

To complete this example numerically, let us take $\alpha = \beta = 1$. Then, ρ^* is the unique positive root of $5\rho^3 + 17\rho^2/2 + \rho/2 - 2$, which is $\rho^* \approx 0.412149553$. Figure 1 shows the expected SE revenue for this example, as a function of ρ . One can also compute that for $\rho = \rho^*$, we have $r = 0.6524696521$ and $g = 0.583089554$, and then $h(r, g) = r/(1 + g) = 0.412149553 = \rho^*$, as expected. \square

In our next example, we consider payments that are all-or-nothing, where the SE gets a revenue when it is also the CP serving the corresponding page. Otherwise, the SE does not have any financial gain when showing the link to the page. An alternative interpretation of this example is that content, instead of being served by the SE, is served by CPs some of whom agree to pay the SE a fixed price, normalized to 1, for each click to their pages served from the SE's output. That price is not tied to a fixed position for the link; it just provides an incentive so the SE favors links with $G_i = 1$ in its ranking.

EXAMPLE 4. Consider an instance with two pages ($M = 2$) where revenues G_i can only take values 0 and 1. Indeed, for $i = 1, 2$, G_i is a Bernoulli random variable with parameter p , R_i has a uniform distribution over $[0, 1]$ independent of G_i , and $\alpha_i = 1$. We let $(\theta_1, \theta_2) = (1, 0)$, $\lambda(r) = r$, $\psi(R) = 1$, and $\varphi(r, g) = \lambda(r)(\beta + g)$. Note that this is equivalent to saying that only one page is displayed in the search output.

The density of Y is a mixture of two uniforms, which verifies Assumption A. Focusing on LO- ρ policies, we derive explicit formulas for $r = r(\rho)$, $g = g(\rho)$, and $\varphi(r(\rho), g(\rho))$. The fixed point ρ^* can be computed from them. To start, we compute the average relevance $r = r(\rho)$. We distinguish two cases for the vector $((R_1, G_1), (R_2, G_2))$:

1. If $G_1 = G_2$, only the most relevant link is displayed, resulting in conditional expected relevance $\mathbb{E}[\max(R_1, R_2) | G_1 = G_2] = 2/3$.

2. If $G_1 \neq G_2$, we can assume (possibly by swapping the roles of pages 1 and 2) that $G_1 = 1$ and $G_2 = 0$. If $R_1 + \rho \geq R_2$, link 1 is displayed and the observed relevance is R_1 ; otherwise, the observed relevance is R_2 . Note that if $\rho > 1$, link 1 is always shown, leading to an expected observed relevance of $1/2$. If $\rho \leq 1$, the expected relevance conditional on (G_1, G_2) is

$$\begin{aligned} \mathbb{E}[R_1 \mathbb{1}_{\{R_1 + \rho > R_2\}} + R_2 \mathbb{1}_{\{R_1 + \rho \leq R_2\}} | G_1 = 1, G_2 = 0] &= \int_{r_1=0}^1 \int_{r_2=0}^1 r_1 \mathbb{1}_{\{r_1 + \rho > r_2\}} + r_2 \mathbb{1}_{\{r_1 + \rho \leq r_2\}} dr_2 dr_1 \\ &= \frac{2}{3} - \frac{\rho^2}{2} + \frac{\rho^3}{3}. \end{aligned}$$

Combining the four possibilities for (G_1, G_2) , the overall expected relevance for the LO- ρ policy is

$$r = r(\rho) = \frac{2}{3} + p(1-p)\bar{\rho}^2 \left(\frac{2\bar{\rho}}{3} - 1 \right), \quad (16)$$

where $\bar{\rho} := \min(1, \rho)$.

Similarly, to compute the expected revenue $g = g(\rho)$ per request, we consider two cases:

1. If $G_1 = G_2$, the expected revenue is 0 if $G_1 = 0$, and 1 otherwise.
2. If $G_1 \neq G_2$, we can assume again that $G_1 = 1$ and $G_2 = 0$. Again, if $\rho > 1$, link 1 is always shown and the revenue is 1. If $\rho \leq 1$, the expected revenue conditional on (G_1, G_2) is

$$\begin{aligned} \mathbb{E}[\mathbb{1}_{\{R_1 + \rho > R_2\}} | G_1 = 1, G_2 = 0] &= \int_{r_1=0}^1 \int_{r_2=0}^1 \mathbb{1}_{\{r_1 + \rho > r_2\}} dr_2 dr_1 \\ &= 1 - \frac{(1-\rho)^2}{2}. \end{aligned}$$

Regrouping all cases, we obtain

$$g = g(\rho) = p^2 + 2p(1-p) \left(1 - \frac{(1-\bar{\rho})^2}{2} \right). \quad (17)$$

Note that both $r(\rho)$ and $g(\rho)$ are constant for $\rho \geq 1$, so we can reduce the search for an optimal ρ to the interval $[0, 1]$, and in that interval $\bar{\rho} = \rho$. With $\lambda(r) = r$, the expected revenue per unit of time is $U_{SE}(\rho) = r(\rho) \cdot (\beta + g(\rho))$, which equals

$$\left(\frac{2}{3} + p(1-p)\rho^2 \left(\frac{2\rho}{3} - 1 \right) \right) \cdot \left(\beta + \left(p^2 + 2p(1-p) \left(1 - \frac{(1-\rho)^2}{2} \right) \right) \right).$$

Figure 2 depicts the expected revenue as a function of ρ , along with $r(\rho)$ and $g(\rho)$, for $\beta = 1$ and $p = 1/2$. While $g(\rho)$ increases and $r(\rho)$ decreases with ρ , the maximal revenue is obtained by taking ρ around 0.4. This optimal ρ uniquely determines the optimal policy (with probability 1). \square

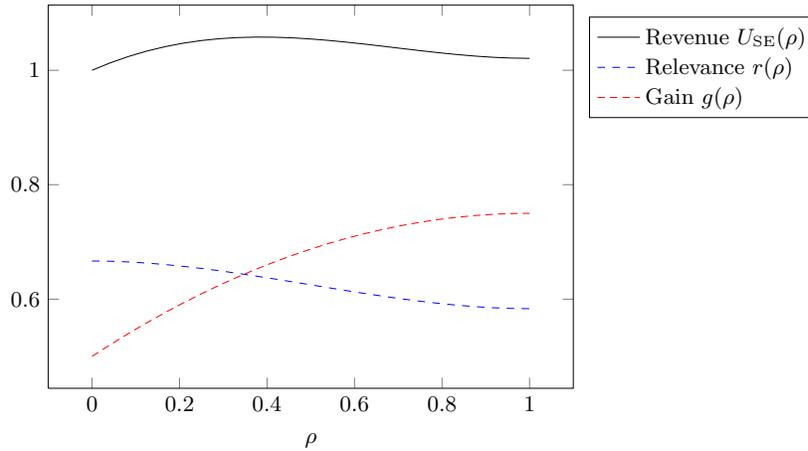


Figure 2 Expected SE revenue per unit time for $\beta = 1$ and $p = 1/2$

The previous example illustrates that to appropriately solve the tradeoff between short-term revenue coming from payments and long-term revenue coming from more exposure due to higher relevance, one must place an appropriate weight on short-term revenues and on relevance (the former being around 40% of the latter in our example). This will provide short-term benefits to the SE without impairing its possibility to attract future users.

4. Finding Optimal Rankings by Computing ρ^*

In this section, we discuss how to find the optimal ρ^* that allows the SE to determine the revenue-maximizing ranking easily. Proposition 1 shows that to achieve an optimal revenue, the SE should rank the items in the request Y by decreasing order of $\tilde{R}_i + \rho\tilde{G}_i$, for a properly chosen $\rho = h(r, g)$. But ρ depends on r and g , which in turn depend on the selected policy μ and are unknown a-priori. Moreover, typically, this dependence is not expressed in a closed-form formula. In the examples of Section 3.3, we were able to derive explicit analytical expressions for $r(\rho)$ and $g(\rho)$, and use them to find the optimal ρ . Unfortunately, instances of real size do not admit such closed-form

derivations and they would usually have to be estimated through simulation. This motivates the following *stochastic root-finding* problem: estimate a root of $\tilde{h}(\rho) - \rho = 0$ when only noisy estimates of \tilde{h} can be obtained, via simulation. Several algorithms have been designed and studied for this type of problem; see, e.g., Pasupathy and Kim (2011) and the references therein. We assume that a root exists and is unique.

An estimator $\hat{h}_n(\rho)$ of $\tilde{h}(\rho)$ at any given value of ρ can be defined and computed as follows. We generate n independent realizations Y_1, \dots, Y_n of Y , with $Y_i = (M_i, R_{i,1}, G_{i,1}, \alpha_{i,1}, \dots, R_{i,M_i}, G_{i,M_i}, \alpha_{i,M_i})$. For each i , we order the triples $(R_{i,k}, G_{i,k}, \alpha_{i,k})$ by decreasing order of $\tilde{R}_{i,k} + \rho \tilde{G}_{i,k}$, and we compute $r_i(\rho) = \sum_{k=1}^{M_i} \theta_k \tilde{R}_{i,\delta(k)}$ and $g_i(\rho) = \sum_{k=1}^{M_i} \theta_k \tilde{G}_{i,\delta(k)}$. Unbiased estimators of $r(\rho)$ and $g(\rho)$ are then $\hat{r}_n(\rho) = (1/n) \sum_{i=1}^n r_i(\rho)$ and $\hat{g}_n(\rho) = (1/n) \sum_{i=1}^n g_i(\rho)$, respectively. They lead to the estimator

$$\hat{h}_n(\rho) = \varphi(\hat{r}_n(\rho), \hat{g}_n(\rho)), \quad (18)$$

which is generally biased for finite n when φ is nonlinear, but is consistent, and the bias typically decreases as $\mathcal{O}(1/n)$ (Asmussen and Glynn 2007). For the special case where $\varphi(r, g) = \lambda(r)(\beta + g)$, this gives $\hat{h}_n(\rho) = \lambda(\hat{r}_n(\rho))(\beta + \hat{g}_n(\rho))$. A confidence interval for $\tilde{h}(\rho)$ can be computed using the Delta method (Asmussen and Glynn 2007), under the assumption that $\hat{r}_n(\rho)$ and $\hat{g}_n(\rho)$ have (approximately) a normal distribution.

When searching for a root of $\tilde{h}(\rho) - \rho$, or if we want to estimate the function \tilde{h} over some interval, we need to compute $\hat{h}_n(\rho)$ at many values of ρ . This can be done using *common random numbers* (CRN), which means that we use exactly the same n realizations Y_1, \dots, Y_n at all values of ρ at which we perform a function evaluation, or using *independent random numbers* (IRN), in which case we draw a fresh independent sample Y_1, \dots, Y_n at each ρ where we estimate $\tilde{h}(\rho)$. In the CRN case, $\hat{h}_n(\rho)$ becomes a deterministic function of ρ and this function typically varies much less than in the IRN case. The *sample average optimization* method consists in optimizing this sample function $\hat{h}_n(\rho)$ defined with CRNs. However, for any fixed n , this sample function is piecewise-constant in ρ , because it depends on ρ only via the selected permutation for each i , and therefore only takes a finite number of values as a function of ρ . As a result, its derivative is zero almost everywhere and (in general) $\hat{h}_n(\rho) - \rho$ has no exact root. Therefore, the best we can do for fixed n is to compute an *approximate* root $\hat{\rho}_n^*$ of $\hat{h}_n(\rho) - \rho$, and for this, any method that relies on the derivative of $\hat{h}_n(\rho)$ must be ruled out. We can compute the approximate root either by a method that does not rely on derivatives (such as binary search), or by a derivative-based method (e.g., a Newton-type method) by approximating the derivative with finite differences. Thus, we can compute $\hat{\rho}_n^*$ such that $\epsilon_n = |\hat{h}_n(\hat{\rho}_n^*) - \hat{\rho}_n^*|$ is small, and do this for an increasing sequence of values of n , in a way that $\epsilon_n \rightarrow 0$ when $n \rightarrow \infty$. This is possible under the assumption that $\hat{h}_n \rightarrow \tilde{h}$ uniformly when $n \rightarrow \infty$,

which usually occurs with CRNs (under mild conditions). For each considered sample size n , we would use the approximate root $\hat{\rho}_n^*$ as a starting point when finding the approximate root for the next (larger) value of n .

Another approach is to use a Robbins-Monro-type *stochastic approximation* (SA) iterative method; see Pasupathy and Kim (2011) for an overview and convergence results. For the situation where $\tilde{h}(\rho) - \rho$ is decreasing in ρ , SA starts from some ρ_0 and generates iterates of the form

$$\rho_{j+1} = \rho_j + a_j(\hat{h}_{n_j}(\rho_j) - \rho_j), \quad (19)$$

where $\hat{h}_{n_j}(\rho_j)$ is an estimate of $\tilde{h}(\rho_j)$ based on sample size n_j . These estimates are independent across values of j , and $\{a_j, j \geq 0\}$ is a slowly-decreasing sequence such that $\sum_{j=0}^{\infty} a_j = \infty$ and $\sum_{j=0}^{\infty} a_j^2 < \infty$. Note that there is no need to have $n_j \rightarrow \infty$; one can take n_j as a small constant independent of j . If we replace a_j by the inverse derivative $1/(\tilde{h}'(\rho_j) - 1)$ and the estimate of $\tilde{h}(\rho_j)$ by its exact value, we obtain the Newton method, which usually converges much faster, but requires knowledge of the function and of its derivative (or accurate estimators and $n_j \rightarrow \infty$), in contrast to SA. On the other hand, without a good choice of the a_j 's, SA might converge extremely slowly.

If we replace a_j by 1 in (19), we obtain

$$\rho_{j+1} = \rho_j + (\hat{h}_{n_j}(\rho_j) - \rho_j) = \hat{h}_{n_j}(\rho_j). \quad (20)$$

If $n_j \rightarrow \infty$, this iteration becomes equivalent in the limit to the mapping $\rho \rightarrow \tilde{h}(\rho)$. Recall that $\rho \rightarrow \tilde{h}(\rho)$ is a *contraction mapping* if there is a constant $\gamma \in [0, 1)$ such that

$$|\tilde{h}(\rho) - \tilde{h}(\rho')| \leq \gamma|\rho - \rho'|$$

for all $\rho, \rho' \geq 0$. A sufficient condition for this to hold is that $|\tilde{h}'(\rho)| \leq \gamma$ for all ρ (in the region of interest). When this holds, we can start from some $\rho_0 > 0$ and iterate: $\rho_{j+1} = \tilde{h}(\rho_j)$, for $j = 1, 2, \dots$. Then, the fixed-point theorem for contraction mappings (Bertsekas and Shreve 1978) guarantees that $\rho_j \rightarrow \rho^*$ at a geometric rate: $|\rho_j - \rho^*| \leq \gamma^j |\rho_0 - \rho^*|$, which provides very fast convergence when $\gamma \ll 1$. In practice, we can replace $\tilde{h}(\rho_j)$ by $\hat{h}_{n_j}(\rho_j)$, and convergence to ρ^* will occur if $n_j \rightarrow \infty$ when $j \rightarrow \infty$. On the other hand, if n_j does not increase with j , ρ_j will generally not converge to θ^* . If n_j is fixed to some large constant n and we use IRN, ρ_j will never converge but wander around in a small neighborhood of θ^* . If we use CRNs, it will converge to a value close to θ^* , but generally different.

It is very common in our model that $\rho \rightarrow \tilde{h}(\rho)$ is a contraction mapping. In particular, this holds in all the examples considered in this paper. Generating iterates of (20), we verified that in all cases it converged very quickly to a very good approximation of ρ^* .

To illustrate the previous discussion about computation we solve the examples provided in Section 3.3 numerically.

Table 1 Values of ρ_j at the first six iterations of (20) for Example 5, with IRN and CRN

Method	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6
IRN	0.4444478	0.4102066	0.4122955	0.4121269	0.4121527	0.4121252
CRN	0.4444478	0.4101850	0.4122638	0.4121351	0.4121428	0.4121424

Table 2 Values of ρ_j at the first six iterations of (20) for Example 6, with IRN and CRN

Method	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6
IRN	0.4444471	0.377720	0.387115	0.3857771	0.3860725	0.3859246
CRN	0.4444471	0.377670	0.387079	0.3857318	0.3859223	0.3858940

EXAMPLE 5. We revisit Example 3 and estimate $r(\rho)$, $g(\rho)$, and $U_{SE}(\rho) = \varphi(r(\rho), g(\rho))$ with IRN. In particular, we take a sample of size $n = 10^7$, and ρ from 0 to 1 with a step size of 0.001. The plot on the left of Figure 3 displays the estimates of $U_{SE}(\rho)$. This gives an idea of the high-frequency estimation noise achieved with IRN. The true maximum of $U_{SE}(\rho)$ is found at $\rho^* = 0.41214955$ whereas the numerical estimate is $\rho = 0.437$. We see that, even with this large sample size, the noise is significant compared with the variation of $U_{SE}(\rho)$ around ρ^* . This illustrates the fact that sample-average optimization with IRN is not a good method to approximate the optimal ranking policy, because of the large high-frequency noise in the sample function.

We also applied the mapping (20) for several iterations, starting at $\rho_0 = 0$, with a fixed sample size of $n_j = 10^7$ for all j , and IRN across iterations. This gave already $\rho_1 = 0.44446$ at the first iteration and $\rho_4 = 0.4121$ (accurate up to four digits) after four iterations. Thus, this method gets close to the optimum very quickly. For comparison, we ran the same method with CRN. The values of ρ_j with both methods are shown in Table 1, for $j = 1, \dots, 6$. With both methods, ρ_j provides a good approximation of ρ^* very quickly. With CRN, it converges to 0.4121425 for $j \geq 7$, which is not the correct value but is accurate to five digits.

To show that we indeed have a contraction mapping, recall that for this example, $\lambda(r) = r$, $r = 2/3 - \rho^2/(6(1+\rho)^2)$, and $g = 2/3 - 1/(6(1+\rho)^2)$. This gives

$$\tilde{h}(\rho) = \frac{2/3 - \frac{\rho^2}{6(1+\rho)^2}}{\beta + 2/3 - \frac{1}{6(1+\rho)^2}} \quad \text{and} \quad \tilde{h}'(\rho) = -\frac{2(4\rho^2 + 6\beta\rho^2 + 6\beta\rho + 7\rho + 4)}{(6\beta + 12\beta\rho + 6\beta\rho^2 + 3 + 8\rho + 4\rho^2)^2}.$$

For $\beta = 1$, one can verify that $\tilde{h}'(\rho)$ is negative and increasing, with $|\tilde{h}'(\rho)| \leq |\tilde{h}'(0)| = 8/81 < 1$. Therefore the mapping $\rho \rightarrow \tilde{h}(\rho)$ is contracting with $\gamma = 8/81$. \square

EXAMPLE 6. We also revisit Example 4 and solve the problem numerically. The plot on the right of Figure 3 shows the estimates of $r(\rho)$, $g(\rho)$, and $U_{SE}(\rho)$, computed with IRN with $n = 10^5$. We also superimpose the corresponding exact curves. Again, we applied (20) for six iterations, starting with $\rho_0 = 0$, and a fixed sample size of $n_j = 10^7$ for all j . The results are in Table 2. We find that $\rho^* \approx 0.3859$.

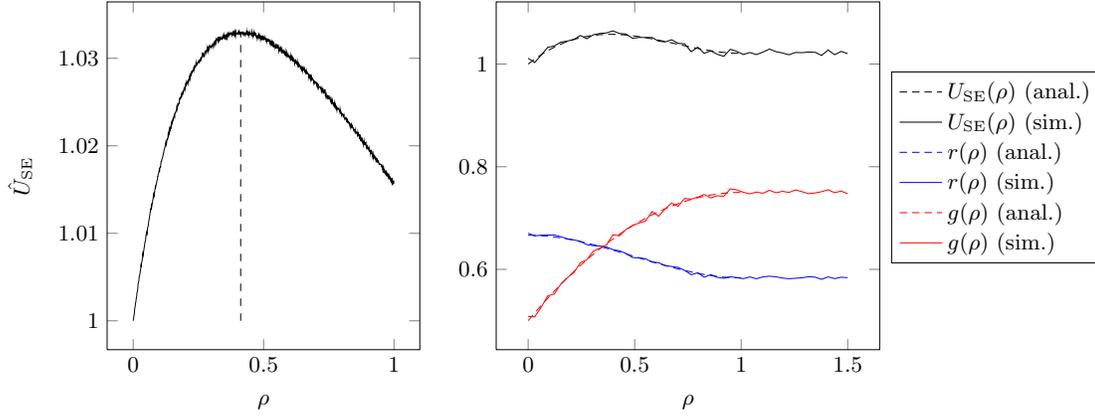


Figure 3 (Left) Estimate of expected SE revenue per unit time in terms of ρ for $\alpha = \beta = 1$ in Example 3, and (Right) Estimate of expected SE revenue per unit time for $\alpha = 1, \beta = 1, p = 1/2$ in Example 4

For this example, with the expressions previously derived for r and g , we get

$$\tilde{h}(\rho) = \frac{2/3 + p(1-p)\bar{\rho}^2(2\bar{\rho}/3 - 1)}{\beta + p^2 + 2p(1-p)(1 - (1-\bar{\rho})^2/2)}$$

and

$$\tilde{h}'(\rho) = -\frac{2}{3}p(1-p)(1-\bar{\rho}) \frac{(3\bar{\rho}\beta + 3p\bar{\rho} + 3p\bar{\rho}^2 - p\bar{\rho}^3 - 3p^2\bar{\rho}^2 + p^2\bar{\rho}^3 + 2)}{(\beta + p + 2p\bar{\rho} - p\bar{\rho}^2 - 2p^2\bar{\rho} + p^2\bar{\rho}^2)^2}.$$

For $\beta = 1$ and $p = 1/2$, one can verify numerically that for $0 \leq \rho \leq 1$, $\tilde{h}'(\rho)$ is negative and achieves a maximum absolute value of approximately $0.15 < 1$ (although the derivative is not monotone). Hence, we have a contraction mapping with $\gamma \approx 0.15$ in that area. \square

5. Comparison of the Neutral and Non-Neutral Ranking Policies

In this section, we show via numerical examples how the theory developed earlier can be used to study the impact of different ranking policies on various performance indicators such as consumer welfare (captured by expected relevance), SE and CP revenue. In particular, we compare neutral ranking policies, where $\rho = 0$, with non-neutral ones, where the SE chooses the optimal ρ^* .

5.1. A Vertically Integrated SE with a CP

EXAMPLE 7. We first focus on a specific type of request which can be served by either third-party CPs or by the SE itself. This is typical for many search categories where the SE also provides content (e.g., video, weather, finance, news, maps, flight information, and so on). In this case, a limited number of CPs compete with the SE, and the parameters r , g , and $\lambda(r)$ for the instance correspond to just this type of request. Let us assume that always ten pages match a request ($M = 10$). Nine of those pages are served by third-party CPs but one of them is served by the SE

Table 3 CTR values used in the simulations of Section 5

θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9	θ_{10}
0.364	0.125	0.095	0.079	0.061	0.041	0.038	0.035	0.03	0.022

directly. Perhaps renumbering CPs, we have that $\alpha_1 = 1$, and $\alpha_2 = \dots = \alpha_{10} = 0$. In addition to the revenue coming from Page 1, the SE also receives an expected revenue of $\beta = 1$ per request from sponsored links. For $i = 1, \dots, 10$, R_i and G_i are all independent random variables uniformly distributed over $[0, 1]$, and $\text{CTR}(i) = \theta_i$ as specified in Table 3. Those numbers were taken from the first table in Dejarnette (2012), which contains the observed relative numbers of clicks according to the position: the actual CTRs should therefore be proportional to those numbers, and the value of the multiplicative constant has no impact on our derivations (hence we take it equal to 1). Finally, we set $\lambda(r) = r$, and ψ to be the unit function.

The M pages are ranked by the SE by decreasing value of $\tilde{R}_i + \rho \tilde{G}_i$, for the correct constant $\rho \geq 0$. Note that for $i > 1$, $\tilde{G}_i = 0$ because $\alpha_i = 0$. To illustrate the dependence on ρ , Figure 4 shows the SE revenue $U_{\text{SE}}(\rho)$, as ρ varies, as well as the relevance $r(\rho)$, the revenue and the visit rate for CP 1 and for third-party CPs. All revenues are expressed as values per time unit. As discussed earlier, the more ρ increases, the more the SE favors CP 1, decreasing the overall relevance and increasing the visit rate to CP 1. The trade-off between short-term revenue and number of visits tells the SE to choose $\rho^* \approx 0.55$. Note that the bias affects only CP 1 and that the relative positions of all other CPs remain the same as in the neutral ranking. Consequently, the relevance $r(\rho)$ is only marginally affected by ρ in this case. If R_1 was stochastically much smaller than the other R_i 's (e.g., uniform over $[0, \epsilon]$ for a small ϵ), then the impact of ρ would be larger. When $\rho \rightarrow \infty$, CP 1 is always ranked first, so the relevance $r(\rho)$ becomes

$$r(\infty) = \left(\frac{\theta_1}{2} + \sum_{i=1}^9 \theta_{i+1} \mathbb{E}[U_{(10-i)}] \right) = \frac{\theta_1}{2} + \sum_{i=1}^9 \theta_{i+1} \frac{(10-i)}{10} \approx 0.517,$$

where $U_{(1)}, \dots, U_{(9)}$ are independent random variables uniformly distributed over $[0, 1]$ sorted by increasing order (the order statistics), and the CP 1 visit rate is $\theta_1 r(\infty) \approx 0.188$.

To assess the sensitivity of the SE strategy to advertising, we now examine how results change for different values of β , i.e., depending on the level of advertisement revenues. This shows the tradeoff that the SE faces for different types of requests. For search keywords related to, e.g., airline tickets, hotel reservations, or retailer products, the SE may expect to make more profit by showing its own content among organic links than through sponsored search because requests of this kind may produce conversions, whereas for keywords that are appealing in the sponsored search market the SE may try to make the search as relevant as possible to boost that revenue stream. Figure 5(a) plots ρ^* as β varies while Figure 5(b) plots the ensuing revenue for CP 1 and for each third-party

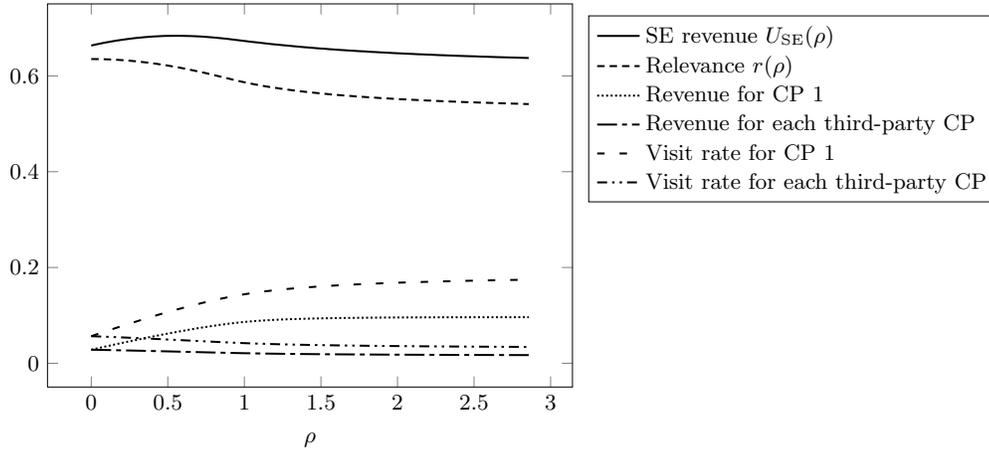


Figure 4 Performance measures as a function of ρ (simulation results)

CP. The curves shown in the figures were estimated by simulation, using the iterative fixed-point method for ρ^* , with a fixed sample size of $n = 10^7$ at each step. When β grows, ρ^* tends to zero, because the revenue from sponsored links dominates, making it rewarding for the SE to improve quality to attract more users. In conclusion, the impact of non-neutrality is small because biasing the ranking only attracts limited additional revenue. Instead, when β is small, sponsored links do not pay off and it becomes worthwhile for the SE to sacrifice relevance to some extent to boost revenue from gains of CP 1. In the extreme case when $\beta = 0$, we have $\rho^* = \infty$, so CP 1 is always placed at the top regardless and the other CPs are sorted by decreasing order of relevance. This gives an average revenue of 0.09619 for CP 1 and 0.01695 for any other CP (even though all CPs have the same relevance and gain distributions). Although not shown in the figure, we remark that U_{SE} tends to grow linearly with β , which means that the increasing revenues of sponsored search dominate the possible revenue coming from CP 1. To illustrate the impact of non-neutrality, Table 4 reports the variations of the most relevant performance metrics when $\rho = \rho^*$ is used instead of $\rho = 0$ (neutral ranking), for different values of β . The table illustrates that while the impact on the perceived quality (relevance) remains small (around 10%), the impact on the visibility and the revenues of the SE-owned CP is substantial: by being non-neutral, the SE can multiply the revenues of its CP by a factor 2.8 and its visit rate by more than a factor of 3. On the other hand, the other CPs see their revenues and visit rates reduced by 14% to 32%, a significant loss that is likely to affect their possibilities of being profitable in the long term.

Finally, we explore the sensitivity of outcomes to the number of available results. Figure 6(a) plots ρ^* as a function of M while Figure 6(b) plots revenues as a function of the number of matching pages M . We include curves for both the neutral ($\rho = 0$) and non-neutral ($\rho = \rho^*$) regimes to compare both situations. As before, we estimate these values using the fixed-point algorithm with

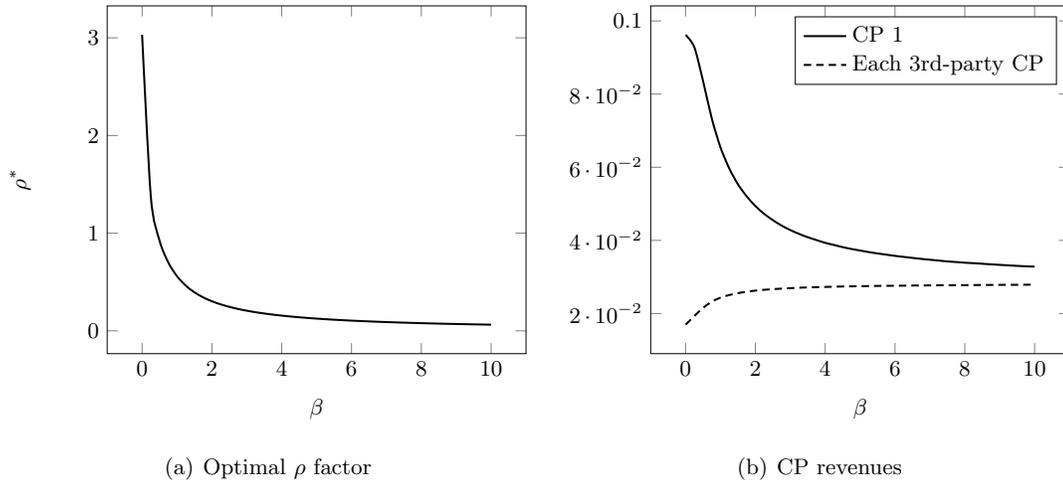


Figure 5 Optimal ρ factor to use for the ranking, and corresponding CP revenues per time unit as a function of β (simulation results)

Table 4 Impacts of a non-neutral ranking for the scenario of Section 5.1

	Relevance	CP 1 revenue	other CP revenue	CP 1 visit rate	other CP visit rate
Neutral, $\rho = 0$ (reference case optimal for $\beta = \infty$)	0.635	0.028	0.0283	0.057	0.057
Non-neutral, $\rho = 0.559$ (optimal for $\beta = 1$)	0.618 (-3%)	0.066 (+136%)	0.0243 (-14%)	0.112 (+96%)	0.049 (-14%)
Non-neutral, $\rho = 0.924$ (optimal for $\beta = .5$)	0.592 (-7%)	0.084 (+200%)	0.0215 (-24%)	0.140 (+146%)	0.043 (-25%)
Non-neutral, $\rho = 1.374$ (optimal for $\beta = .25$)	0.568 (-11%)	0.093 (+232%)	0.0193 (-32%)	0.158 (+177%)	0.039 (-32%)

$n = 10^7$ at each step. As M increases, ρ^* increases too: The SE can give more weight to CP 1 and increase its revenue while making less damage to the relevance, because placing CP 1 higher has less impact on the overall relevance when M is larger. As a result, the revenue of CP 1 when $\rho = \rho^*$ increases with M , and so does the advantage of CP 1 over the other CPs. The loss of revenue of the other CPs seems close to constant as a function of M .

5.2. Vertical Integration and Investment

EXAMPLE 8. Continuing with the example of vertical integration, we now assume that one of the nine third-party CPs, say CP 2, invests in quality and manages to improve the relevance distribution. More specifically, we assume that when it invests $z > 0$, the relevance of CP becomes uniformly distributed over $[0, 1 + 20z]$ (instead of over $[0, 1]$). The other parameters and distributions, including the distribution of its gain G_2 , are unchanged. Figures 7(a) and 7(b) show simulation results

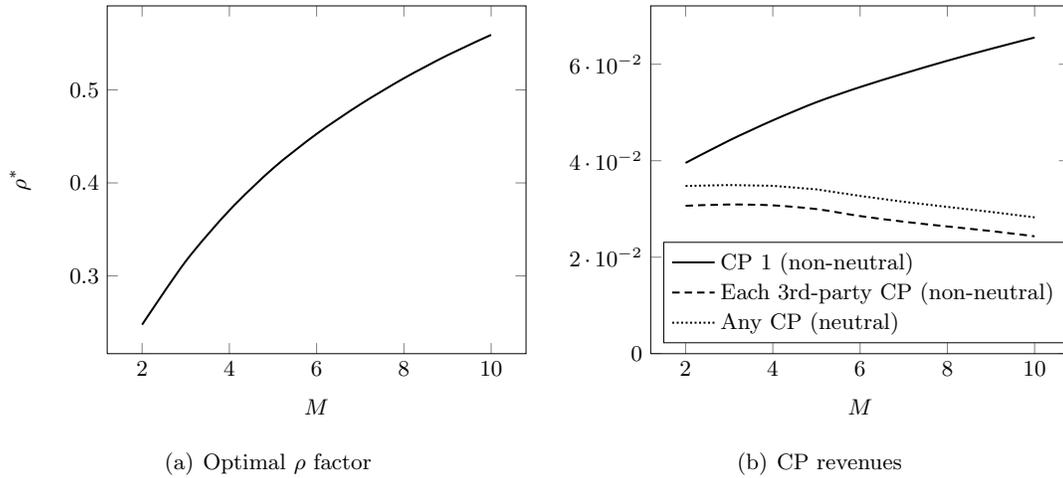


Figure 6 Optimal ρ factor to use for the ranking, and corresponding CP revenues per time unit as a function of M (simulation results)

when the SE ranks CPs according to $\tilde{R}_i + \rho \tilde{G}_i$, for varying values of ρ , and when $z = 2$. For a neutral ranking ($\rho = 0$), CP 2 logically makes more revenue than the other CPs, since it regularly gets higher ranking. However, when ρ increases and exceeds about 0.8, CP 1 becomes the one with highest revenue, despite its (stochastically) lower relevance.

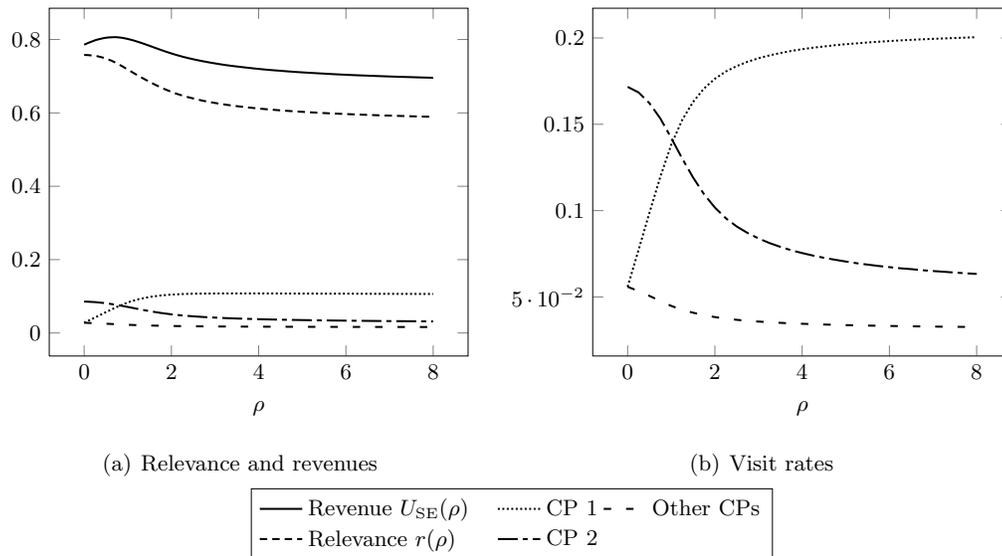


Figure 7 Average relevance, revenues and number of visits per time unit for the case of vertical integration with investment

We now take the perspective of CP 2, and compute its optimal decision. CP 2 invests z in quality to modify its relevance distribution to $[0, 1 + 20z]$, anticipating that the SE is going to rank

requests according to ρ^* . (We assume that the SE can learn the distribution of relevance of all CPs quickly.) Therefore, CP 2's profit equals the revenue from the search market minus z . To find the optimal value of z we simulated the outcomes for $z \in [0, 0.45]$. Figures 8(a) and 8(b) plot the resulting curves. In both figures, we see that differences between neutral and non-neutral revenues are small, except for CP 1. This is particularly true for CP 2. This means that, at least in this case, non-neutrality does not deter innovation. Actually, the optimal investment level under both regimes coincide and is equal to $z^* = 0.025$. Optimal profits, though, vary. They are 0.037 for the neutral case and 0.0296 for the non-neutral one; see Figure 9 where we show CP 2 profits as a function of the investment z .

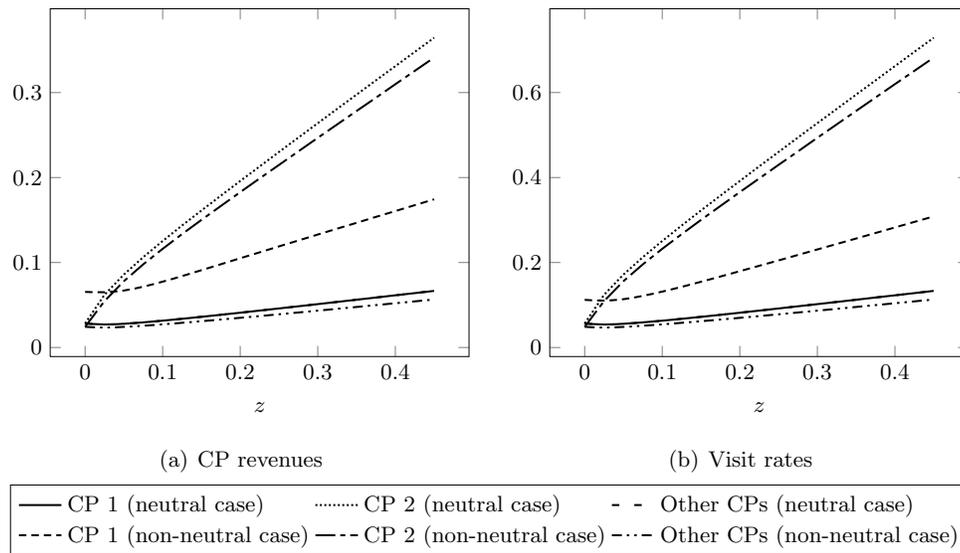


Figure 8 Revenues and visit rates to various CPs as a function of CP 2 investment

6. Conclusion

We have introduced a new modeling framework that allows online platforms to rank items accounting for both short-term and long-term revenues. The long-term impact is captured by the arrival rate of requests, which is an increasing function of the average relevance of displayed results. Under appropriate regularity conditions, we proved that although we have to choose an ordering among an exponential number of possibilities and the objective function is nonlinear, the task reduces to computing a linear combination between relevance and short-term profits for each item and then sorting items with respect to those numbers. Henceforth, the whole problem reduces to finding the appropriate constant used in the linear combination. We have also provided algorithms to find such constant.

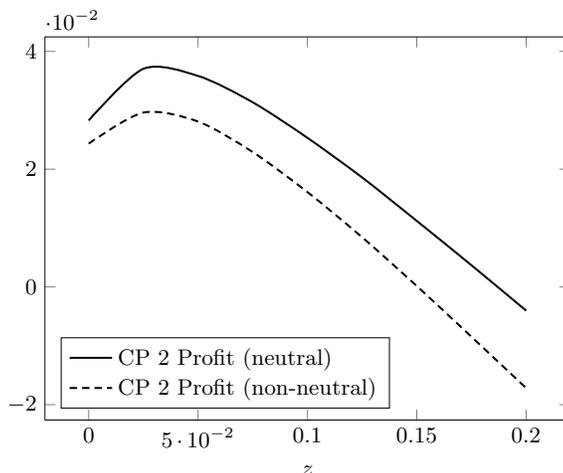


Figure 9 Profit per time unit as a function of CP 2 investment

Our results might prove useful to platform owners (search engines, classified ads websites, online retailers) to navigate the tradeoff between short-term and long-term effects when defining their ranking strategies. They can also be of interest to regulators, seeking to understand the behavior of revenue-oriented platforms and to anticipate the impact of regulatory interventions, which is of particular importance with regard to the current *search neutrality* debate.

Future work will take several directions. In particular, we plan to (i) work on the design of optimal randomized policies as highlighted in the case of discrete distributions of requests; this will also allow us to relax Assumption A in the case of a continuous distribution for requests; (ii) gather real data and providing a practical case study; (iii) perform a profound study of the implications of profit-maximizing platforms on the online economy to shed light on the search-neutrality debate.

Acknowledgments

This work has been supported by Inria's associated team MOCQUASIN, as well as an Inria International Chair, a Canada Research Chair, an NSERC Discovery Grant to P. L'Ecuyer and by CONICET Argentina Grant PIP 112-201201-00450CO and FONCYT Argentina Grant PICT 2012-1324.

References

- Asmussen, S., P. W. Glynn. 2007. *Stochastic Simulation*. Springer-Verlag, New York.
- Athey, S., G. Ellison. 2011. Position auctions with consumer search. *Quarterly Journal of Economics* **126**(3) 1213–1270.
- Auction Insights. 2008. Decoding eBay's best match. https://web.archive.org/web/*/http://www.auctioninsights.info/decoding-ebays-best-match.html . Last accessed Oct 2014.
- Austin, D. 2006. How google finds your needle in the web's haystack. *American Mathematical Society Feature Column* **10**(12). <http://www.ams.org/samplings/feature-column/fcarc-pagerank>.

- Avrachenkov, K., N. Litvak. 2004. Decomposition of the Google PageRank and Optimal Linking Strategy. Rapport de recherche RR-5101, INRIA. URL <http://hal.inria.fr/inria-00071482>.
- Bertsekas, D. P. 2005. *Dynamic Programming and Optimal Control, Volume I*. 3rd ed. Athena Scientific, Belmont, Mass.
- Bertsekas, D. P., S. E. Shreve. 1978. *Stochastic Optimal Control: The Discrete Time Case*. Academic Press, New York.
- Brill, J. 2013. Statement of the Commission regarding Google's search practices. <http://www.ftc.gov/public-statements/2013/01/statement-commission-regarding-googles-search-practices>. Last accessed Oct 2014.
- Crowcroft, J. 2007. Net neutrality: The technical side of the debate: A white paper. *ACM SIGCOMM Computer Communication Review* **7**(1).
- Dejarnette, R. 2012. Click-through rate of top 10 search results in Google. URL <http://www.internetmarketingninjas.com/blog/search-engine-optimization/click-through-rate>.
- Edelman, B., B. Lockwood. 2011. Measuring bias in "organic" web search. <http://www.benedelman.org/searchbias>.
- Edelman, B., M. Ostrovsky, M. Schwarz. 2007. Internet advertising and the generalized second-price auction: Selling billions of dollars worth of keywords. *The American Economic Review* **97**(1) 242–259.
- Google. 2011. Facts about Google and competition. <http://web.archive.org/web/20111104131332/http://www.google.com/competition/howgooglesearchworks.html>. Last accessed Oct 2014.
- Hanson, W.A., K. Kalyanam. 2007. *Internet Marketing and E-Commerce*. International student edition, Thomson/South-Western, Mason, Ohio.
- Inria. 2012. Inria's response to ARCEP consultation about network neutrality.
- Lahaie, S., D. Pennock, A. Saberi, R. Vohra. 2007. Sponsored search auctions. N. Nisan, T. Roughgarden, E. Tardos, V. Vazirani, eds., *Algorithmic Game Theory*, chap. 28. Cambridge University Press, 699–716.
- Maillé, P., E. Markakis, M. Naldi, G. Stamoulis, B. Tuffin. 2012. Sponsored search auctions: an overview of research with emphasis on game theoretic aspects. *Electronic Commerce Research* **12** 265–300.
- Maillé, P., B. Tuffin. 2014. *Telecommunication Network Economics: From Theory to Applications*. Cambridge University Press. To appear.
- Odlyzko, A. 2009. Network neutrality, search neutrality, and the never-ending conflict between efficiency and fairness in markets. *Review of Network Economics* **8**(1) 40–60.
- Pasupathy, R., S. Kim. 2011. The stochastic root finding problem: Overview, solutions, and open questions. *ACM Transactions on Modeling and Computer Simulation* **21**(3) Article 19.
- Rushe, D. 2012. Eric Schmidt Google senate hearing – as it happened. <http://www.guardian.co.uk/technology/blog/2011/sep/21/eric-schmidt-google-senate-hearing>. Last accessed Oct 2014.

Varian, H.R. 2007. Position auctions. *International Journal of Industrial Organization* **25**(6) 1163–1178.

Williams, H. 2010. Measuring search relevance. <http://www.ebaytechblog.com/2010/11/10/measuring-search-relevance>.

Wright, J. D. 2012. Defining and measuring search bias: Some preliminary evidence. George Mason Law & Economics Research Paper 12-14, George Mason University School of Law.