Universidad de la República
Facultad de Ingeniería

Technological and Economic Aspects for Quality of Service in Multidomain Alliances

Tesis presentada a la Facultad de Ingeniería de la Universidad de la República por

María Isabel Amigo

en cumplimiento parcial de los requerimientos para la obtención del título de Doctor en Ingeniería Eléctrica.

Directores de Tesis
Pablo Belzarena .... Universidad de la República
Sandrine Vaton .... Télécom Bretagne, Institut Mines-Télécom

Tribunal
Dominique Barth (Revisor externo)............. Prism, Université de Versailles
Peter Reichl (Revisor externo)........................... University of Vienna
Hélia Pouyllau .... Thales Research and Technology
Bruno Tuffin... INRIA

Director Académico
Pablo Belzarena .... Universidad de la República

Montevideo
12 de julio de 2013

Acknowledgements

It is with no hesitation that I believe that this thesis would not have been possible without the aid and company of so many people around me. I will try to summarize this in a few paragraphs, although I am sure that an exhaustive list would need several pages. Therefore, I apologize before-hand to those who are not explicitly mentioned.

First of all, I would like to thank my advisors, Sandrine and Pablo, who trusted in me from the beginning, stood beside me all the time, and helped me beyond technical aspects. I would also like to thank Thierry, Serge and Federico, whose help was extremely valuable and instructive.

I also wish to thank my family and friends in Uruguay. Firstly, because they encouraged me to start this project, even if it would keep me away from them during long periods. Secondly, for all the support I received from them through the distance during these three and a half years, which demanded a hard work on their side. I have always felt them with me, and trusting in me, sometimes even more than myself. Gracias.

But having to leave things behind makes room for new ones and I am now happy to have new people in my life to whom I am so very grateful. Thus, I want to thank all my friends I met in France, who are now a bit of my family as well. In particular, I would like to thank Lara who was literally my "hada madrina" from the very first day of my arrival in France, my room-mates, Cris, Javi and Marie, and all my friends, Amina, Amparo, Angela, Elisa, Juan David, July, Oscar, Patricia, Sandra, Santiago, Soraya and a large *et al.*. I want to thank them all because they did not get tired (or maybe they did so, but continued anyway) of repeating encouraging phrases during my most negative periods. I want to thank them because they always shared my joy during the positive ones. But above all, I want to thank them because they gave me a place in their lives, making this thesis a most wonderful and memorable experience. Merci.

I would also like to thank Armelle, Laura and María, who helped me in every administrative procedure during my thesis, always with excellent willingness to help, and all colleagues at IIE and Informatique, in particular to my office mates.

Last but not least, I would like to thank the reviewers and members of the jury who kindly accepted to review my work.

Abstract

Providing end-to-end quality-assured services implies many challenges, which go beyond technical ones, involving as well economic and even cultural or political issues. In this thesis we first focus on a technical problem and then intent a more holistic regard to the whole problem, considering at the same time Network Service Providers (NSPs), stakeholders and buyers' behaviour and satisfaction.

One of the most important problems when deploying interdomain path selection with Quality of Service (QoS) requirements is being able to rely the computations on metrics that hold for a long period of time. Our proposal for solving that problem is to compute bounds on the metrics, taking into account the uncertainty on the traffic demands. In particular, we will explore the computation of the maximum end-to-end delay of traversing a domain considering that the traffic is unknown but bounded. Since this provides a robust QoS value for traversing the NSP or Autonomous System (AS), without revealing confidential information, we claim that the bound can be safely conceived as a metric to be announced by each AS to the entities performing the path selection, in the process of interdomain path selection. We show how the maximum delay value is obtained for an interdomain bandwidth demand and we propose an exact method and a numerical approximation method for computing it, neither of which rely on a complex monitoring infrastructure. Simulations with real data that illustrate the problem and validate our results are also presented.

In the multidomain context economics and policies become more complex. In this regard, AS alliances or federations are envisaged to emerge in the near future as a means of selling end-to-end quality-assured services through interdomain networks. This collaborative paradigm mainly responds to the ever increasing Internet traffic volumes that requires assured quality, and constitutes a new business opportunity for NSPs. However, current Internet business rules are not likely to satisfy all involved partners in this emerging scenario. How the revenue is shared among NSPs must be agreed in advance, and should enforce economic incentives to join an alliance and remain in it, so that the alliance remains stable.

Inspired by this scenario, we propose a complete framework for selling interdomain quality-assured services, and subsequently distributing revenues, in an AS alliance context. We state the problem as a network utility maximization problem with QoS constraints and show that a distributed solution can be carried out.

With respect to the revenue sharing problem, we formally formulate the properties the revenue sharing method should fulfil and argue why the existing methods are not suitable. We propose a family of solutions to the revenue sharing problem such that the economic stability and efficiency of the alliance in the long term is guaranteed. The proposed method is based on solving a series of Optimization Problems and considering statistics on the incomes.

We then move to a more holistic approach and consider the interactions with the monitoring plane and the buyers' behaviour. We propose a simple pricing scheme and study it in detail, in order to use QoS monitoring information as feedback to the business plane, with the ultimate objective of improving the seller's revenue. In our framework, assured-quality Services are sold through first-price auctions, and in case of failure, a percentage of the price paid for the service is given back to the buyers. We derive the expression for the willingness to pay and we model the pricing problem through a Stackelberg game. We solve the game to show that the equilibrium that maximizes the seller's revenue implies reimbursing 100% in case of failures.

iv

The previous study is built upon a strong symmetry hypothesis, that is, buyers are assumed to be symmetric, and so do services in sale. In order to relax this assumption we present a simulative approach and evaluate the proposed pricing scheme with its aid. Results of simulations are shown in different scenarios, which in particular have shown results that are coherent with the analytical ones. That is to say, that in the evaluated scenarios reimbursing 100% provides more revenue to the seller than when no reimbursement is in place.

Key words: Interdomain Quality of Service, Alliances, Bandwidth Auctions, Revenue Sharing, Pricing, Reimbursement

Résumé

La provision de services de réseau avec qualité garanti suppose un nombre de défis qui s'étendent au-delà des aspects techniques. Au fait, ces problématiques comprennent également des aspects économiques et même culturels et politiques. Dans cette thèse, nous adressons d'abord le problème technique et ensuit nous nous lançons sur une regard plus holistique, en considérant au même temps les intéresses de plusieurs acteurs, tels que les fournisseurs de service (NSP, pour la sigle en anglais), les actionneurs et les acheteurs, y compris leur comportement et satisfaction.

L'un des problèmes le plus important quand il s'agit adresser la sélection de chemin inter-domaine avec des contraints de qualité de service (QoS) garanti, c'est d'être capable d'avoir des métriques sur les quelles baser les calculs, et d'ailleurs, d'avoir des métriques qui restent valables pendent une durée suffisamment longue. Notre proposition pour adresser ce problème est de faire un calcul basé sur des bornes de ces métriques. Ces bornes prennent en compte les incertitudes qui existent dans les demandes de trafic. Plus précisément, nous explorons le calcul de la valeur maximale du délai de traverser un domaine d'Internet, en considérant que le trafic dans le domaine est inconnu mais borné. Étant donné que cette valeur donne une métrique de QoS de traverser un domaine ou NSP, sans révéler de l'information confidentiel, nous considérons que cette borne peut être considérée sans aucun risque comme une métrique à être annoncé par chaque NSP aux autres NSPs, qui vont s'en servir pour faire des calculs dans le contexte de la sélection de chemins interdo-maines. Nous montrons comment définir formellement le problème de trouver la valeur maximale du délai et nous proposons une méthode exacte pour trouver cette valeur. Ensuite, nous proposons une méthode approximative qui permet de trouver une valeur arbitrairement proche à la valeur exacte et qui a l'avantage de consommer un temps de calcul beaucoup moins élevé que la méthode exacte. Des résultats des simulations qui utilisent des données des réseaux réels sont également présentés.

Dans le contexte multidomain les aspects économiques et les politiques à y appliquer deviennent plus complexes. Dans cet égard, l'on envisage l'émergence des alliances ou fédérations de NSPs comme un moyen pour permettre la vente de services de qualité garanti à travers des réseaux multidomains tels que l'Internet. Cette paradigme collaborative répond principalement au fait que le trafic d'Internet a à présente une tendance constante de croissance qui a besoin de qualité et qui signifie un nouveau niche de business pour les NSPs. Cependant, les lois de business actuelles de l'Internet ne sont très probablement pas intéressantes pour les différents acteurs de ce business. En particulier, comment le revenue de ces alliances sera partagé parmi leur membres doit être défini au préalable et doit encourager les NSPs, du point de vu économique, de rejoindre et rester dans l'alliance. Autrement dit, les alliances doivent être conçues de façon stable.

Inspirés par ce scénario, nous proposons une architecture complète pour permettre la vente de services de qualité garanti dans des réseaux interdomaine et pour ensuite faire le partage de revenue, dans des contextes des alliances de NSPs. Nous établissons le problème de la vente comme un problème de maximisation d'utilités (Network Utility Maximization problem) avec contraints de qualité de service et nous montrons que ce problème peut être résolu de façon distribuée.

En ce qui concerne le partage des revenues, nous formulons de façon formelle les propriétés qu'une telle méthode devrait accomplir et nous justifions le fait de que aucune méthode existante est adaptée à notre scénario. En conséquence, nous proposons une famille de solutions au problème de partage des revenues de façon telle que la stabilité de l'alliance du point de vue économique

et sa efficacité est assurée dans le long terme. La méthode proposée se base sur la résolution des problèmes d'optimisation et considère des statistiques sur les revenues.

Ensuite nous nous concentrons sur un approche plus holistique en considérant des interactions entre la couche de mesure de la qualité de service et la couche du business, et les effets que ce a sur le comportement des acheteurs. Plus en détail, nous proposons un schéma de tarification qui est simple et nous l'étudions en grand détail. Ce schéma propose de faire la vente des services en utilisant les enchères de premier prix et en assurant que si jamais la qualité de service n'est pas atteinte, l'acheteur est remboursé d'un pourcentage de ce qui il a payé pour le service. Nous trouvons l'expression mathématique pour la volontée de payer des acheteurs et nous modélisons le problème de tarification, en nous servant de la théorie des jeux, comme un jeu de Stackelberg. Nous trouvons l'équilibre du jeu qui montre que rembourser 100% maximise le revenue du vendeur.

Cette étude est développée sur la forte hypothèse des acheteurs symétriques, ce qui signifie que tous les acheteurs suivent un même modèle qui donne la valeur au service. Par contre, cette hypothèse n'est pas réaliste. C'est pour quoi ensuite nous étudions le même schéma de tarification mais en enlevant cette hypothèse. Dans ce cas là nous nous trouvons face à la problématique que quand l'on considère des acheteurs qui ne sont pas symétriques développer des résultats analytiques devient impossible dans le cas général. En réponse à cela nous proposons une étude simulative. Nous montrons sur des différents scénarios que dans le cas non symétriques les résultats obtenus sont cohérents avec le cas symétrique, c'est à dire, rembourser 100% donne plus de revenues au vendeur que ne pas de tout rembourser.

Mots clés : Qualité de service interdomaine, alliances, enchères de bande passante, partage de revenue, tarification, remboursement.

Resumen

La provisión de servicios de red con calidad garantizada de extremo a extremo implica varios desafíos, que se extienden más allá de los técnicos, involucrando aspectos tanto económicos como culturales y hasta incluso políticos. En esta tesis primero abordamos el aspecto técnico y luego intentamos una mirada más holística al problema en su conjunto, considerando al mismo tiempo proveedores de red (NSPs, por su sigla en inglés), accionistas, y el comprador o usuario, considerando su comportamiento y satisfacción.

Uno de los problemas más importantes cuando se desea implementar la selección de camino interdominio con calidad de servicio (QoS, por su sigla en inglés) es el poder contar con métricas en las cuales basar los cálculos, y que estas métricas sean válidas durante un período de tiempo suficiente. Nuestra propuesta para atacar este problema es computar una cota a estas métricas. Esta cota tiene en cuenta las variaciones que pueden haber en el tráfico por la red, siendo éstas desconocidas *a priori*. En particular, exploramos el cálculo del retardo máximo de atravesar un NSP o sistema autónomo (AS), sin revelar información confidencial del AS. Esta métrica puede ser anunciada por cada AS para hacer posible el cómputo de caminos interdominio con calidad de servicio. Mostramos cómo formular el problema de obtener el retardo máximo para una demanda de ancho de banda y proponemos un método exacto para obtener dicho valor. A su vez proponemos un método numérico que provee una aproximación de dicho valor con un menor tiempo de cómputo. Ninguno de estos dos métodos, tanto el exacto como el aproximado, requiere de una infraestructura de monitoreo compleja. Finalmente ilustramos el problema y la solución con simulaciones que utilizan datos obtenidos de redes reales.

Cuando varios dominios o ASes interactúan entre sí, los aspectos económicos y las políticas a aplicar se vuelven más complejas. En este sentido, se prevé que las alianzas o federaciones de ASes emerjan en el futuro próximo, de manera tal de que faciliten la venta de servicios con calidad de extremo a extremo a través de redes interdominio. Este paradigma colaborativo responde principalmente al constante crecimiento del tráfico de Internet que requiere calidad de servicio, y a su vez constituye una oportunidad de negocio para los NSPs. Sin embargo, las leyes que rígen actualmente el mercado de interconexión de Internet no son necesariamente atractivas para todos los actores que participarían de este escenario emergente. Por ejemplo, cómo las ganancias de la colaboración serán repartidas entre todos los NSPS tiene que ser acordado de antemano, y la manera de hacer dicho reparto debería fomentar a los NSPs a formar alianzas y permanecer en ellas. En otras palabras, se busca alianzas estables.

Inspirados por este escenario, proponemos un esquema completo para la venta de servicios con calidad asegurada en redes multidomino, y un método para el reparto de las ganancias que de estas ventas resulten, en el contexto de alianzas de ASes. Planteamos el problema como un problema de maximización de utilidades de la red (Network Utility Maximization problem) con requerimientos de calidad de servicio y mostramos que una solución distribuida a este problema puede ser construida.

Con respecto al reparto de las ganancias, formulamos formalmente el problema, establecemos las propiedades que un método de reparto debería tener en este contexto y mostramos por qué los métodos existentes no son apropiados. Proponemos una familia de soluciones al problema de reparto de ganancias tal que la estabilidad en términos económicos de la alianza sea garantizada. El método propuesto se basa en la resolución de una serie de problemas de optimización y estadísticas

en las ganancias de la alianza. Exploramos luego otras propiedades que este método provee.

Luego adaptamos una mirada más holística y estudiamos una interacción entre el plano de monitoreo y el comportamiento de los compradores. Proponemos un método de tarificación simple y lo estudiamos en detalle. El mismo utiliza información de monitoreo de la calidad de servicio como realimentación al plano de negocios, y tiene como objetivo mejorar la ganancia del vendedor. En el método que proponemos, los servicios de calidad garantizada son vendidos a través de subastas de primer precio, y en caso de que el servicio no alcance la calidad esperada un porcentaje de lo que se pagó por el servicio es reembolsado al comprador. Deducimos la expresión de la voluntad de pagar para los compradores y modelamos el problema de tarificación a través de un juego de *Stackelberg*. Luego resolvemos el juego para mostrar que el equilibrio del mismo, que maximiza la ganancia del vendedor, implica reembolsar 100 % en caso de fallas.

El estudio anterior supone la fuerte condición de simetría de los compradores, es decir que para derivar resultados analíticamente los compradores son modelados todos de la misma manera, lo que en particular significa que valoran el servicio a comprar de la misma forma. Con el objetivo de levantar esta hipótesis poco realista realizamos el mismo estudio con un abordaje basado en simulaciones, lo que permite sacar conclusiones sobre compradores no necesariamente simétricos. En particular mostramos que escenarios diversos arrojan resultados coherentes con los resultados analíticos. Es decir, que en los escenarios evaluados un reembolso de 100% provee más ganancia al vendedor que no establecer ninguna política de reembolso.

Palabras clave: Calidad de Servicio interdominio, alianzas, subastas de ancho de banda, reparto de ganancias, tarificación, reembolso

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Internet traffic consumption tendencies are evolving along two main axis. On the one hand, the continuous growth in terms of volume as well as in terms of Quality of Service (QoS) demanding applications, such as telepresence, video or gaming [21, 22]. On the other hand, the need for QoS connectivity from end-to-end across several domains or Autonomous Systems (ASes), which possesses political, economic and technical issues [114].

Indeed, according to recent studies [21], the applications which are envisioned to have the greatest increase are those with real time characteristics, thus those needing special quality in order to be delivered in a proper way across heterogeneous networks. In addition, emerging technologies such as telepresence or cloud computing not only generate large volumes of traffic with real time requirements, but are also used to interconnect sites around the globe. As a consequence, in addition to a QoS capable network, this kind of services require an end-to-end QoS enabled chain crossing heterogeneous carrier networks [114].

At the same time, there is a need for Network Service Providers (NSPs) to find new business cases and technology for fulfilling customer needs and maximizing revenues. Moreover, nowadays, the focus of telecommunication market is on best effort content. In order to meet customer expectations telecommunications companies are forced to invest in capacity, without getting sufficient return on these investments to have sustainable businesses. The ever evolving features provided by the handset terminals, and the growing number of connection capable equipments, constitute more evidence in favour of the forecast of Internet traffic increase.

In this scenario, current Internet business rules for domain interconnection may not be able to provide a sustainable economy for all actors in the value chain (Application Providers, Network Service Providers, etc.). Indeed, these rules (peering agreements) are not aware of the QoS capabilities of the domains and most of them are based on a traffic-symmetry assumption that may no longer be valid in evolving services (for instance HD video on demand, which intrinsically produces asymmetric traffic flows and is foreseen as one of the services that is going to grow the most [21]).

Furthermore, customer-providers agreements are based on a per-traffic consumption charging mechanism, while a common way of pricing for Internet connection to end users is a monthly flat rate. At the same time, other actors, e.g.Application Providers or the so-called Over the Top Providers (OTTs) receive revenues on a per bandwidth-consumed basis, relying their services on the existent network infrastructure but not remunerating NSPs adequately [96]. All together, there is a disagreement between money flow and traffic flow.

Even if the necessity of a QoS market from the point of view of providers has been identified, and if benefits from the point of view of the users can be expected (through a wide variety of quality assured enhanced services), the QoS market can not still be said to have succeeded. A possible issue that has been identified as the cause of this, is the fact that there is a lack of transparency between

providers and clients, driving high quality out of the market place. A similar problem has been identified decades ago, in the so-called *market for lemons* [23] phenomenon. In the connectivity market, this means that since buyers are not sure about the quality of the service they could get, they are not encouraged to pay for a differentiated service, even if quality could ultimately results satisfactory. The bad quality services leave the good ones out of the market.

Besides this economic context, technical issues related to interdomain QoS provisioning are not negligible at all. The interconnection of heterogeneous network poses technical challenges itself. Though best-effort traffic has overcome these issues, that is not the case when providing QoS. Even if several standards and protocols for Traffic Engineering have been developed such as MPLS-TE [11] and DiffServ [10], and their intedomain flavours such as the PCE hierarchy framework in [17], interdomain MPLS-TE framework [14] and its companion signalling protocol RSVP-TE [15], when it comes to interdomain each AS still has its own rules, and respecting markings, priorities or tags is not mandatory. In addition, the lack of end-to-end coordination results in an inefficient routing under congestion, with overloaded paths and underused ones at the same time. This lack of coordination stems as well from the fact that multidomain coordination is delicate, having to deal with scalability and confidentiality issues.

As already mentioned, over-provisioning has been typically the way to address Traffic Engineering in complex interdomain scenarios. However, that relied on a premise that may no longer be true: the bottleneck at access networks. With the advent of big pipes to the final user, as Fiber-to-the-Home initiatives [5] and the ubiquitous presence of connectivity-capable terminals, the validity of this assumption is seriously questionable. The context asks thus for more complex Traffic Engineering techniques enabling a smarter use of resources.

Taking into account the previous considerations many companies and academic groups are analysing future scenarios so as to meet the end-to-end requirements and business models. As a possible architecture to provide these services, the ASes alliances or federations have emerged (see for instance [4]). In this kind of interconnection market there exists a cooperation of infrastructure, policies and incentives for rational usage of resources and agreements for providing end-to-end QoS. At the same time, interesting issues arise, such as priorities and revenue sharing [124].

This collaborative paradigm aims to handle the difficulties stated above. Firstly, to provide a sustainable way to the proper development of the connectivity market, where business rules are revisited in order to meet the interests of all involved actors. Secondly, to establish trustworthy communities selling verifiable quality assured services (ASQ), so as to overcome the lack of transparency on the provided QoS driving quality goods out of market. Thirdly, to provide coordination principles in order to overcome poor traffic congestion management, technical interconnection difficulties, to allow for traffic monitoring and automatic, seamlessly quality-assured end-to-end path computation and establishment.

To sum up, the emergence of NPSs alliances, a collaborative framework where NSPs work together in order to provide end-to-end quality assured services, is likely to take place in few years to come. This new paradigm is expected to enable new business opportunities, providing a trustworthy marketplace with an increase in geographical coverage and number of clients, and making it possible to perform interdomain monitoring and to coordinate traffic routing and forwarding policies. In few words, the alliances are expected to work upon the synergy of several NSPs.

In this thesis we shall work in this context, addressing the problem from the engineering standpoint, while considering both economic and technological aspects of this emerging scenario.

1.2   Work Context: the ETICS Project

This thesis was developed in the context of the ETICS project [4]. ETICS (Economics and Technologies for Inter-Carrier Services) is a European research project, which is supported by the European Commission within the 7th Framework Program of the European Union. ETICS aims to create a new ecosystem of innovative QoS-enabled interconnection models between NSPs allowing for a fair distribution of revenue shares among all the actors of the service delivery value-chain.

To achieve these objectives, ETICS has focused on the analysis, specification and implementation of new network control, management and service plane technologies for the automated end-to-end QoS-enabled service delivery across heterogeneous carrier networks.

The participation in the ETICS project has provided us the opportunity to have rich exchanges and fruitful discussions with several industrial and academic partners.

1.3 Thesis' Contributions

This thesis has studied subjects related to the provisioning of end-to-end QoS. As aforementioned, the solutions in this regard must deal with at least both technical and economic aspects. We have assumed the working scenario of a multidomain alliance, and derived our contributions tailored to that scenario. Figure 1.1 presents in a schematic way the topics this thesis has addressed, where darker-red boxes indicate the topics on which contributions have been provided. In the lower layer, closer to the network, Traffic Engineering takes place to perform interdomain quality assured paths.
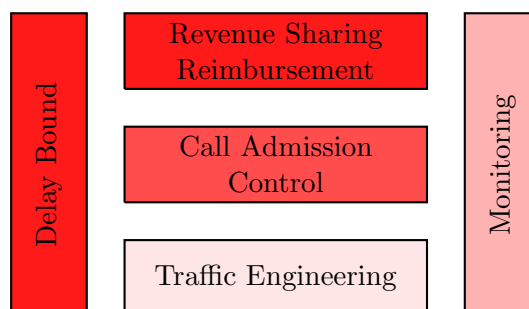
Figure 1.1: Layer diagram of the subjects addressed by the Thesis.

On top of this technical layer, the multidomain overlay alliance is built. We have worked on a call admission control mechanism, over the already established paths by the lower layer. This accounts for the bandwidth allocation mechanisms.

On top of it, and thus at a layer closer to the business plane, we develop studies related to revenue sharing within the alliance and a pricing scheme based on first price auctions and reimbursement, which interacts with the monitoring plane. The bound on the delay tool has been conceived as a transversal tool to all planes. In particular, we shall presented it as a tool aiding the Traffic Engineering mechanisms, as we shall shortly comment on. However, its use could be exploited up to the business plane. We shall now summarize each of the main contributions.

1.3.1   Part I: A tool for Interdomain Traffic Engineering

Traffic Engineering techniques for interdomain QoS have as main objective to establish end-to-end paths that can guarantee the quality needed by the service to deliver. The decisions of these techniques are mainly based on QoS metrics related to each domain, which are exchanged during the computation process. Different mechanisms have been proposed for the selection and establishment of interdomain QoS-constrained tunnels, that mainly rely on RSVP-TE [12] and the PCE architecture [8] (e.g. [17, 43, 121, 124]), or in the de facto interdomain routing protocol BGP. These mechanisms are based on metrics announced by each AS but no specification about how those metrics are computed or obtained is provided. Even if retrieving such information from an available monitoring infrastructure were possible, the announced metrics have to hold for some period of time, ideally as long as the service is provided. That is to say, measured values could not be enough since they can not be guaranteed to hold for a certain period of time.

Rather than relying on measured values, that can abruptly change if traffic anomalies arise, we propose to rely on robust bounds for such metrics. In that sense, we present a means to compute a bound on the end-to-end delay of traversing a domain, considering that the traffic varies within a given uncertainty set. This provides a robust and a verifiable quality of service value for traversing the AS, without revealing confidential information. Consequently, the bound can be safely conceived as a metric to be announced by each AS in the process of interdomain path selection.

We state the problem of computing the maximum delay for an interdomain bandwidth demand and show that it turns to be a non-convex optimization problem, thus it is not easy to solve. We reformulate the problem and propose both an exact method to solve it, and a numerical approximation method, neither of which rely on a complex monitoring infrastructure. We prove that the numerical approximating method provides a solution arbitrary close to the exact solution, while presenting lower computational cost than the exact one.

1.3.2   Part II: An Overlay Alliance

Inspired by nowadays tendencies described in Section 1.1, we propose a complete framework for selling interdomain quality assured services, and subsequently distributing revenues, in an AS alliance context. Finally, we propose to use feedback from monitoring to the business plane, through a pricing mechanism that reimburses the buyer in case of service disruption.

The problem of selling services is stated as a resource allocation problem, where the revenue of the whole alliance is maximized, while end-to-end QoS requirements are respected. This is formulated as a classical network utility maximization (NUM) problem to which we incorporate end-to-end QoS constraints. We propose as an application the use of first-price bandwidth auctions, which allows to determine the utility functions to be maximized by the NUM problem. We show that a distributed algorithm can be carried out to solve the bandwidth allocation problem.

Regarding revenue sharing, we study the case where bandwidth is allocated such that the revenue of the whole alliance is maximized, that is through the NUM problem, and we formally formulate the properties the revenue sharing method should fulfil in the context of multidomain alliances. Some of these usually sought properties are Stability, Efficiency, Fairness and Monotonicity in the resources. Stability, for instance, implies that no sub-group of NSPs within the alliance will

have economic incentives to break up the alliance. We shall argue that it is reasonable to consider this property as mandatory, since it constitutes the essence of the creation and sustainability of an alliance. Regarding Fairness, it can be conveniently defined related to some measurement of revenue produced by each member of the alliance, but its theoretical principles should be agreed on by all members of the alliance. We propose a formal definition of fairness based on the revenue contribution of each NSP to the alliance. In addition to providing the already mentioned properties, through an appropriate revenue sharing mechanism, the alliance can provide incentive to its members to improve the features they dedicate to it, for instance, capacity. This property is usually referred to as Monotonicity in the resources.

We review existing revenue sharing methods and argue on why they are not suitable to our problem. In particular, maybe one of the most famous revenue sharing methods provided by coalitional game theory, the Shapley value [138], is not adequate for our needs, since it does not guarantee one of the aforementioned properties: the stability of the alliance. We propose a family of solutions to the revenue sharing problem such that the economic stability and efficiency of the alliance in the long term is guaranteed. The proposed method is based on solving a series of Optimization Problems and considering statistics on the incomes. We as well evaluate the behaviour of the method and whether further properties are fulfilled by it through simulation studies. In particular, we propose a choice for the objective function of the aforementioned involved Optimization Problems and show through simulations that that function provides with Monotonicity and Fairness.

The dissertation closes with a proposition of a pricing scheme that makes it possible to use QoS monitoring information as feedback to the business plane, with the ultimate objective of improving the seller's revenue. In this study we assume an NSP alliance where collaboration among NSPs is tight and where a Network Performance Monitoring infrastructure is in place, i.e. the monitoring plane. In case that the monitoring plane detects that the service has failed, the buyer is reimbursed a certain, pre-announced, percentage of what he has paid for the service. Once more, we propose to sell AQS through first-price auctions, which allows both to determine the market price of the quality-assured good and to have guidelines to model the buyers' behaviour towards failures and reimbursement. In order to study this pricing scheme, we model buyers willingness to pay as a function of the quality expected by them (the service probability of failure assumed by them) and the percentage of reimbursement (announced by the seller).

Regarding the seller's standpoint, that is the alliance's interests, we assess the seller's outcome with respect to the expected revenue and find the optimal percentage of reimbursement, that is to say the percentage of reimbursement that maximizes the alliance's expected revenue. The analysis is split into two different cases, namely when buyers perform some level of service monitoring or have some knowledge about service performance (complete information), and when they do not. The latter is an asymmetric information situation, where the seller has more knowledge about the service than buyers have. We model the pricing problem in this asymmetric case through a Stackelberg game [150], where the seller initially announces a percentage of reimbursement and buyers bid according to it.

We show that when buyers are uncertain about the performance of the service they plan to buy, and assuming they act rationally, setting the percentage of reimbursement equal to 100% maximizes the seller's revenue. In addition, in that case, we show that if percentage of reimbursement is to be set to a value smaller than 100%, the *market for lemons* phenomenon appears, where the bids decrease until market disappearance. Conversely, if the percentage of reimbursement is set to a value greater than 100%, the so-called *moral hazard* behaviour is observed, where buyers take the risk of assuming a very good service performance, since in case of failure the seller would bear the costs. In both cases the seller's expected revenue diminishes. Setting the reimbursement equal to 100% overcomes these problems, and the resulting outcomes for seller and buyers are the same as when there is complete information.

Finally, we also provide a simulator that computes an approximate of the best bidding strategies when a number of the assumptions of the previous model are relaxed. The simulator makes it possible to evaluate the proposed pricing scheme in diverse scenarios, more realistic ones, where buyers are allowed to value the service on sale in a different way. A friendly graphical interface was developed, which allows to easily obtain an estimation of the best bidding strategies.

We present the application of the simulator to three different case studies and find for them the optimal percentage of reimbursement. In particular for all studied scenarios we obtain results that coincide with the theoretical ones, i.e. the seller's expected revenue when reimbursing 100% is greater than when no reimbursement is in place.

1.4   Document Structure

The remaining of this dissertation is organised as follows.

Part I is dedicated to the technical aspects related to intedomain path selection and provisioning. More precisely, Chapter 2 presents the study for computing a bound on the delay as a tool for interdomain path selection mechanisms.

Part II is devoted to the overlay alliance aspects. Chapter 3 acts as introduction to this second part, introducing the overlay NSP alliance: the working scenario for the rest of the thesis. Chapter 3 also presents the bandwidth allocation problem, whose objective is to allocate bandwidth in the overlay alliance such that the revenue of the whole alliance is maximised, while end-to-end QoS constraints are respected. In this regard, we formulate the problem as a NUM problem with per-service end-to-end constraints. We show that a distributed solution to solve that optimization problem is possible. The chapter presents as well an introduction to bandwidth auctions, and its use in the NUM problem.

Chapter 4 deals with the revenue sharing mechanism to be used within the alliances, when services are sold as in Chapter 3. The chapter discusses the properties the revenue sharing method should fulfil in order to be attractive for all the NSPs member of the alliance, from an economic point of view. Different existing methods are reviewed, which rely on the coalitional game theory literature. Finally a novel method is proposed. Tools of convex optimization are used to formulate the problem. Chapter 4 also discusses implementation considerations related to the scalability and confidentiality of the method, in the particular context of NSP alliances.

Chapter 5 and Chapter 6 close the document, by putting together ideas used throughout the thesis to consolidate a pricing scheme for AQS goods. It presents the pricing scheme based on first-price auctions and reimbursement, in case of not fulfilment of the quality. The proposed mechanism analyses buyers' willingness to pay through their bidding strategies, and interacts with the monitoring plane in order to determine if reimbursement must take place, i.e. if an SLA violation is detected. In particular, Chapter 5 is dedicated to an analytical analysis where simplifying assumptions are made. Chapter 6 in turn, relaxes some of those assumptions and presents a simulative approach. The simulative approach is applied to the study of the proposed pricing scheme in three different scenarios.

Part III is dedicated to the conclusions of the thesis and perspectives. Finally, further proofs and simulations are presented in Part IV, in Appendix A to Appendix C.

Part I

A Tool for Interdomain Quality of Service

Chapter 2

Computing a Bound on the Delay as a Tool for Traffic Engineering Techniques

2.1 Introduction

As stated in Chapter 1, there is increasing interest in value-added services, such as videoconferencing or other bandwidth-on-demand services. In this context QoS and how to guarantee it becomes a crucial issue for all involved actors, i.e. the NSP, the Service Provider (SP) and the Customer. This is especially difficult when the service traverses several domains, or ASes. In this case, QoS must be provided by all the ASes involved, which raises several technical, economic and political issues. Concerning the technical aspects, achieving scalability, preserving confidentiality and providing interoperability is paramount in any solution [157].

In this chapter we focus on point-to-point services with QoS requirements. In this case the service may be abstracted to a QoS guaranteed tunnel (for instance an MPLS tunnel [14]). The path might cross those domains through which destination is reached and whose combination of QoS parameters fulfils the service requirements.

In a NSP collaborative framework, as, for instance, the NSPs alliances, to be introduced in Chapter 3, carriers are envisioned to work together in order to achieve a common interest. In this scenario QoS values related to each domain are exchanged, and Traffic Engineering decisions are taken after them. Different mechanisms have been proposed for the selection and establishment of interdomain QoS-constrained tunnels, that mainly rely on RSVP-TE [12] and the PCE architecture [8] (e.g. [17, 43, 121, 124]). These mechanisms are based on metrics announced by each AS but they do not specify how to compute such metrics. In any case, the announced metrics must hold for some period of time, ideally as long as the service is provided. Hence, ASes would be interested in being able to provide QoS values that are guaranteed to hold for a certain period of time.

Other approaches providing methods for end-to-end QoS can be found in the literature. For instance, some propose extensions to the de facto standard interdomain routing protocol BGP (see for instance [125]). Others propose ad-hoc functions to BGP, like [93, 154]. These are based on self-adaptive methods and perform routing decisions at the edge routers level in order to maintain certain QoS parameters below some given bounds. They monitor the network state obtaining feedback which acts as an input to the self-adaptive engine. These methods are conceived to work in a pure BGP network. However, we are interested in the case of explicitly signaled tunnels, like the PCE-based mechanisms, since they are more suitable in the context we are working on. For instance, we seek a method that strictly achieves the QoS needed and not only soft QoS. In addition, for reasons explained below, we seek for a method with light dependence on monitoring.

We shall focus on services for which available bandwidth and end-to-end delay are critical parameters. The latter is composed of the sum of the delays introduced by each transit AS and the terminal ones, from source to destination. As illustrated in Fig. 2.1, where we show a situation with two terminal ASes and one transit AS, the delay in each of the ASes depends on the traffic
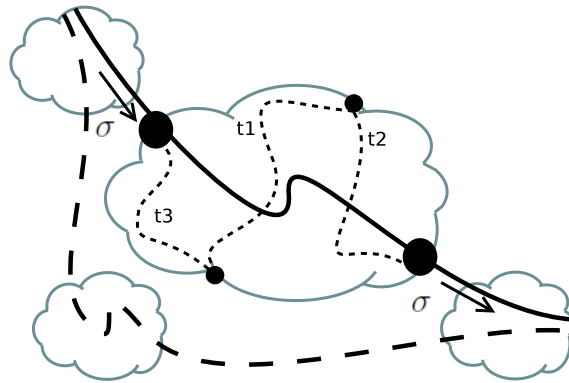
Figure 2.1: Scenario

already present in the AS (t_* flows in Fig. 2.1), the topology, the routing configuration, and the traffic coming from the new tunnel (flow σ in Fig. 2.1).

Naively, we may think that the problem of choosing the delay value to advertise can be reduced to simply advertising the current one. However, this presents two main problems, as commented in the following.

Monitoring the delay is itself a complex task. Several techniques have been proposed in the literature, mainly based on passive measurements, where some data packets are timestamped and sent to a collector (see for instance [68, 118]), or on active ones, where probe packets are sent along the network and the delay is inferred from the one experienced by the probe packets (see for instance [18, 52, 131]). These techniques present several drawbacks, just to mention the

most common of them, they usually present issues of bandwidth consumption and need for synchronization, for instance, based on specific equipment as GPS devices. Moreover, all techniques need for a monitoring architecture, which can become complex when accuracy is needed [54].

In addition, even if we were able to accurately measure the delay, the announced value, as mentioned above, should hold for a certain period of time. In this scenario the complexity is mainly due to the existence of uncertainty. This uncertainty can be classified into two types: network state uncertainty and traffic uncertainty.

Uncertainty in network state refers to the situation where the topology changes or is partially known. This may be due to information arriving out of date or not synchronized to the entity performing the computation, or simply to link failures. In the literature some approaches have been proposed for performing QoS routing under this kind of uncertainty [66, 91, 105]. However, in this chapter we assume that the topology does

not change, and considering this uncertainty is out of the scope of the present thesis.

On the other hand, we consider uncertainty in the traffic. This refers to the fact that the flows traversing the domain are not perfectly known. Knowing exactly the volumes of these flows, which we shall call Origin Destination traffic flows, requires a measurement infrastructure that is not always present, or could be expensive to implement. Techniques based on flow-level measurements, like Netflow [13] are very expensive for routers in terms of computational cost, while their sample-based version can lead to errors in traffic volume estimation. Techniques based on SNMP data considerably reduce the CPU load on routers. In that case, the measured data consists of volumes of traffic traversing the different links of the network. In order to estimate the Origin Destination traffic volumes an ill-posed linear inverse problem has to be solved, though several methods exist in the literature for doing so, for example [51, 146, 159]. Moreover, traffic uncertainty is not only related to the complexity on measuring the Origin Destination flows, but also to the fact that traffic may change rather frequently. There can be several reasons for these changes, for instance, external routing modification, the presence of unexpected events such as network equipment failures outside

the domain, large-volume network attacks or flash crowd occurrences [142].

In summary, in this chapter we aim to find a valid end-to-end QoS metric. Thus, two approaches could compete. Either we follow a dynamic approach, in which network state is continuously monitored and the metric value is updated, or we use a robust approach, in which a bound for the metric is provided. Reactive approaches make it possible to tightly follow the variations of the traffic but they require a monitoring infrastructure to be present and some sophisticated algorithms to process the measurement data. Moreover, reactive approaches are able to detect variations in the traffic demand such as abrupt changes but they are not able to forecast them [50]. On the contrary, proactive mechanisms provide pessimistic values of QoS metrics but they are able to provide metrics values which should hold for a given period of time since in that case uncertainty is taken proactively into account.

In this thesis we employ the proactive approach and consider the situation where traffic variation is the principal cause of delay variation. Thus, we shall focus on the computation of a bound for the end-to-end delay of traversing an AS through a particular path as a function of the AS parameters we mentioned before: the routing configuration, the traffic demands and the traffic injected through the new tunnel. We assume that the topology and the routing configuration are fixed. However, we consider that traffic is non-static, and that it is contained in a so-called uncertainty set [42]. The question of how to choose this set is discussed later on in this chapter.

In this context, we provide an exact method an approximate solution for solving the problem, which renders a solution arbitrarily close to the exact solution and lower computation complexity than the exact one. These solutions do not require any complex monitoring infrastructure to be deployed.

The obtained value can be afterwards advertised in the context of ASes path selection, since it is a QoS parameter bound that does not introduce confidentiality vulnerabilities. The latter refers to the fact that no topology information is delivered, just the delay of traversing the AS, where the AS is seen from the outside as a black box.

The remainder of this chapter is organized as follows. Section 2.2 introduces the assumptions and notations and formally states the problem. In Section 2.3 we show an exact solution to the problem and evaluate it through simulations. Section 2.4 presents an approximate solution with lower computational complexity than the exact one. Section 2.4.1 shows numerical results. Finally, a summary of the chapter is provided in Section 2.5. The results shown in this chapter are based on [30, 31].

2.2 Maximum Delay Problem Statement

In this section we formally present the problem of finding the maximum end-to-end delay experienced by a bounded amount of traffic traversing an AS through a particular path. As mentioned before, we will consider that traffic varies within an uncertainty set.

2.2.1 Assumptions and Notations

First, let us introduce the notations that are going to be used throughout the chapter and state some assumptions. The network is compounded of n nodes and of a set L of links, $L = \{l_1 \ldots l_{|L|}\}$, where the notation $|\cdot|$ refers to the cardinality of the set. Traffic demands are represented by the so-called traffic matrix $TM = \{tm_{i,j}\}$, where $tm_{i,j}$ is the mean amount of traffic from node i to node j. We shall use as well the term Origin Destination (OD) flows to refer to them. We reorder every traffic demand and rewrite the OD flows $(tm_{i,j})$ in vector form as t, $t = \{t_k\}$, $k = 1 \ldots n(n-1)$. The amount of traffic coming from the interdomain injected into the new tunnel will be σ.

The link load $Y = \{y_i\}_{i \in L}$ is a vector containing in the i-th entry the load on link l_i without considering σ. With these definitions we can see that $Y = R.t$ where R, a $|L| \times m$ matrix ($m = n(n-1)$), is the routing matrix, which means that $R_{i,j} = 1$ if flow j traverses link i, and 0 otherwise.

The flow that carries σ will traverse the AS from an origin to a destination node following a certain path. We will call this path P. We will equally refer to the set of links that belong to that path as P, in this case P is a subset of L.

The mean link delay is approximated by the M/M/1 model, that is to say $D_l = \frac{K}{c_l - y_l}$, where c_l is the capacity of the link l and K the mean packet size. We then obtain the delay of a path as the sum of the mean delay of the links it traverses:

$$Delay_P = \sum_{l \in P} \frac{K}{c_l - y_l}. \tag{2.1}$$

The propagation delay may be ignored in our formulation since it does not change with the load and may be added as a constant later on. Moreover, the M/M/1 model is used for illustrating the procedures towards a solution. In fact, any convex function may be used instead. For instance, the interested reader should consult [87] for a method to obtain a good convex approximation of the delay function based on measurements. The same procedures explained in this chapter should be then repeated but with the new function. We will as well ignore the constant K in the following formulations, for the sake of notations simplification.

2.2.2   Modelling Traffic Uncertainties

As mentioned above, we will not make any assumptions on the traffic matrix except that it always belongs to a certain uncertainty set. In particular we will follow the approach presented in [42] and define the uncertainty set as a polytope formed by the result of the intersection of several half-spaces. Consequently, all constraints can be written as $At \leq b$, where A is a certain matrix that can be defined according to different models, and b is a given bound. We now present four examples of polytope definition.

The Hose Model.   This particular case of the general polytope definition was presented in [57] in the context of VPN services specification. The model establishes bounds in the ingress and egress points of a network. It is suitable for the case of VPN, where the ingress and egress values are easily known, but no detailed information regarding the network is available. However, in the context of interdomain path selection, we would like to have a smaller polytope, which is obtained with more detailed information, which would allow us to have a tighter bound.

Links Capacity Model.   This model results from the application of bounds on the total traffic traversing the different links of the network. Its definition can vary from a simple static one, imposing the physical constraints, i.e. links capacity, to a more dynamic one, allowing the constraints to be obtained from historical metrics. In the latter, the constraints can be written as $R^h t \leq b$, where $b = \{b_i\}_{i \in L}$ is the vector of an historical link load and R^h is the routing matrix at the moment when the measurements were taken. This approach is used for example in [75] where a polyhedral definition of the traffic matrix is preferred to its estimation because of non stationary artefacts and estimation errors.

The Links Capacity Model with historical bounds, for instance considering the maximum observed link load, provides more detailed information than the Hose Model, along with dynamism, while it is still simple to obtain. The polytope can be frequently updated but does not require complexity for its computation.

Known Statistical Values.   If statistical values such as the mean, the variance and the covariance of link loads are known, we can compute the variance ellipsoid as $\{w = \varrho + \alpha \mid \alpha^T \Omega \alpha \leq 1\}$ where ϱ is the expected value of the link loads, and Ω its covariance matrix. Therefore, the variables w describe an ellipsoid. Several half-planes tangent to the ellipsoid can be defined in order to obtain linear constraints. Fig. 2.2 illustrates this example. The polytope can then be written as
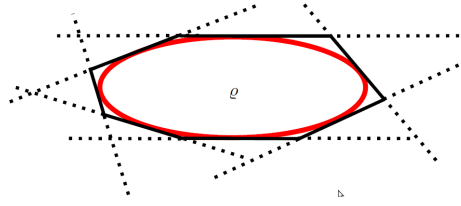
Figure 2.2: Example of a polytope definition using known statistical values.

$A \times R \times t \leq b$, were R is the routing matrix and A and b define the polytope in which the ellipsoid is inscribed.

Prediction Based Model. Yet another alternative for computing tighter polytopes are prediction based mechanisms. In this case the polytope is defined through imposing bounds on the value of traffic demands which are based on traffic prediction. The prediction of future demands is based on past observations. For example artificial intelligence methods such as neural networks or time series analysis can be used in order to forecast the future values of the traffic demand; see for example [63] for prediction based on a seasonal ARIMA model. These mechanisms provide a more dynamic polytope, which must be updated according to predictions but involves more complexity. The result is a tighter polytope that provides, in turn, a tighter bound.

The choice of the model for defining the polytope implies a trade-off between complexity and tightness of the bound. As we have shown above, simpler approaches could be used providing looser bounds, or more complex ones, needing in addition to be updated frequently, to provide tighter bounds. In the remainder of the chapter we shall use the Links Capacity Model and the Known Statistical Values one, though the solutions provided are still valid for any other model. We shall consider historical maximums for the bounds, thus measurements have to be carried out. These measurements can be performed using SNMP, which is a widely deployed protocol. Since the value needed is just the overall network interface traffic volume, we can safely assume that these values are going to be available on any AS.

2.2.3 Mathematical Formulation

For the path traversed by the new tunnel the maximum link delay is going to be computed allowing the flows t to vary within a polytope. Therefore, we will work with a maximization problem with linear constraints. In order to have a more compact notation of the problem we shall define the m-dimensional column vector w_l, $l \in P$, as $w_l = \{w_{l,i}\} = R_{l,i}/c_l$.

The optimization problem is described by Problem (2.1), where A and b define the polytope.

Problem 2.1

$$
\max_{t} \quad \sum_{l \in P} \frac{1}{c_l} \frac{1}{1 - w_l^T t - \sigma/c_l}
$$
$$
s.t. \quad At - b \leq 0.
$$

Please note that if some additional linear constraints must be taken into account they can be integrated in the definition of the polytope $At \leq b$. Example of such constraints can be $w_l^T t + \sigma/c_l < 1$, for $l \in P$, which simply states that there should be enough link capacity in order to accommodate all the traffic, including the new tunnel.

The objective function in the maximization problem defined by Problem (2.1) is not a concave function, consequently, the problem is not a convex one. On the contrary, the problem is the maximization of a convex function over a polytope. This is a very difficult problem, all the more so since the objective function is not strictly convex.

Intuitively we can see that the function is not strictly convex due to the difference between the number of links and the number of OD flows. Indeed, while the number of links grows linearly with the number of nodes in the network, the number of OD flows squares with the number of nodes in the network. This means that for different values of the vector t the objective function of Problem (2.1) can have the same value, while its gradient remains always non-negative.

More formally, we state the following proposition.

Proposition 2.1 *The function $f(t)$, objective function of Problem (2.1), is a convex function over the set $T = \{t \in \mathbb{R}^m | A \times t \leq b\}$, but not a strictly convex one.*

Proof We explore if the following inequality holds [110]

$$f(t_1) \geq f(t_2) + \nabla f(t_2)^T (t_1 - t_2),\ t_1, t_2 \in T. \tag{2.2}$$

Applying the definition of f to Equation (2.2) we obtain the following inequality for t_1, $t_2 \in T$:

$$\sum_{l \in P} \frac{1/c_l}{1 - w_l^T t_1 - \sigma/c_l} \geq \sum_{l \in P} \frac{1/c_l}{1 - w_l^T t_2 - \sigma/c_l} + \sum_{l \in P} \frac{1/c_l \times w_l^T (t_1 - t_2)}{(1 - w_l^T t_2 - \sigma/c_l)^2}. \tag{2.3}$$

Let us now define $g_l(t)$, an auxiliary function in order to simplify the notations, as

$$g_l(t) = 1 - w_l^T t - \sigma/c_l,\ t \in T. \tag{2.4}$$

Substituting the latter definition in Equation (2.3) and performing some regular math operations we obtain the following inequality

$$\sum_{l \in P} 1/c_l \frac{(g_l(t_2) - g_l(t_1))^2}{g_l(t_1) g_l(t_2)^2} \geq 0,\ t_1, t_2 \in T. \tag{2.5}$$

Each term on Inequality (2.5) is either zero or greater than zero for all t_1, $t_2 \in T$. Therefore, the function f is convex over T. It remains to show if the function is strictly convex or not. Which is equivalent to showing if there exist t_1 and $t_2 \in T$ such that $< w_l, t_2 - t_1 >$ is equal to zero for all $l \in P$, that is to say, having all vectors w_l, $l \in P$ orthogonal to the vector $(t_2 - t_1)$, or not. Since the vectors w_l do not form a basis of $\mathbb{R}^m$ it is possible to find t_1 and $t_2 \in T$ such that their difference is orthogonal to all vectors w_l, $l \in P$. $\square$

Proposition (2.1) shows that f is a convex function, but not a strictly convex one. In the following section we reformulate the problem and show a way to find its solution.

2.3 Finding the Exact Solution

2.3.1 Formulation

We now state the problem in a different way and propose a method to find the exact solution. We aim to formulate the problem in such a way that the objective function is strictly convex and the dimension of the problem is reduced. For doing so we shall decompose the vector t over a particular basis of $\mathbb{R}^m$.

The procedure consists in decomposing the vector t over the vectors w_l, $l \in P$, and their orthogonal complement. We define matrix W_1 as an m by $|P|$ matrix, whose columns are the vectors w_l, with $l \in P$, and W_2, an m by $(m - |P|)$ full rank matrix such that it verifies

$$W_1^T W_2 = 0. \tag{2.6}$$

In other words, the columns of W_2 form a basis of the space orthogonal to the one spanned by the columns of W_1.

Provided that the columns of W_1 are as well linearly independent, the columns of the matrix W defined after W_1 and W_2 as

$$W = [W_1 W_2] = [w_1, \ldots, w_l, \ldots, w_{|P|}, \ldots w_m] \tag{2.7}$$

represent a basis of $\mathbb{R}^m$.

We shall decompose vector t over the defined basis using the auxiliary variables $z \in \mathbb{R}^{|P|}$ and $h \in \mathbb{R}^{m-|P|}$ as

$$t = W_1 z + W_2 h. \tag{2.8}$$

Multiplying both sides of Equation (2.8) by w_l^T, and using Equation (2.6) we obtain

$$w_l^T t = w_l^T W_1 z = \nu_l^T z, \tag{2.9}$$

where we have defined $\nu_l^T = w_l^T W_1$, for all $l \in P$. Note that both ν_l and z are column vectors of dimension $|P|$.

Equation (2.9) will directly lead us to rewriting the objective function of Problem (2.1) as a function of z. We shall now redefine the polytope by writing it in the basis W. For doing so, the change of variables defined by Equation (2.8) needs to be done in the constraints of Problem (2.1). This leads to defining a new matrix denoted D and computed as AW. The polytope over the new basis can be compactly written as $D[z^T \ h^T]^T \le b$.

All in all, Problem (2.1) can be rewritten in the form of Problem (2.2). Please note that the objective function depends only on the variable z.

Problem 2.2

$$\max_z \quad \sum_{l \in P} 1/c_l \frac{1}{1 - v_l^T z - \sigma/c_l}$$

$$s.t \quad D\begin{pmatrix} z \\ h \end{pmatrix} \le b.$$

Let us call the objective function of Problem (2.2) as $J(z)$ and the new polytope as H (i.e. $H = \{[z^T \ h^T]^T \in \mathbb{R}^m : D[z^T \ h^T]^T \le b\}$). Let us as well define the polytope H_z as

$$H_z = \{z \in \mathbb{R}^{|P|} \mid \exists \, h \in \mathbb{R}^{m-|P|} : D[z^T \ h^T]^T \le b\}. \tag{2.10}$$

Let $\mathcal{W}_1 = span\{w_1 \ldots w_{|P|}\}$, where span refers to the set of all linear combinations of vectors $w_1 \ldots w_{|P|}$. Clearly H_z is the projection of H onto $\mathcal{W}_1$.

Since H is a convex polytope by definition, it is easy to check that H_z is also a convex polytope. More precisely, H_z is the convex hull of the projection of the extreme points of H onto $\mathcal{W}_1$ [49].

Then, since $J(z)$ does not depend on h, Problem (2.2) can be represented in the space $\mathcal{W}_1$ as follows:

Problem 2.3

$$\max_z \quad J(z)$$

$$s.t. \quad z \in H_z.$$

The following statement summarizes our development of the problem.

Proposition 2.2 *The optimization problem defined by Problem* (2.1) *is equivalent to the one defined by Problem* (2.3).

We now show that $J(z)$ is a strictly convex function over H_z, which will in turn allow us to prove that the solution of Problem (2.3) is attained at an extreme point of the polytope H_z.

Proposition 2.3 *The objective function of Problem* (2.2), *$J(z)$, is a strictly convex function over the set H_z defined as in Equation* (2.10).

Proof We define $\lambda_l(z)$ as

$$\lambda_l(z) = (1 - \nu_l^T z - \sigma/c_l)^{-2}, \ \forall l \in P \tag{2.11}$$

and matrix Λ as

$$\Lambda(z) = diag(\lambda_1, \ldots, \lambda_{|P|}). \tag{2.12}$$

For all $z \in H_z$ and $l \in \{1 \ldots |P|\}$, $\lambda_l(z) > 0$. Thus, $\Lambda(z)$ is a positive-definite matrix[1].

In addition, we can check that $[\nu_1 \ldots \nu_{|P|}] = W_1^T W_1$ is also a positive-definite matrix. Thus, the Hessian of $J(z)$, which is

$$\nabla^2 J(z) = (W_1^T W_1)\Lambda(z)(W_1^T W_1) \tag{2.13}$$

is as well a positive-definite matrix. $\qquad\square$

We are now able to show that the solution to Problem (2.3) is attained at an extreme point of H_z.

Theorem 2.1 *The solution to Problem* (2.3) *is attained at an extreme point of the polytope H_z, defined by Equation* (2.10).

Proof We prove by contradiction that the maximum of $J(z)$ over H_z must be reached at an extreme point of H_z. Since, by Proposition (2.3), J is a strictly convex function, the following inequality holds [110]:

$$J(h_1) > J(h_2) + \nabla J(h_2)^T (h_1 - h_2), \ \forall \ h_2, h_1 \ \in H_z. \tag{2.14}$$

Now, let $\bar{h}_2 \in H_z$ be an optimal point of Problem (2.3). Therefore, $\bar{h}_2$ is a strict maximum, since J is strictly convex, and, for all $h_1 \in H_z \setminus \{\bar{h}_2\}$, we must have:

$$J(h_1) - J(\bar{h}_2) < 0. \tag{2.15}$$

Together with Inequality (2.14), we get

$$\nabla J(\bar{h}_2)^T (h_1 - \bar{h}_2) < 0, \ \forall h_1 \in H_z \setminus \{\bar{h}_2\}. \tag{2.16}$$

By contradiction we suppose that $\bar{h}_2$ is not an extreme point of H_z. Then there exists $\mu \in \mathbb{R}^{|P|}$ such that $||\mu|| > 0$ and $\bar{h}_2 + \mu, \ \bar{h}_2 - \mu \in H_z$. By letting $h_1 = \bar{h}_2 - \mu$ and $h_1 = \bar{h}_2 + \mu$ at a time, we would get:

$$\nabla J(\bar{h}_2)^T \mu < 0 \ \text{and} \ -\nabla J(\bar{h}_2)^T \mu < 0, \tag{2.17}$$

which is not possible. $\qquad\square$

Problem (2.3) allows us to work with a strictly convex function, and to reduce the dimension of the feasible region, in some cases, considerably. In order to find the solution, we need to be able to perform the projection of a polytope, and afterwards enumerate its extreme points. Methods for doing so are available (see for instance [76]), although these can be computationally expensive tasks. In the following subsection we explore this solution by performing simulations in a real topology.

2.3.2   Simulations

In order to assess the proposed method we now present some simulation studies. The simulations are carried out using two different research networks. Namely, the Abilene network, whose topology, historical traffic demands and routing matrix are available from [158], and the GÉANT network [20]. In order to compute the polytope projection and enumerating its extreme points we use the MPT library [7] and the ET library [86], distributed along with the former.

---

[1] A $n \times n$ real symmetric matrix M is positive-definite if $z^T M z > 0$ for all non-zero vectors z, $z \in \mathbb{R}^n$.

2.3.2.A   The Abilene Network

The Abilene network consists of 30 internal links and 12 routers, all exchanging traffic among them. Fig. 2.3 shows a traffic trace of the Abilene Network. This example shows how the traffic matrix is prone to sudden traffic variations. Fig. 2.3a shows the traffic for some OD flows corresponding to 2016 consecutive measurements (where each color corresponds to one OD flow), while Fig. 2.3b shows the link load (each color corresponds to the load of a certain link).

For illustrative purposes we compute results for three different types of services. Namely, a VoIP service with 1 Mbps of bandwidth, a broadcast quality HDTV service with 19.4 Mbps and a VPN service with a demand of 270 Mbps. We compute the maximum delay suffered by a flow traversing the AS through a particular path and carrying each one of these services at a time. The path is chosen arbitrarily, from one origin to one destination node. This choice and its impact on the delay are out of the scope of the present thesis.

In the first place, we define the polytope using the Links Load model introduced in Subsection 2.2.2. That is to say, the polytope is defined by imposing bounds on each link load, which are based on the maximum values obtained historically.

The values obtained for the defined path and the three services are shown in Fig.2.4a (dotted line) along with the current delay value. The current delay value corresponds to a value obtained instantaneously. For this particular case the maximum delay value is approximately 3 times more than the current one which illustrates the weakness of the current value as a metric on which rely. We will come back to this kind of comparisons later on this section.

In the second simulation, we define the polytope based on the Known Statistical Values model, introduced in Subsection 2.2.2. We compute the variance ellipsoid using a historical traffic trace (the same trace used for the first simulation) and we approximate the ellipsoid by a polytope, by intersecting several half spaces tangent to it. The maximum delay of traversing the AS is computed for the same path used in the previous simulation.

The results are shown in Fig. 2.4a (dashed line) for a flow traversing the same path as in the previous simulation and carrying the three defined services, one at a time. We can see that in this case the bound obtained is smaller than the one obtained in the first place and closer to the instantaneous value.

We now compare the two bounds with the real delay suffered by the path during the two weeks after the computation of the polytopes, in all the cases assuming an interdomain bandwidth demand of 1 Mbps. The results are shown in Fig. 2.4b which illustrates the behaviour of the bounds with respect to the real values. We can see that there is a trade-off between assuring a delay value for most of the time, by using a big polytope, or having a tighter bound most of the time, but having delays that outstrip the bound. Nevertheless, the polytope could be reduced in a safe way if we had additional information, for example by using as well the hose model which imposes bounds to the traffic coming from other clients, which may be limited by a contract and traffic shaping.

These results were computed on a recent machine with good computational power (two processors Intel Xeon X5660 2.80GHz, 24GB of RAM). The time consumed to perform the computation of all demands for one path varied between 4 and 38 minutes, which for a moderately sized network is rather high. In fact, even if in several topologies we were able to find the exact solution through these means, it is still an open question whether there exists an algorithm for enumerating all extreme points of a polytope of an arbitrary dimension in polynomial running time [80]. Through the following simulations we empirically explore the time consumed by the method in a larger network.
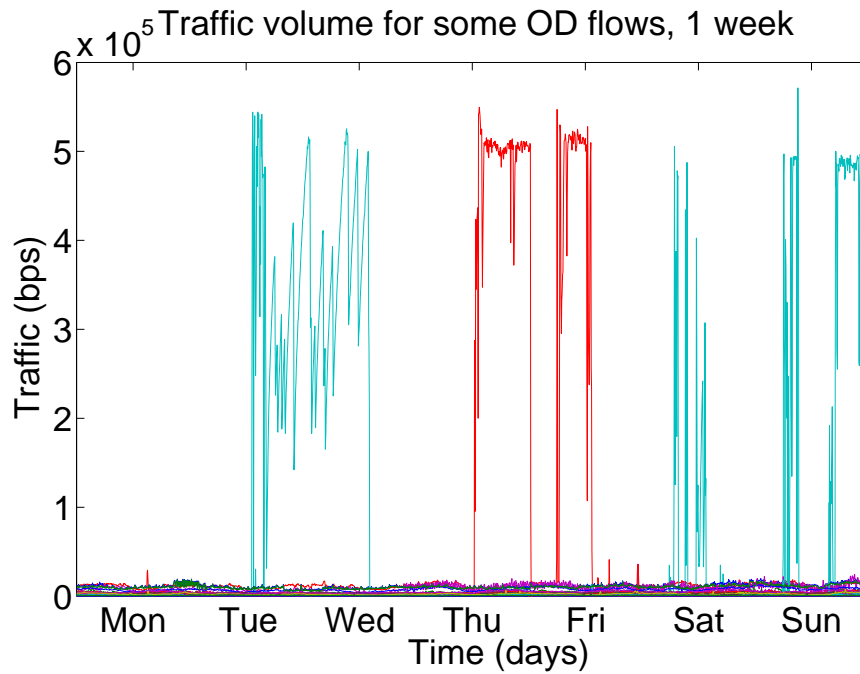
(a) Traffic volume per OD flow.
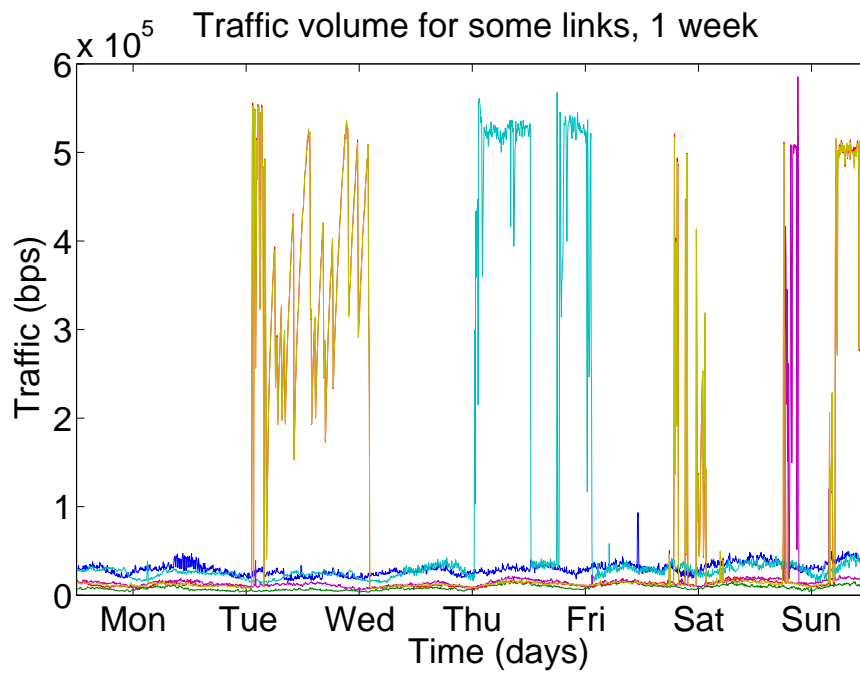
(b) Link load.

Figure 2.3: Example of traffic variation in the Abilene network, one week of traffic.
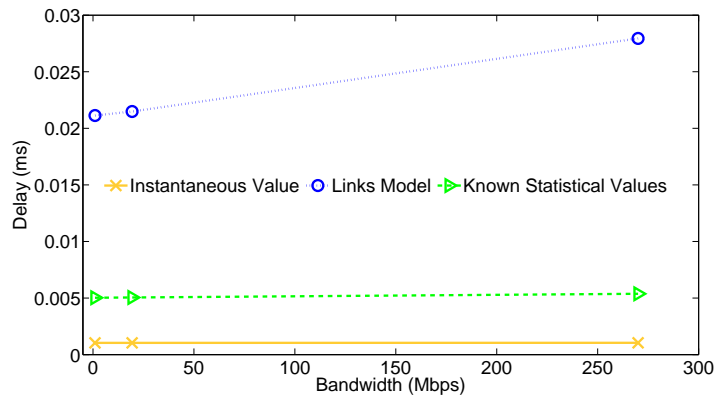
(a) Instantaneous delay value and Maximum delay value for two different polytopes and three bandwidth demands.
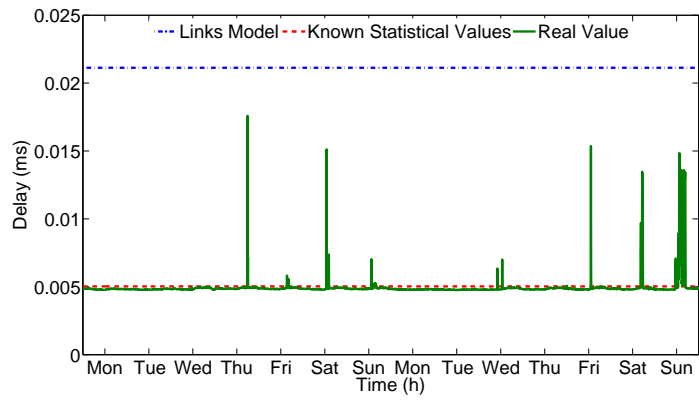
(b) Maximum value for two different polytopes and real values during two weeks for one bandwidth demand.

Figure 2.4: Simulations in the Abilene network.

2.3.2.B   The GÉANT Network

In order to test the proposed solution on a larger topology, we use the GÉANT network. This network is compounded of 23 nodes and 74 links. Thus, we can define up to 506 independent OD flows. As we have already mentioned, the computation complexity of the proposed solution is likely to grow with the dimension of the network (i.e. the number of links in the path and the number of OD flows in the network). The simulations with this network aid as to assessing the performance of the method when the number of OD flows grows. We perform the simulations considering several subsets of OD flows, containing each of them 170, 200, 230 and 260 OD flows. The polytope is defined using the Links Load model and historical data. Results in this case were computed on a regular computer (Intel Pentium Dual-core 1.86GHz, 2GB of RAM).

Fig. 2.5 shows the time consumed by each phase of the procedure, that is to say obtaining the polytope in the new basis, projecting the polytope and finding its extreme points. We can see that in all the cases, when we increase the number of OD flows considered, the task that consumes most of the time is the projection of the polytope.

The procedure has shown rather high computational times, though it was still feasible in all the tests. It is because of this that we think of this method as of great aid when developing approximate, but less time consuming, methods, since it provides the ground truth, thus a validation tool for such methods. In the following section, we present an approximate method that can be used as an alternative to the exact one introduced in this section when its computational time becomes excessive.
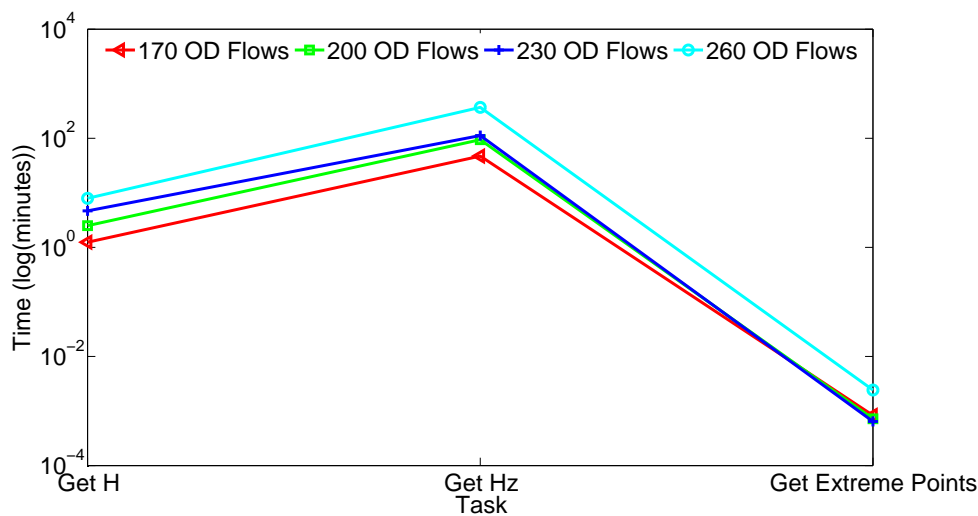
Figure 2.5: Computation time as a function of the number of OD flows considered on the GÉANT Network.

2.4 Finding an Approximate Solution

In Section 2.3 we have presented a method that allows to find the exact solution. Nevertheless, we have pointed out that its complexity remains open. In this sense, we now present a method that provides an approximate solution to Problem (2.1), while reducing the computational time. More precisely, we present a numerical method based on the approximation of the objective function by a piecewise linear function. This method provides a value that is arbitrarily close to the exact solution (up to some controlled error).

Let us introduce the method with a detailed description of the procedure to design it. First of all, we transform each link's delay function, $\frac{K}{c_l}\frac{1}{1-y_l/c_l-\sigma/c_l}$, into a piecewise linear function over y_l. For this, we partition each function's domain into η_l subintervals and approximate the function in each subinterval by its first order Taylor polynomial. We shall note the subintervals of link l as $\Delta_{l,j}$, $j = 1 \ldots \eta_l$.

Secondly, we obtain the delay of the path, as before, by summing up the delay on each link belonging to it. Therefore, we obtain a maximization problem similar to Problem (2.1) but now with a piecewise linear objective function. Let us use the indicator function, defined as

$$\mathbb{1}_\Delta(z) = \begin{cases} 1 & \text{if } z \in \Delta \\ 0 & \text{otherwise.} \end{cases} \tag{2.18}$$

Then, the new problem can be written as in Problem (2.4), where α and β are taken from the Taylor polynomial of the original function.

Problem 2.4

$$\max_t \quad \sum_{l \in P} \sum_{j=1}^{\eta_l} \left(\alpha_{l,j} w_l^T t + \beta_{l,j} \right) \mathbb{1}_{\Delta_{l,j}}(w_l^T t)$$
$$s.t. \quad At - b \leq 0.$$

The next step is to redefine Problem (2.4) in order eliminate the indicator function and to obtain a linear objective function. To do so, we decompose Problem (2.4) into $\prod_{l \in P} \eta_l$ problems, each of them having a linear function as objective one. This linear function stems from the consideration of one of the linear functions that compound each link delay's approximation, and summing them up. Let us now use the index $j(l)$, $j(l) = 1 \ldots \eta_l$, for all $l \in P$, to denote the linear function chosen for link l, corresponding to subinterval $\Delta_{l,j}$. In order to consider each linear function only on the corresponding domain we introduce new constraints to the problem. That is to say, the solution has to be restricted to belong to the original polytope and at the same time to the set $\{t \in \mathbb{R}^m | w_l^T t \in \Delta_{l,j}\}$. It can be readily proved that this set is equivalent to imposing restrictions on the load on each link. Thus, it is itself a polytope. We represent the intersection of the original polytope and the new one, which is also a polytope, in the matrix form as the set $\{t \in \mathbb{R}^m | A^* t \leq b^*\}$, where the matrix A^* and the vector b^* define the intersection polytope.

Finally, in order to have a problem equivalent to Problem (2.4) we consider all combinations of linear functions for each link, find the maximum over t for each combination and keep the combination which leads to the greatest value of the objective.

The mathematical formulation of the equivalent problem can be seen in Problem (2.5), where the maximum on $j(1), j(2), \ldots, j(|P|)$ means that we consider the maximum obtained when we let each value $j(l)$ vary between 1 and η_l.

Problem 2.5

$$\max_{j(1),j(2)\ldots j(|P|)} \begin{cases} \max_t & \sum_{l \in P} \alpha_{j(l)} w_l^T t + \beta_{j(l)} \\ s.t. & A^* t - b^* \leq 0. \end{cases}$$

As we have claimed above, this method leads to a solution that is arbitrary close to the exact solution of the original problem (Problem (2.1)). The following theorem proves such statement.

Theorem 2.2 *The solution to Problem* (2.5) *provides a solution that can be made arbitrarily close to the solution to Problem* (2.1).

Proof Let us call the original function, i.e. the objective function of Problem (2.1), as $f(t)$ and the piecewise linear approximation of f as $\tilde{f}$. Let $\tilde{t_f}$ and t_f be the values at which the maximum of $\tilde{f}$ and f are attained respectively. Let us as well note the feasible region of Problem (2.1) as T.

We set the hypothesis that for a given real positive ϵ, the approximation of f can be made such that the difference between f and $\tilde{f}$ is bounded by ϵ. That is to say that

$$f(t) - \tilde{f}(t) \leq \epsilon \ \forall t \in T. \tag{2.20}$$

Under the conditions of Equation (2.20) and with the definitions of $\tilde{t_f}$ and t_f provided above we can prove that

$$f(t_f) - f(\tilde{t_f}) \leq \epsilon. \tag{2.21}$$

Let us first prove that Equation (2.20) holds for the case of the M/M/1 model mean delay function. Please note that for other functions this is an hypothesis to be checked before applying the algorithm. We provide a constructive proof showed in the following. For the sake of simplicity on the notations we will not include σ in the formulation, but the whole procedure can be reproduced in an analogous way considering σ.

Let $e(y) = \frac{1}{1-y}$ be such that $f(t) = \sum_{l \in P} \frac{1}{c_l} e(w_l^T t)$, $t \in T$. Let us note the partition of the domain of e over y ($[0, 1)$) as the set of subsets Δ_i where

$$\Delta = \{\Delta_i : i = 1 \ldots \eta\}. \tag{2.22}$$

Let $\tilde{e}$ be the piecewise linear approximation of e over each one of the subsets defined in (2.22), such that the following inequality holds

$$|e(y) - \tilde{e}(y)| \leq \frac{\epsilon}{\sum_{l \in P} 1/c_l} = \delta \ \forall y \in \Delta. \tag{2.23}$$

This will ensure that Equation (2.20) holds since $\tilde{f}(t) = \sum_{l \in P} 1/c_l \tilde{e}(w_l^T t)$.

Consider the graphic displayed in Fig. 2.6. Let us note Δ_i as $\Delta_i = [y_{i-1}, y_i]$, with $y_0 = 0$. We define $z_i \in \Delta_i$ as the linearization point of function e in Δ_i. Let us define ϵ_{i-1} and ϵ_i as the difference between e and $\tilde{e}$ at each y_{i-1} and y_i respectively. That is to say

$$\epsilon_{i-1} = e(y_{i-1}) - \tilde{e}(y_{i-1}) \text{ and } \epsilon_i = e(y_i) - \tilde{e}(y_i), i = 1 \ldots \eta. \tag{2.24}$$

It is not difficult to see that the maximum of the difference between e and $\tilde{e}$ will be attained at either y_{i-1} or y_i, $i \in 1 \ldots \eta$.

Given y_{i-1} fixed, let z_i increase from y_{i-1}. We define y_i such that $\epsilon_i = \delta$. As z_i increases, y_i and ϵ_{i-1} increase. Therefore, the maximum subinterval size under the constraint $\epsilon_{i-1}, \epsilon_i \leq \delta$ is achieved when

$$\epsilon_{i-1} = \epsilon_i = \delta. \tag{2.25}$$

Hence, given y_{i-1} and δ, we can find a value z_i imposing that

$$e(y_{i-1}) - \tilde{e}(y_{i-1}) = \delta. \tag{2.26}$$

Once z_i is known, we can compute y_i by imposing

$$e(y_i) - \tilde{e}(y_i) = \delta. \tag{2.27}$$

All in all, it appears that e can be approximated over its domain by means of a piecewise linear function. Thus, $f(t) = \sum_{l \in P} 1/c_l e(w_l^T t)$ can be approximated by means of a sum of piecewise
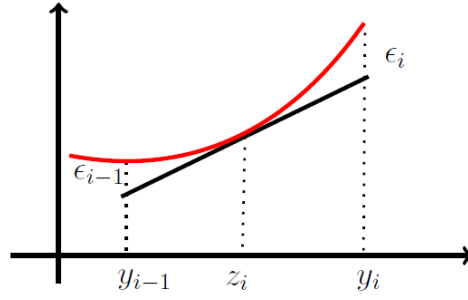
Figure 2.6: Difference between function e and its approximation $\tilde{e}$.

linear functions, and this approximation, which we note as $\tilde{f}(t)$, is such that Equation (2.20) holds for all $t \in T$.

We have provided a constructive proof of Equation (2.20). We are now able to show that Equation (2.21) holds. Indeed, let again $\tilde{t_f}$ and t_f be the values at which the maximum of $\tilde{f}$ and f are attained respectively. The following inequalities are obtained straightforwardly from the definition of the maximum

$$\tilde{f}(\tilde{t_f}) \geq \tilde{f}(t_f) \tag{2.28}$$

$$f(t_f) \geq f(\tilde{t_f}). \tag{2.29}$$

We are interested in finding a bound to the difference $f(t_f) - f(\tilde{t_f})$ which can be rewritten as in the following equation

$$f(t_f) - f(\tilde{t_f}) = \overbrace{\tilde{f}(\tilde{t_f}) - f(\tilde{t_f})}^{\leq 0} + \overbrace{\tilde{f}(t_f) - \tilde{f}(\tilde{t_f})}^{\leq 0} + \overbrace{f(t_f) - \tilde{f}(t_f)}^{\leq \epsilon}. \tag{2.30}$$

Equation (2.30) immediately leads to the inequality $f(t_f) - f(\tilde{t_f}) \leq \epsilon$, which completes the proof. Please note that in Equation (2.30) we have used the fact that $f(t)$ is greater than $\tilde{f}(t)$ for all $t \in T$, which is true since $\tilde{f}$ is the piecewise linear approximations of a convex function. □

Problem (2.5) can be solved computationally by performing a loop of $\prod_{l \in P} \eta_l$ iterations. Please note that the problem solved on each iteration is a linear one, which is very easy to solve.

In order to obtain the partition needed to define the piecewise linear function, we propose to iteratively compute the subintervals such that within each of them the maximum difference between the approximate function and the original one is a given ϵ, at most. This constructive procedure corresponds to the one shown in the proof of Theorem (2.2).

In order to reduce the number of iterations, we pre-compute the maximum value that the load can achieve at each link according to the constraints imposed by the polytope. Table 2.1 shows the number of subintervals needed to define the piecewise linear function, for different percentage errors and maximum link utilization (LU). This gives an idea of the complexity of the procedure. For example, for a 6-link length path, at most $4^6 \approx 4000$ linear problems need to be resolved for obtaining a result with 10% of error, thus the numerical complexity is still feasible.

	Error (ϵ)	
LU	1 %	10 %
60%	5	2
80%	9	3
90%	12	4

Table 2.1: Number of subintervals needed to define the piecewise linear function.

2.4.1   Numerical Results

In order to assess the results of the numerical approximation method we shall first use the Abilene network topology as before. We will as well use one of the polytopes used before, so as to be able to compare results afterwards. In particular we shall use the polytope computed after the links model. As optimization software we use CPLEX [6]. We shall secondly perform further simulation studies on the GÉANT network topology, so as to be able to have more information about computation time. All results shown in this subsection are obtained using the same machine used for computing the results over the Abilene network in Subsection 2.3.2 (two processors Intel Xeon X5660 2.80GHz, 24GB of RAM).

We compute the maximum delay for the four paths and three interdomain demands used in Subsection 2.3.2, using the numerical approximation method. The results along with the exact solutions are displayed in Fig. 2.7, where the bars indicate the maximum error (10% in this case). Overall, the computation of each of the aforementioned values takes in mean 2.28 seconds. The maximum link utilization (imposed by the topology) was between 30% and 80%. These computation times are dramatically smaller than the ones necessary for obtaining the exact solution (approximately a 1000× decrease) while providing a very tight bound.

The previous simulations allowed us to validate the approximate method and show that its computational time is much smaller than the one obtained through the exact method. We shall now explore this computational time when varying different parameters of the problem, namely the error, the maximum links' load and the number of links in the path.

The computation time depends on the accuracy needed by the application, which is not established a priori since it is a decision to be taken by each AS. It depends, in addition, on the
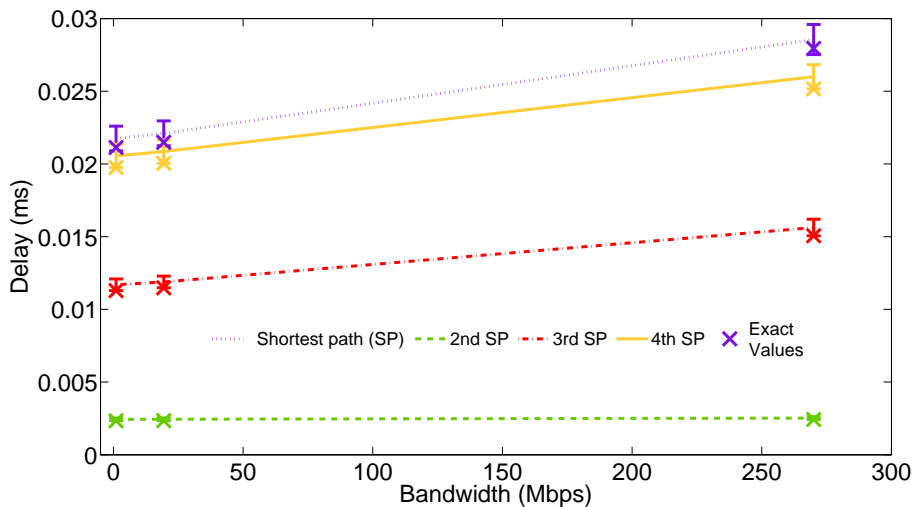
Figure 2.7: Approximate and Exact Solution: Maximum delay for one Origin-Destination pair for different Bandwidth demands and paths.
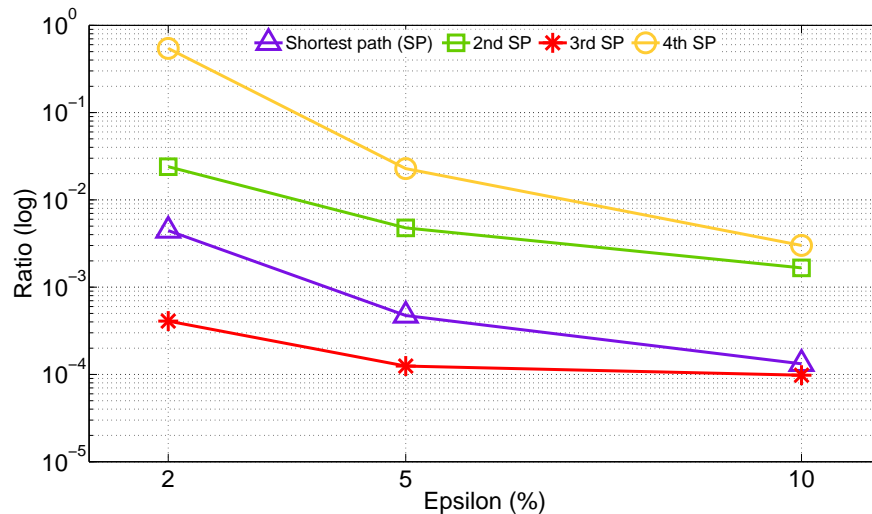
Figure 2.8: Ratio of the time consumed by the Approximate Method to the Exact Method: one interdomain demand for different allowed error and maximum link utilization between 80 and 90%. Bound computed for four different paths over the Abilene network.

maximum link utilizations allowed by the polytope along with the topology, and on the number of links in the path.

In order to asses the impact of the accuracy in the time consumed by the procedure we repeat the simulations allowing a maximum error of 5% and of 2%. The mean time needed for computing the maximum delay over one path for one interdomain demand is of 1 and 7.9 seconds, for obtaining a solution within 5% and 2% of error respectively, which implies approximately a 90× decrease for the 2% error, with respect to the exact method.

The previous results were obtained using historical traffic demands over the Abilene Network. This implies that the maximum link load, imposed by the polytope, is between 30 and 80%, as mentioned before. We now present further simulation results using synthetic data to define the polytope, so as to obtain results on a scenario with higher maximum link utilizations.

Fig. 2.8 shows the ratio of the time consumed for computing the delay bound through the Approximate Method to the time consumed by the Exact Method. The bound was computed for one interdomain demand of 19.4 Mbps for a polytope imposing maximum link utilizations between 80 an 90% and three different values of errors (i.e. ϵ). This was repeated for the same four different paths presented above. Results of these simulations show that the time consumed by the Approximate Method is much less than the one consumed by the Exact Method. In the worst case, that is, $\epsilon = 2\%$ and the 4th shortest path, the time consumed by the approximate method is of 19 minutes while the time consumed by the Exact Method is of 38 minutes.

Finally, we explore the influence on the computation time of the number of links in the path. For doing so, we use the GÉANT network, whose topology is larger than the Abilene's topology. Fig. 2.9 shows the computational time for one interdomain demand of 19.4 Mbps from an origin to a destination node, through different paths. Results are presented for a link utilization between 20% and 90% and different values of the allowed error. Is it worth clarifying that the topology does not allow to have high link utilizations in all links at the same time, since there are a number of bottleneck links on it. Results show that the computation time is not very sensitive to the path-length. For the case of a 11-link path, which greatly exceeds the maximum path length on a domain, is of approximately 7 minutes, providing a value within the 2% of error.

The Approximate Method has shown, through extensive simulation studies, to present low computation times in most of the cases. The methods proposed in this chapter are conceived to be
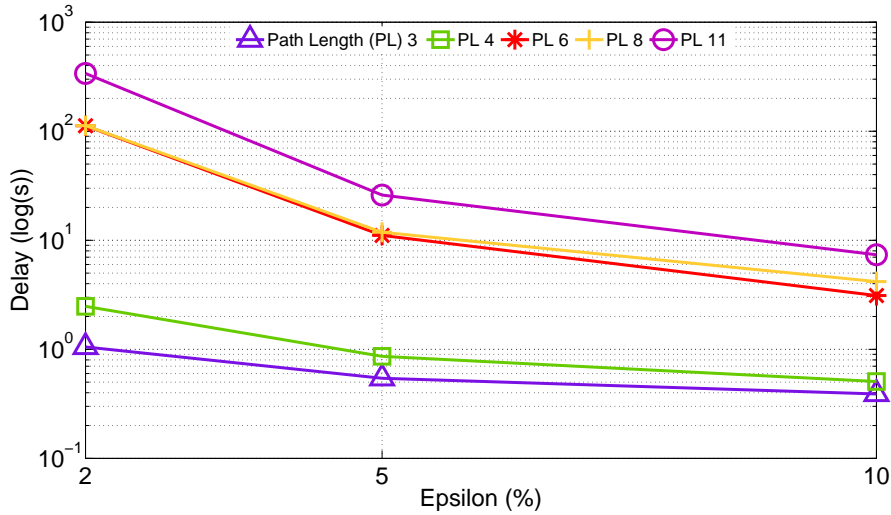
Figure 2.9: Time consumed by the Approximate Method: one interdomain demand for different errors and maximum link utilization between 20 and 90%. Bound computed for different paths over the GÉANT network.

used by each AS in order to obtain a value of a metric to announce in the process of Interdomain Path Selection. The announced bound is supposed to hold for a long period of time, for instance, several hours. In this context, the time consumed by the Approximate Method is considered totally acceptable. However, we have not focused our work on optimizing the computation time, which could be, for instance, diminished through parallelizing the code, since its nature allows it (it solves an optimization problem over several independent feasible regions).

2.5 Summary

In this chapter we have presented a means to compute a bound on the end-to-end delay. The method considers uncertainties in the traffic traversing the AS, which have been modelled as a polytope. Therefore, it is a value that the AS can guarantee for a certain period of time. The problem was mathematically stated and different solutions were provided. A method for finding the exact solution was given, and an alternative approximate method was proposed, as a remedy for the high computational cost of the former. Such approximate method renders rather than a value of the delay, an interval to where the exact maximum delay is guaranteed to belong. The latter was theoretically proven and numerically validated. Both methods were tested over real topologies using measurement and synthetic data. The Approximate method was shown, through simulations, to provide acceptable computation time on several scenarios. All together, we have proposed a method to enhance PCE-based interdomain path selection mechanisms, which can be implemented with low computation complexity and little monitoring infrastructure.

Part II

An Overlay Alliance

Chapter 3

The Alliance Model and Bandwidth Allocation

3.1 Introduction

This chapter acts as introduction to the remaining of the document. Its objective is twofold, we introduce the NSP alliance model, which is the framework used throughout the remaining of the dissertation, and the bandwidth allocation mechanism, used in order to sell enhanced services on top of the alliance infrastructure. In NSPs alliances, NSPs are no longer competitors but they work together in a collaborative way. For the alliance to make economic and technical sense, several aspects need to be agreed on. In particular, coordination principles regarding bandwidth allocation must be put in practice, revenue or penalty sharing must be in place as well as policies assuring QoS.

In our framework, the alliance sells end-to-end quality assured bit pipes to intermediate actors like brokers or Over-the-top providers (OTT), which will in turn resell them to the final user, by providing their own services through a quality assured path. We propose a selling mechanism that is conceived to allocate bandwidth so that the revenue of the whole alliance is maximized, while fulfilling end-to-end QoS requirements. The mechanism is based on solving a problem that has its roots on a Network Utility Maximization (NUM) problem, widely used in network congestion control.

We shall introduce as a particular use case of the proposed allocation mechanism, a bandwidth-auctions based mechanism, which uses first-price auctions as a means of discovering the buyers willingness to pay for the services on sale. This wilingness to pay will determine the utility functions to be maximized. A similar mechanism has been proposed in [41], where the classical NUM problem is solved to allocate bandwidth on a network where utilities come from the offered bids. We shall extend this application to the interdomain case and consider in addition end-to-end QoS constraints. This original modification of the NUM problem is one of the main contributions of this chapter.

In addition, our aim is to be able to solve the allocation problem in a distributed way. We shall argue that at the allocation stage a decentralized mechanism is needed, since, among other reasons, this stage is likely to occur at a time-scale where centralization is not scalable. The classical NUM problem with separable objective functions and capacity constraints is easily shown to be solved in a decentralized way. However, when end-to-end delay constraints are introduced, a priori it is not straightforward to decouple the problem. Nevertheless, in this chapter we shall show that even when including end-to-end QoS constraints a decentralized solution can be found, which constitutes the second major contribution of the chapter.

The chapter is organized as follows. Section 3.2 introduces in more detail the NSP alliances, its motivation and general characteristics. Section 3.3 provides the mathematical model of such alliances. In Section 3.4 we propose our approach for interdomain bandwidth allocation with QoS

constraints. Section 3.5 is dedicated to the particular application based on first-price auctions, presents a brief introduction to auctions mechanism and reviews previous work concerning network bandwidth auctions. Section 3.6 shows the results of simulative studies of the bandwidth allocation mechanism and comments on implementation considerations. Finally, Section 3.8 summarises the chapter. The results shown in this chapter are based on [27].

3.2   Mutidomain Alliances

In different contexts where services must be provided throughout several geographical regions, with high availability and assured quality, collaboration arises as a viable way to reach the requirements of high-demanding clients without the need of ubiquitous deployment of redundant infrastructure. This statement is valid in a wide set of cases, beyond the Internet connectivity provisioning one. Indeed, it happens in the legacy telephony network, in cloud computing clusters, or even in railway companies and airlines. Consider as an example the case of railway companies. Such companies would not deploy a network to reach all destinations, neither would they constrain their service offer to their own country or area. On the contrary, companies work together sharing their networks. This collaboration ranges from merely network interconnection to common coordination and cooperation principles. The former occurs already in the Internet, the network of networks, for best-effort traffic, while examples of the latter are airline alliances. Of course, different entities would be willing to collaborate as long as the cost of their interconnection does not exceed the revenues due to the cooperation.

Coming back to NSPs, in addition to benefits such as, for instance, services' offer diversification, increase of the regions reached by those services, increase of the available resources and back-up paths, one of the main motivations for an alliance is the increase of the revenue of the NSPs that belong to it. This revenue increase comes from the fact that given a collaborative environment, services can be sold (or what is the same in this context, resources can be allocated) so as to maximize the revenue of the alliance as a whole, and moreover, possibly increase it with respect to what all its members would earn by acting by themselves.

In one word, synergy is one of the main attractions for the creation of an alliance, though it must be noticed that it is not achievable by the joint effort *per se*. Indeed, in order to provide higher revenues, coordination principles must be set in place, so as to allow for optimum resource allocation, along with the use of an appropriate revenue sharing mechanism, which guarantees that each member's share is at least equal or greater than what he would obtain by acting alone. We address the problem of resource allocation in this chapter, while we shall take care of the revenue sharing one later on in Chapter 4.

From an economic standpoint, an alliance might as well be a means to foster the market of ASQ goods. As it has been identified in [2], one possible reason for the lack of QoS market in the current Internet is that buyers are not sure about the quality they would get. A well-reputed alliance, with transparent QoS monitoring mechanisms, would provide a trustworthy marketplace for quality-assured services, reactivating the market and increasing buyers willingness to pay. These ideas are highly related to the *market for lemons* theory, which states that when buyers are not sure about the quality of the good, the market for quality services disappears. We shall comment more on this phenomenon in Chapter 5 and Chapter 6.

As mentioned in Chapter 1, there exists concrete evidence of potential implementation of these conceptual ideas. For instance, different levels of collaboration are expected to emerge in the near future as proposed by the ETICS project [4]. ETICS proposes three different levels of collaboration through what they call the *community* paradigm: a technology-agnostic economic and business layer on which NSPs collaborate to sell end-to-end ASQ goods. These three levels of communities have received the names of *open association*, *federation* and *alliance*. The range of options aims to cover different context on which several business models or mechanisms could arise, taking into account different degree of willingness to collaborate, trust or market maturity [2]. Among the different levels, the open association is the community type that asks for the least collaboration, and the alliance is the one that states the tightest collaboration, with common coordination and

business principles and even a mandatory monitoring infrastructure.

The community concept, though defined at the business layer, implies consequences on the technical lower layers. In particular, a complete set of rules defining each level of collaboration is provided in [2]. We shall reproduce here, for convenience, some of them related to the alliance community, which depict to what extent collaboration is envisaged:

1. Alliance members are prohibited from selling ASQs bypassing the alliance.

2. Each member commits to the resources contributed to the alliance; deviation from the agreed resource contribution will result in penalties for the underperforming NSP.

3. The validity (duration and expiration period), allocation of each NSP's contributed resources and the respective price updates, traffic engineering and policy decisions adhere to the alliance rules.

4. Routing, admission control and resource allocation are regulated by the alliance so that the alliance agreed optimization goals are met.

5. All members of the alliance agree on a common coordination model, which might necessitate a centralised facilitator.

In this remaining of this thesis we propose to work in an alliance context closely related to the ETICS's alliance concept introduced above. Indeed, we assume an important degree of collaboration and trust among NSPs. However, this trust may be represented in some cases by a centralized entity, and we shall not ask NSPs to disclose confidential information among them but only ask them to trust information to the centralized entity. We shall assume, that a common coordination principle exists in order to sell services, and that common business policies related to revenue sharing rules in particular are agreed. Our community concept is as well technology-agnostic, in order to ease the trading of network resources. This gives the name of overlay alliance. In addition, a monitoring infrastructure is assumed to be in place.

3.2.1 A Word on Interdomain Path Computation

The overlay alliance sells services over already established quality-assured paths. Taking into account that our objective is to remain independent of a specific technology and a specific procedure to enable the path computation and provisioning, we shall only briefly comment on this.

The whole cycle of computation and provisioning of interdomain quality-assured services has been specified by the ETICS project, a detailed description is provided in [3]. In particular, the ETICS project has adopted the hierarchical Path Computation Element architecture (H-PCE) [17] to perform interdomain quality-assured path computation. The H-PCE is a a recent standard that allows the selection of an optimum domain sequence and the optimum end-to-end path, through the use of a hierarchical relationship between domains.

In a typical deployment of this architecture, each domain computes internally its internal interdomain network routes, that is to say, the routes within the domain that are going to be used for carrying the interdomain traffic. The Service PCE, a centralized PCE instance, combines Traffic Engineering indicators about the interdomain network topology (obtained from the H-PCEs instances in each NSP) together with service-layer information about prices and customers to compute a sequence of domains.

Once the sequence of domains is determined, the Backward-Recursive PCE-based Computation (BRPC) model [16] is used, where the path computation request is forwarded along the NSP chain and the end-to-end path is computed.

Communication among entities is enabled by the PCE communication Protocol (PCEP). In order to support path computation relying not only in technical aspects but also in business ones, for instance, considering pricing information, some modifications to the standard protocol should be performed.

3.3   The Alliance Model

In this section we shall introduce the mathematical model of the alliance used throughout the remaining of the document. The overlay alliance sells services over already established quality-assured paths, which we shall call simply as *QoS pipe* or *path*. In this scenario, the capacity dedicated by each NSP to sell by this means is a portion of their already deployed capacity. That is to say, NSPs have their infrastructure through which traditional services are sold following the best-effort paradigm and they decide to dedicate some portion of their capacity to the alliance.

3.3.1   Topology Abstraction

An important aspect is to be able to find a suitable model that represents the interdomain topology. For that purpose, we propose to perform some level of topology aggregation or abstraction. This abstraction has to be done mainly because of confidentiality and scalability issues. Abstracting the topology allows as well to have tractable problem formulations.

Different approaches could be used to keep different degree of topology information in the model. For instance, the topology aggregation methods proposed in the ATM framework [19] to solve the problems of scalability. In such framework three levels of topology aggregation are proposed: a simple node one called *Symmetric-point*, a *Star* one and a *Full-mesh* one. In the first case, the domain is abstracted to a simple node. In the star topology, the nodes in the aggregated topology are the border nodes of the domain and they are connected by links through a middle point. In the full mesh topology, the nodes of the aggregated topology are as well the border nodes of the domain, and there is one link representing each path between an origin-destination pair. Some methods to compute the equivalent cost of the links have been proposed (see [74,151] and references therein). Yet another possibility to model a domain in an interdomain network is presented in [77], where an approach similar to the full-mesh one is used, but each path is modelled as a link between an origin-destination pair and an interdomain link. In [67], each domain is abstracted to a logical node and links to their neighbours nodes with capacity. Their justification is that for commercial issues bandwidth on interdomain links will always be the bottleneck, and not internal domain nodes.

As far as our problem is concerned, the mathematical formulation would be exactly the same either considering a more complete topology or a more aggregated one. The question should be then posed over which aggregation method is used, and how this information is updated. These two aspects typically imply a trade-off between scalability and accuracy of the aggregation. Since topology abstraction can be studied separately from the problem tackled in this thesis we leave it out of the scope, while any of the existing methods could be used. Hereafter we will use the simplest approach, which is the Symmetric point scheme.

3.3.2   Definitions and Notations

Let us now introduce some notations so as to formally represent the scenario described above. As aforementioned we shall use the Symmetric point scheme to abstract the NSPs topologies, thus each NSP belonging to the alliance is abstracted to a node indexed by n with an equivalent capacity of c_n. The complete set of nodes is denoted by N. We shall hereafter indistinctly refer to a NSP as node or NSP.

The QoS pipes on sale are the ones in the set S and are indexed by s. The constraint on the delay on path s (i.e. the maximum admissible delay) is denoted by D_s. We assume that the routes within the alliance are fixed and single-path. We represent these routes with the $|N| \times |S|$ matrix R. The entry $R_{i,j}$ is equal to 1 if the route of pipe j traverses node i and is equal to zero otherwise. We denote pipe's s route as $r(s)$. The bandwidth allocated to pipe s (i.e. the amount of traffic sold to the buyers associated to path s) is denoted by a_s. Vector a is defined as $a = \{a_s\}_{s \in S}$. Each path s has a utility function associated to it, which is called $U_s(a_s)$. We assume that $U_s(a_s)$ is known and, as usual in this context, it is a strictly concave function of the bandwidth. We shall address

later on in this chapter a way to determine these utilities.

Please note that the QoS pipes are defined by an ingress and egress point along with an amount of bandwidth and a maximum delay. This implies that two QoS pipes are considered different even if they share exactly the same physical path but provide different delay bounds or amount of bandwidth.

On top of this alliance ASQs goods or services are to be sold in a coordinated way. Our proposed mechanism for such purposes is provided in the following section.

3.4 Bandwidth Allocation with end-to-end QoS Constraints

We assume that for each QoS pipe there is a group of users or buyers interested in getting a portion of bandwidth on that pipe. The amount of money this group is willing to pay as a function of the amount of bandwidth is the so-called utility function. The objective is to sell the available resources in such a way that the revenue of the whole alliance is maximized while the end-to-end constraints are respected. In particular, we shall work with the end-to-end delay, but other additive metrics could be as well easily considered.

Let us first state some additional assumptions. The delay introduced by each node in a path is an increasing convex function of the bandwidth carried by all the paths traversing the node. We assume that this function can be somehow learnt or estimated by the NSP, and we provide an example to model it later on in this chapter. The delay function of node n is denoted as f_n and is a function of the total traffic traversing the node, that is $f_n(\sum_{s \in S} R_{n,s} a_s)$ where $a = \{a_s\}_{s \in S}$, for brevity we shall also refer to it as $f_n(a)$.

We shall address the resource allocation problem following the approach proposed in the seminal work of Kelly [79]. In such work, a Network Utility Maximization (NUM) problem is proposed to solve the resource allocation in such a way that a global network utility function is maximized while links capacities are respected. In our case we apply the same principle to our model of interdomain network and consider in addition end-to-end quality constraints. Indeed, in our scenario the amount of traffic sold to all paths must be such that the revenue perceived by the alliance is maximized, while the QoS constraints are fulfilled. This is formalized in the following bandwidth allocation problem:

Problem 3.1

$$\max_{a_s} \sum_{s:s \in S} U_s(a_s)$$
$$s.t. \sum_{n:n \in r(s)} f_n(a) \leq D_s, \ \forall s \in S.$$

Remark 3.1 In Problem (3.1) we have not included a capacity constraint which is assumed to be taken into account in f_n. Indeed, if f_n is a barrier function (i.e. it approaches infinity as the bandwidth approaches the capacity) we can safely ignore any capacity constraint.

Remark 3.2 Problem (3.1) does not consider either the cost of selling bandwidth on the different paths. However, we can model this situation by defining a cost function of the allocated bandwidth $\kappa_s(a_s)$ for each service $s \in S$ and modifying the objective function in Problem 3.1 by $\sum_{s \in S} [U_s(a_s) - \kappa_s(a_s)]$. This function could be, for instance, compounded by the cost of using the different interdomain links in the path. Provided the cost function is strictly convex, the new problem would be analogous to Problem 3.1. For the sake of notations simplicity we shall not consider the cost function hereafter.

A classical way to solve such convex optimization problem is to look for a saddle point of its

Lagrangian, which in this case is given by:

$$L(a, \lambda) = \sum_{s:s \in S} \left[U_s(a_s) + \lambda_s \cdot \left(D_s - \sum_{n:n \in r(s)} f_n(a) \right) \right], \tag{3.1}$$

where $\lambda = \{\lambda_s\}_{s \in S}$ is the vector of Lagrange multipliers.

In the context of the multidomain alliance we envisage that the services are to be sold in a quite dynamic way. In order to gain in scalability we need to solve Problem (3.1) with the least coordination possible, and not relying on central computations if possible. We shall thus seek for a distributed solution. For such purpose we explore the associated primal-dual gradient dynamics. A priori, it could seem not straight forward to find a distributed solution, since in contrast to the formulation in the classical NUM problem, each term s of the sum in Equation (3.1) is not a separable function of a_s. The distribution in the NUM problem is straight forward since the utility function is assumed to be a separable function of its arguments and the constraints are linear. However, we shall show that the distributed solution is still possible in our case.

Indeed, computing the partial gradients of Equation (3.1) we obtain:

$$\frac{\partial L}{\partial a_s} = \left(U_s^{'}(a_s) - \sum_{n:n \in r(s)} \sum_{v:n \in r(v)} \lambda_v f_n^{'}(a) \right) \tag{3.2}$$

$$\frac{\partial L}{\partial \lambda_s} = \left(D_s - \sum_{n:n \in r(s)} f_n(a) \right) \tag{3.3}$$

We shall propose an iterative approach, in which each iteration is indexed by variable t. Let us define $\dot{a}_s = \frac{da_s}{dt} \equiv \frac{\partial L}{\partial a_s}$ and $\dot{\lambda}_s = \frac{d\lambda_s}{dt} \equiv \frac{\partial L}{\partial \lambda_s}$. Applying a gradient-projection algorithm where the primal and dual variables are iteratively updated in the direction of the respective partial gradient renders the following iterations:

$$a_s^{t+1} = [a_s^t + \gamma_s \dot{a}_s]^+ \tag{3.4}$$

$$\lambda_s^{t+1} = [\lambda_s^t - \alpha_s \dot{\lambda}_s]^+ \tag{3.5}$$

where $[\cdot]^+ = \max\{0, \cdot\}$, α_s, γ_s are positive step sizes and signs were appropriately chosen to find the saddle point of Equation (3.1) (maximum in a and minimum in λ).

The previous manipulation makes it possible to decouple the problem. Indeed, updates (3.4,3.5) are performed iteratively on each edge router of a pipe, which we call the *source*. Every source sends an initial value for λ_s and a_s through route $r(s)$. Each node receives all the values and computes the delay, the derivative of the delay times the sum of the multipliers it has received and sends them back to the source. All these values can be accumulated in two sums in the way back to the source, thus only two values are needed to be sent along the return path on each iteration. Once the source receives such values it proceeds to update the value in λ_s and in a_s. This is repeated iteratively in the control plane and it is run prior to any resource allocation.

The following theorem states the convergence of the algorithm.

Theorem 3.1 *Convergence of the primal-dual algorithm. Given Problem (3.1) let $\sum_s U_s(a_s)$ be a strictly concave function and $f_n(a) \forall n \in N$ convex functions. Then the iterations $a_s^t, \lambda_s^t \forall s \in S$ as defined in Equation (3.4) and Equation (3.5) respectively converge asymptotically to the solution to Problem (3.1).*

The proof can be performed in an analogous way as introduced in [59,60]. Note that the problem can be written as $\max \sum_s U_s(a_s)$ s.t. $g_s(a) \leq 0, \forall s \in S$, where $g_s(a) = \sum_{n \in r(s)} f_n(a) - D_s$. The

proposed primal-dual laws are proved to converge to the optimum of this general problem, provided that U_s is strictly concave and g_s convex, both $\forall\, s$, and for $a \in \mathbb{R}^{|S|}$, see [60]. In our case, the rate vector $a \in \mathbb{R}^{|S|}_+$. An extension of the proof can be readily done by carefully considering the cases when the projection in a coordinate of a gets active (i.e. $[a_s]^+ = 0$, $s \in S$), as detailed in [59].

We shall provide a proof directly adapted from [59], which relies on the Krasovskii's method reviewed for convenience in Appendix A. The proof can be as well carried out using a classical quadratic Lyapunov function as shown previously in [33].

Let us first introduce the following notations.

$$U'(a) = \{U'_s(a_s)\}_{s \in S} \tag{3.6}$$

$$G(a) = \{g_s(a)\}_{s \in S} \tag{3.7}$$

$$D = \{D_s\}_{s \in S} \tag{3.8}$$

Hence, the updates can be expressed by the following dynamics.

$$\dot{a} = K \left[\frac{\partial L}{\partial a}\right]^+ = K\left[U'(a) - \nabla G\lambda\right]^+, \tag{3.9}$$

$$\dot{\lambda} = \Gamma \left[-\frac{\partial L}{\partial \lambda}\right]^+ = \Gamma\left[G - D\right]^+, \tag{3.10}$$

where $K = diag\{k_s\}_{s \in S}$ and $\Gamma = diag\{\gamma_s\}_{s \in S}$ are diagonal matrix with positive step sizes.

The dual of Problem (3.1) with the notation introduced above can be expressed as

Problem 3.2

$$\min_{\lambda \geq 0} \mathcal{D}(\lambda)$$

where $\mathcal{D}$ is defined as:

$$\mathcal{D}(\lambda) = \max_x \left\{\sum_{s \in S} U_s(a_s) - \sum_{s \in S} \lambda_s(G_s(a) - D_s)\right\}, \tag{3.12}$$

We state the following proposition proved in [116], that will aid as in the proof of Theorem (3.1).

Proposition 3.1 *Let $\hat{\lambda}$ be a vector such that the set $\mathcal{M}_{\hat{\lambda}} = \{\lambda \geq 0 : \mathcal{D}(\lambda) \leq \mathcal{D}(\hat{\lambda})\}$ is nonempty. Then $\mathcal{M}_{\hat{\lambda}}$ is bounded.*

We shall now present a proof of Theorem (3.1).

Proof We choose a Lyapunov function V as in the Krasovskii Method (KM), in the state $z = (a^T, \lambda^T)^T$, for the system $\dot{z} = F(z)$. We shall first assume that there are no projections.

Let us define the following

$$Q = 1/2 \begin{bmatrix} K^{-1} & 0 \\ 0 & \Gamma^{-1} \end{bmatrix} \tag{3.13}$$

$$V(z) = \dot{z}^T Q \dot{z} \tag{3.14}$$

We shall now prove that the expression in the KM is negative semidefinite.

$$\dot{V}(z) = \dot{z}^T \left[\left(\frac{\partial F}{\partial z} \right)^T Q + Q \left(\frac{\partial F}{\partial z} \right) \right] \dot{z} \tag{3.15}$$

$$\frac{\partial F}{\partial z} = \begin{bmatrix} K[diag\{U_s''(s)\}_{s \in S} - \nabla^2 G(a)\lambda] & -K\nabla G(a) \\ \Gamma \nabla G(a) & 0 \end{bmatrix} \tag{3.16}$$

Thus,

$$\dot{V}(z) = \dot{z}^T \begin{bmatrix} diag\{U_s''(s)\}_{s \in S} - \nabla^2 G(a)\lambda & 0 \\ 0 & 0 \end{bmatrix} \dot{z} \tag{3.17}$$

Since $U(a_s) \, \forall s$ is assumed to be strictly concave and $G(a)$ is the sum of concave functions ($f_n(a) \, \forall n$ are convex) then the matrix in (3.17) is negative semidefinite and $\dot{V} \leq 0$ over the trajectories.

The above is true for the case where there are no active projections, we will study the case of active projections in the following.

Let $\sigma = (\sigma_a, \sigma_\lambda)$ represent the subset of services which have the projections active. We can write the Lyapunov function as:

$$V(\sigma, z) = \frac{1}{2} \sum_{s \notin \sigma_a} k_s \left(U_s'(a_s) - \sum_s \frac{\partial G(a)}{\partial a_s} \lambda_s \right)^2 + \frac{1}{2} \sum_{s \notin \sigma_\lambda} \gamma_s \left(G_s(a) - D_s \right)^2 \tag{3.18}$$

For any interval where σ is constant, the expression of $\dot{V}$ in Equation (3.17) is still valid, since it means only that there are less terms in Equation (3.18). Therefore V is also decreasing in the situation where there are active projections and where σ is constant. It remains to rule out the case of projection switch.

At switching times, a new term is added or deleted from the expression (3.18). If a term is deleted, V decreases and we are in LaSalle's conditions, that is $V(t^-) \geq V(t^+)$. If a term is added there is no discontinuity.

We are in the conditions of the LaSalle Generalized Invariance Principle (LGP). We shall now show that any trajectory satisfying conditions i or ii of LGP must be an equilibrium.

Imposing condition i. we see from Equation (3.17) and the fact that U_s is strictly concave $\forall \, s \in S$ that $\dot{V} = 0$ implies $\dot{a} = 0$. Thus, a is at equilibrium, say $\hat{a}$.

We now look at the dual variable when $a = \hat{a}$. We shall study $\dot{\lambda}$ from its definition in Equation (3.10). If $G_s(\hat{a}) - D_s < 0$ for any given $s \in S$, the corresponding projection must be active, that is $\lambda_s = 0$, otherwise, λ_s would converge linearly to zero, producing a discontinuity of V when the projection becomes active, violating ii. in LGP.

Finally, we shall see that $G_s(\hat{a}) - D_s > 0$ for any given $s \in S$ is not possible either. If $G_s(\hat{a}) - D_s > 0$, the corresponding multiplier grows linearly in time. In addition, the dual problem is $\max_{\lambda \geq 0} \mathcal{D}(\lambda(t))$, where $\mathcal{D}$ is defined as in Equation (3.12). Hence, $\lambda(t)$ is moving within the optimum of :

$$\mathcal{D}(\lambda(t)) = \sum_{s \in S} U_s(\hat{a}_s) - \sum_s \lambda_s (G_s(\hat{a}) - D_s),$$

and $\mathcal{D}$ is strictly decreasing.

From Proposition (3.1) we know that $\lambda(t) \in \mathcal{M}_{\lambda(0)} \, \forall t \geq 0$. In addition, $\mathcal{M}_{\lambda(0)}$ is a bounded set. Thus, λ can not increase linearly with t, and hence $G_s(\hat{a}) - D_s > 0$ is not possible.

Therefore, the invariant set has λ at an equilibrium $\hat{\lambda}$ as well, which must satisfy either $\hat{\lambda}_s = 0$ or $G_s(\hat{a}) - D_s = 0$, the complementary slackness conditions. $\qquad \square$

3.5 Interdomain Bandwidth Auctions

We now discuss an application that fits to the model proposed before. In particular, we propose to use a first-price auction mechanism to sell quality-assured services and maximize the alliance's revenue. Let us first briefly introduce the main ideas of the trading mechanisms based on auctions, then review how these ideas have been applied to the bandwidth allocation problem and finally propose a means to integrate the use of bandwidth auctions to our NUM problem.

3.5.1 Generalities of Auctions Mechanisms

An auction is a trading mechanism where the price is set through bids. Many flavours of these mechanisms exist, being characterized by two main rules. Namely, they are characterized by the rule that determines which bid wins the auction, and the rule that states how the price is determined. Besides, the mechanisms can differ in the way bids are submitted, namely sealed-bid auctions or open auctions. In the latter the bids can be collected in an ascending order or the selling price can be decreased until a bidder decides to pay the price. Besides, bidders may bid for obtaining one object, or submit multiple bids for obtaining multiple objects (multi-bid).

Auctions have been used for trading during ages. While the oldest recorded use of auctions dates from the Roman Empire period, it is still widely used as a trading means. Nowadays, they are typically used for selling antiquities, artworks and by governments to allocate, for instance, the exploitation of natural resources or the usage of electromagnetic spectrum for communications. Their somehow open and transparent character has made them attractive for those purposes.

Typically, the objective of an auction mechanism is either to maximize the social welfare or to maximize the seller's revenue. Auctions maximizing seller's revenue are referred to as *optimal* and those maximizing social welfare are referred to as *efficient.*

In an auction mechanism, the bid is modelled as a function of the valuation the bidder attaches to the good, and this valuation is in the general case kept secret to each bidder. For that reason, auctions can be seen as a game with incomplete information, where the equilibrium depends on the auctions' rules stated above, as we shall present in detail in Chapter 5 and Chapter 6 for the First-price auctions case.

A usually sought property for mechanisms seeking welfare maximization, is the so-called incentive compatibility property. An auction is incentive compatible if bidders reveal their true valuations.

Probably the most well known auction mechanism is that one where an auctioneer announces ascending selling prices (also called asks) and bidders raise their hands if they are willing to pay the ask. The Auction finishes when nobody is ready to further increase the payment. This is the so-called English auction. Another well-known mechanism is the Dutch auction, where the mechanism is similar to the English one but the selling prices are announced in descending order and the object is sold as soon as a buyer is ready to pay the priced announced by the auctioneer. These constitutes examples of open auctions.

In sealed auctions, bidders privately submit their bids to the auctioneer, who determines, according to already agreed on rules, who wins the auction (generally the highest bid as in first-price or second-price auctions) and the amount to be paid (typically the value of the winning bid for the first-price auction and the second-highest bid for the second-price auction).

Second-price auctions, which have also received the name of Vickrey auctions [149], can be proved to provide welfare maximization and incentive compatibility. The extension to multi-unit auctions, like the Vickrey-Clark-Groves (VCG) (see e.g. [53]) mechanisms, as well keep this property of incentive compatibility, and are the only type of auctions that keep this property when extended from single unit to multi-unit.

In first-price auctions, seller's revenue maximization is sought and the implementation for allocating network resources is much less complex than in second-price auctions, as we shall see in the following subsection. Although first-price auctions do not provide incentive compatibility, since

the winning bidder pays his or her bid and it is thus more rewarding to submit a bid that is smaller to their true valuation for the service, they are suitable to our problem, where the objective is to maximize seller's revenue and keep implementation complexity low. It is for these reasons that in this thesis we shall focus on first-price auctions, as commented in the following subsection.

Related work of network bandwidth auctions and further justification for our choice is provided in the following subsection, while a more detailed introduction to auction theory and its application to telecommunication networks can be found in [53].

3.5.2   Related Work: Network Bandwidth Auctions

We shall now review the literature related to bandwidth allocation and in particular to network bandwidth auctions for that purpose, keeping in mind that our aim is to propose an allocation mechanism that allows for a scalable implementation, maximizing the seller's revenue and assuring end-to-end QoS constraints. Another important aspect of bandwidth allocation in our scenario is how to handle the multi-period. That is to say, that bandwidth is allocated at given periods, but chances are that some services finish before others, and consequently the capacity available at each period that the allocation is carried out may vary. A trade-off between allocating all capacity in the present period and not allocating capacity waiting for future higher bids exists. In the former, if all capacity is allocated, the risk is of not being able to sell that capacity to future, possibly higher, bids. In the latter, the risk is of not receiving such higher bids and having in the end idle capacity. We shall review related work paying attention to this aspect as well.

The literature addressing these topics is very rich, we thus not pretend to exhaustively cover all related work. Indeed, several works in the literature have proposed bandwidth network auctions for solving the bandwidth allocation problem, most of them seek bids' truth revealing mechanisms, that is to say, they seek mechanism that incentivisze the buyers to reveal their true valuation (incentive compatibility) and thus allowing to allocate the object efficiently (i.e. to that one that values it the most). These proposals are mainly based on second-price auctions and derivatives from that mechanism, like the VCG mechanism.

A first proposal appeared in an unpublished paper by Mac Kie-Mason in 1995 [97], where second-price auctions are used at packet level in order to allocate resources in a multiservice network. The mechanism is simple, however it cannot provide end-to-end guarantees for users' traffic.

In [88] the Progressive Second Price (PSP) mechanism was introduced, where several rounds of bids occur till convergence is reached. The authors proved analytically that the method can converge to an equilibrium where incentive compatibility is assured. However, this mechanism presents a high convergence time and significant signalling overhead. In addition, since the mechanism actually presents several equilibria, the one achieved in practice may not provide incentive compatibility as argued in [100].

In [129] the problem of resource allocation is analysed, assuming that connections have a long duration. The approach is based on second-price auctions. In order to address the multiperiod case the same authors propose in [127] a second chance scheme, which means that all bidders compete in the auction (new connections and ongoing previously-established connections) but the previous one have a second chance to bid in case they do not win the bid.

In [98] a proposal based on multi-bid auctions is made, which is based on second-price auctions and provides incentive compatibility and efficiency, and better convergence time and less overhead than previous methods. However, the proposal is derived under the assumptions of an over-provisioned backbone and an access network with tree topology, which does not fit our case. In addition, the multi-period imposes that the connections of users already in the network may change from period to period, which is not our desire.

In [56] a mechanism based on descending price auctions (i.e. Dutch auctions) with a VCG like payment rule is proposed for allocating the bandwidth of the network. The mechanism provides almost efficient allocation and incentive compatibility. The mechanism is applicable to any general

topology network and the authors claim that its implementation complexity is lower than that one of the mechanism in [98]. However, users are supposed to place bids at all links on the path through which they need the bandwidth. A similar approach is proposed by the authors to address the multiperiod case; users are requested to submit bids not only for all links in the path but for all timeslots in which they want to reserve bandwidth. We consider that this approach is not suitable to our scenario, where the customer is not expected to have any knowledge about the topology supporting the service he or she wants to buy.

In [45] a proposal based on second-price auction that allows for a distributed implementation is presented. Actually, the decentralized part consists of collecting the bids at the edge routers, and then sending to a centralized entity performing the optimization some few parameters that can rebuild an approximate utility function. No considerations of end-to-end delay are present in the admission of traffic, thing that we do in this chapter.

In [143] the implementation complexity shortcomings of previous second-price-based proposals were addressed and a double-sided auctioning mechanism (bids and asks are allowed) for bandwidth allocation in interdomain network allowing distributed implementation is proposed. The trade-off is not fulfilling efficiency neither incentive compatibility. This work does not present QoS constraints, thus it is not suitable for our problem.

In the aforementioned cases the objective is welfare maximization and that is the reason why mechanism which provide incentive compatibility are sought. On the contrary, few works are based on First-price auctions. However, when seller maximization is sought it becomes reasonable to prioritize optimality rather than efficiency, if the latter supposes a higher, or even prohibitive, implementation complexity. For instance, in [41] the proposal is to work with First-price auctions. In that case, the authors claim that the complexity of the mechanism based on second-price auctions is not justified, since their objective is to maximize the seller's revenue. Indeed, the implementation complexity when using First-price auctions is much smaller than the one when using Second-price auctions, and thus more suitable to our multidomain case. Moreover, in [99] it is shown that VCG mechanisms can hardly be applied on multidomain networks.

In addition, the Revenue Equivalence Theorem (see e.g. [84]) states that under certain assumptions (mainly risk-neutral symmetric bidders, definitions which we shall see in Chapter 5) all types of auctions have the same expected revenue for the seller.

Regarding how the multi-period case is tackled, in most of previous work either the issue is not addressed, either second chances schemes, as introduced in [127] and adopted as well in [98], are proposed. In those schemes, at every new allocation period the services which are already being carried out compete among new ones, and have a second chance to bid if they loose. Our aim is, however, to guarantee that the service is not going to be interrupted, while not compromising too much the seller's revenue. In [41], the issue is addressed in both senses, that is to say guaranteeing the services are not interrupted, and guaranteeing the seller to maximise his revenue in the long term, by proposing policies that consider future bids and statistics on the bids. While we do not focus on this point, this approach is applicable to our case.

For the reasons exposed above, our auctions proposal is to use fist-price auctions. The idea is aligned with the one in [41]. However, we consider a multidomain federation scenario rather than a single domain and we incorporate an end-to-end QoS constraint rather than only considering capacity constraints. With respect to this last aspect [135] states a similar problem, but its context and the way it is solved differ significantly from ours.

3.5.3 Application: Network Utility Maximization with QoS constraints and First-Price Auctions

We now introduce the use of first price auctions along with the NUM problem. Let us associate to each pipe a service to be sold which has a certain bandwidth σ_s and an assured delay D_s (for instance, this service can be a VoD movie). Several instances of a service are sold through the same pipe. These services are sold by means of first-price network bandwidth auctions.
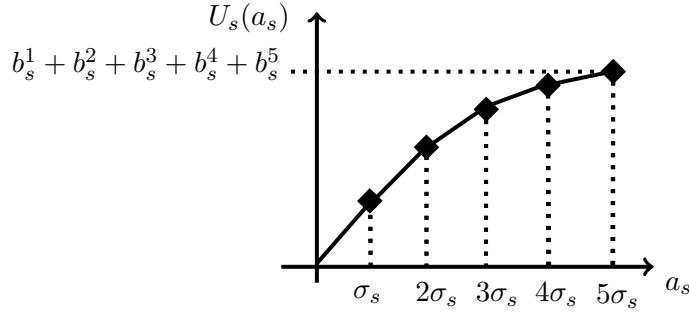
Figure 3.1: Example of an utility function for service s built-up with the ordered bids received for service s.

We shall first consider the case of one-shot bandwidth auctions. That is to say, that the whole capacity available for providing the services is allocated at one certain moment.

Let us introduce some new notations. For each service s there are M_s buyers or bidders, which participate in the auction for obtaining an instance of the service. Each of the M_s buyers bids $b_s^{(i)}$ which we order as

$$b_s^{(1)} \geq b_s^{(2)} \geq \cdots \geq b_s^{(M_s)}. \tag{3.19}$$

The resource allocation decision is to find which of these bids to accept, so as to maximize the revenue of the whole alliance while the per-route delay remains smaller than a given bound, under a first-price auction. Since for each s all bids are for the same bandwidth and delay constraint, the optimal solution is accepting the highest bids per service. We define the variable $\psi_{s,i}$ which is equal to 1 if bid i for service s is accepted, and zero otherwise. Then, defining the variable m_s as the number of bids accepted for service s we have the following equality:

$$\sum_{i=1}^{M_s} b_s^{(i)} \psi_{s,i} = \sum_{i=1}^{m_s} b_s^{(i)}. \tag{3.20}$$

Accepting m_s bids would render a total accepted rate of a_s where $a_s = \sigma_s m_s$. Thus, the utility per service can be defined as a function of a_s as

$$U_s(a_s) = \sum_{i=1}^{a_s/\sigma_s} b_s^{(i)}. \tag{3.21}$$

Equation (3.21) is defined for discrete values of a_s (the multiples of σ_s). We extend it to a piecewise linear concave function of a_s by linear interpolation. An example of this obtained utility function for a service s with bandwidth σ_s is shown in Fig. 3.1.

All together, we can write the allocation problem as follows:

Problem 3.3

$$\max_{a_s} \sum_{s \in S} U_s(a_s)$$

$$s.t. \sum_{n:n \in r(s)} f_n(a) \leq D_s, \, \forall s \in S, \, a_s/\sigma_s \in \mathbb{Z}.$$

In Problem (3.3) the objective function is concave but not strictly concave (as in Problem (3.1)) and an integer restriction has been added. These two changes need to be addressed carefully. Indeed, since integer programming is NP hard, we have strong indication of the difficulty of this problem, not easy to overcome even allowing for centralized computation. We will thus accept a

sub-optimal allocation which involves solving the convex relaxation, and rounding off to satisfy the integer constraints.

Besides, the not strictly concavity of the utility function may lead to two problems. The maximum might not be unique, and oscillations may appear. Regarding the former, it would occur in the case of non strictly convex constraints when the solution, for each a_s would be a set $\{a_s \in \mathbb{R}_+ : [a_s/m_s\sigma_s] \in [0,1]\}$, where notation $[\cdot]$ means the integer part. However, this problem is readily overcome when rounding down in order to have integer values of $a_s \, \forall s \in S$. Regarding the convergence of the algorithm, Theorem (3.1)'s hypothesis ask for strict concavity in U_s and convexity in g_s, $\forall s \in S$. Since U_s is now a piece-wise linear concave function, the oscillations in the accepted bandwidth can occur at points where the derivative of the utility function does not exist (that is at multiples of σ_s for each $s \in S$). This means that oscillations merely imply accepting one bid more or one bid less. In consequence, they have not a big impact on the result.

In any case, removing oscillations is possible through different methods. For instance, the objective function could be modified, as in the so-called proximal method, proposed in [44] and used in [41,94], which changes the objective function so as to make it strictly concave while keeping unchanged the optimum.

With respect to the constraints, so far we have assumed that the delay function f_n is a somehow learnt convex function of the bandwidth, and more precisely a function of the bandwidth traversing the node, that is $f_n(a) = f_n(\sum_{s \in S:n \in r(s)} a_s)$. We have as well assumed that it is a barrier function, that is to say, that it approaches infinity when the bandwidth traversing the node approaches its equivalent capacity. Indeed, any function verifying such conditions is suitable. For instance, one can model each node as a M/M/1 queue and assume the delay of traversing such node as the average delay of the queue. In such case the delay function takes the form:

$$f_n^{M/M/1}(a) = \frac{K}{c_n - \sum\limits_{s:n \in r(s)} a_s},\qquad(3.22)$$

where we have assumed a mean packet size equal to K. In Section 3.6 we shall present some simulation studies where the M/M/1 model is assumed.

To finalize this section let us say a word on multi-period allocations. We have shown so far how to allocate bandwidth through auctions in a distributed manner and in one-shot. A more realistic scenario would be considering that the allocation occurs for a given period of time, and that the allocation mechanism occurs periodically. Indeed, for selling the services we repeat the process described above in a periodic fashion. Every period of time T, bids are collected and bandwidth is allocated. As mentioned in subsection 3.5.2, most previous work on multi-period auctions (e.g. [56]) allow future bidders to compete with incumbent ones, albeit given the latter some advantage (e.g. [127]). A different approach (e.g. [41]) is to impose the condition that once bandwidth has been allocated in an auction, the successful bidder has a reservation for the duration of his or her connection. We shall not present the specific solution for the multi-period auctions problem since any of the previous proposals can be adopted. However, because of the nature of the services the alliance is expected to sell, we consider that the proposals that avoid connection or service interruption are the more appropriated to our problem.

3.6 Simulations

We now revisit the services examples of Chapter 2. Namely, a VPN service, a VoIP service and a HDTV Internet service. For illustrative purposes we shall refer to concrete services even if the alliance might just sell QoS pipes to brokers or intermediate buyers, without distinguishing which services are carried by those pipes. The set of services available is thus $S = \{VPN, VoIP, HDTV\}$ and their characteristics are the following ones: $\sigma_{VPN} = 270Mbps$, $D_{VPN} = 200ms$, $\sigma_{HDTV} = 19.4Mbps$, $D_{HDTV} = 300ms$ and $\sigma_{VoIP} = 1Mbps$, $D_{VoIP} = 150ms$. The alliance topology and services routes are shown in Fig. 3.2a. For each service 60 buyers bid for obtaining a chunk of

bandwidth. We assume random bids and proceed as explained in Section 3.5 to build the utility functions. More precisely, for each service, bids are independently drawn from a common exponential distribution. The means of the exponential distributions for the different services are such that the VPN service is the one getting the highest bids, followed by the HDTV service. The VoIP service is the one with the smallest offers, the less demanding in terms of amount of bandwidth but the most demanding with respect to the end-to-end delay. The utility functions are shown in Fig. 3.2b, where for convenience in the display they are plotted as functions of the number of bid instead of as functions of the bandwidth.

We then solve the allocation problem through the iterations in Equation (3.4) and Equation (3.5). Fig. 3.2c and Fig. 3.2e show the evolution of the primal and dual variables respectively. We can observe that both variables have converged by 2000 iterations. We can check that when the algorithm converges the delay bounds are respected for all the services. Moreover, the necessary Karush–Kuhn–Tucker conditions for the optimum can be checked. That is to say, at the optimum the following conditions hold: 1) the Lagrangian multipliers are all greater than or equal to zero, 2) the delay for each service is smaller than or equal to the corresponding delay bound, and 3) the so-called complementary slackness condition, for each service at least one of the following occurs: a) the Lagrangian multiplier is equal to zero, or b) the delay is equal to the bound. All conditions can be readily verified to occur once the iterations on Fig. 3.2 have converged. In particular, we observe that for the VoIP and the VPN services the delay is equal to the bound, while for the HDTV service the delay is lower than the bound for such service, but the corresponding Lagrange multiplier is equal to zero.

(a) Topology.

(b) Utility functions.

(c) Evolution of the rate (i.e. a_s) for each service.

(d) Detail of the rate for the HDTV service.

(e) Evolution of the Lagrange multipliers.

(f) Evolution of the delay.
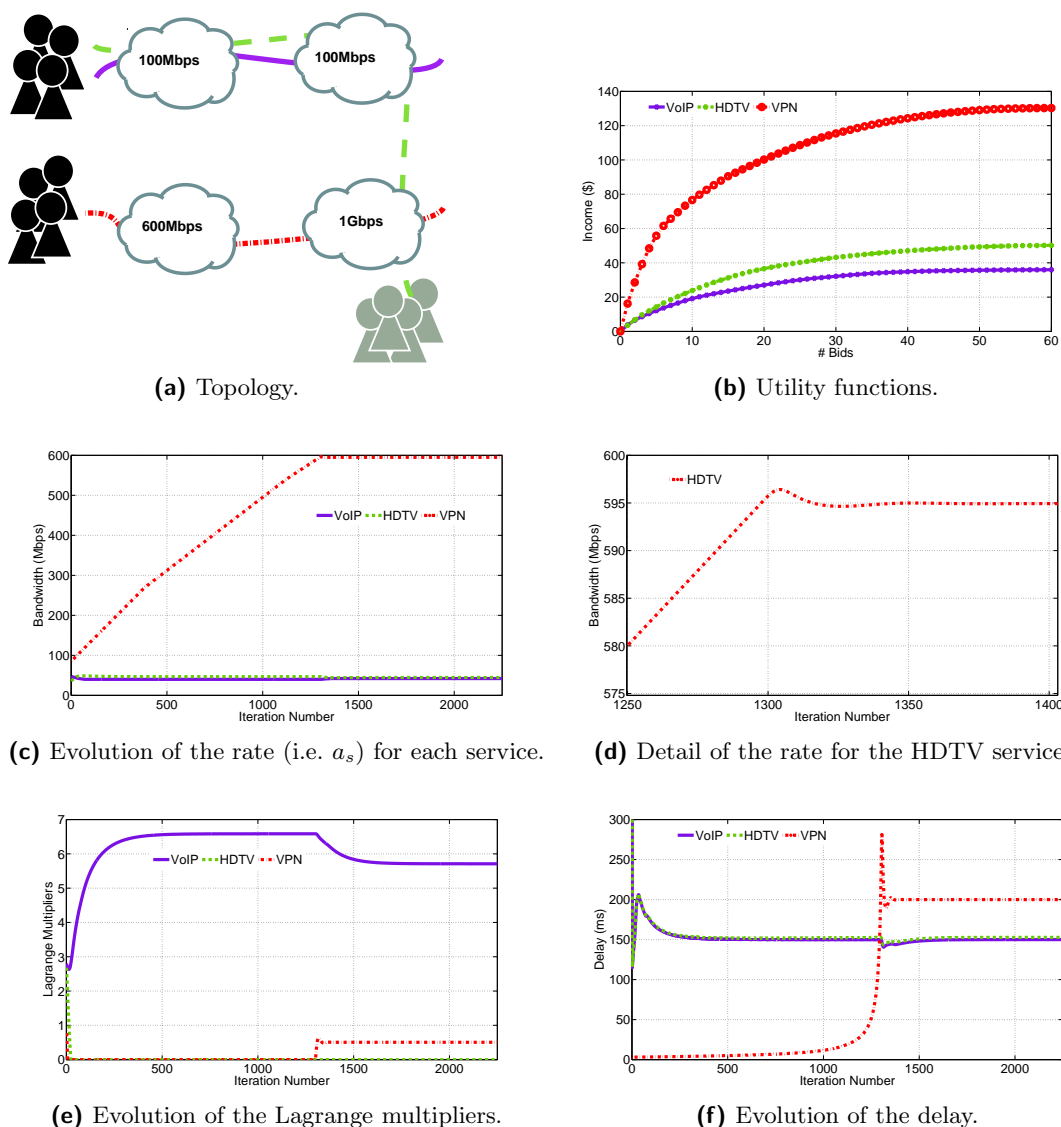
Figure 3.2: Bandwidth auctions with QoS constraints, one-shot allocation. Simulation example.

3.7 Implementation Considerations

The proposed mechanism allows to allocate bandwidth respecting end-to-end QoS constraints and in such a way that the revenue of the alliance is maximized. The use of auctions provides a flexible pricing mechanism suitable for new services, where market price is not necessarily known. In that regard, we claim that from the theoretical viewpoint the proposed mechanism is appealing. We now briefly comment on implementation issues. As stated in Chapter 1, the multidomain scenario poses new problems that are not experienced in the context of the intradomain one. For instance, political aspects, such as confidentiality and trust among domains, technical aspects, such as interoperability and scalability, and economic ones as, for instance, business models' coordination and revenue sharing.

In the NSP alliance context, it is reasonable to considered that the NSPs tell the truth and fulfil their common interests. Nevertheless, they might ask for confidentiality, privacy on committed agreements and freedom on pricing [123].

We claim that the proposed framework introduced in this chapter preserves confidentiality. Indeed, the bandwidth allocation is performed in a distributed fashion, where the values of the delay

of traversing the NSPs and its derivative are passed from one NSP to another being accumulated in a sum. This makes that the QoS metrics of each domain are not propagated along the routes, only the second node in a route knows this information about the source node, and for the remaining of the nodes information is not separable by domain. In addition, the disclosure of domain topology information is not needed at all.

Pricing can be freely defined at the per service level for the premium services, through, for instance, the so-called reserve prices (a minimum acceptable value for bids imposed by the seller). For best-effort traffic, prices can still be defined at a per NSP level.

Finally, the proposed solution appears to scale well. For the rate allocation only a few bytes in the forward and backward direction are needed during a preallocation iteration phase. These are the values of the primal and dual variables in Equation (3.2) and Equation (3.3) for the forward path, and the values of the delay and its derivative of traversing each NSP in the way back to the source. The latter can be accumulated along the route, thus only two values per service need to be sent back to the source node. Nevertheless, we have not focused on optimizing the convergence time of the algorithm, such as for instance considering an adaptive step value for updating the primal and dual variables.

3.8 Summary

In this chapter we have introduced two key aspects that define our working scenario. First, we have formally introduced the concept of NSP overlay Alliances. We have discussed their pertinence when selling end-to-end assured quality services, and we have presented a mathematical representation for the alliance, which relays on a topology abstraction. Second, we have proposed a means for allocating bandwidth in such alliances with the main objective of maximizing the alliance's revenue. We have stated the problem of network bandwidth allocation as a network utility maximization problem with end-to-end QoS constraints, which constitutes one of the main contributions of this chapter. We have then shown that a distributed solution to that problem can be performed. We have then discussed the pertinence of using first-price auctions for selling enhanced network services and proposed its usage along with the network utility maximization problem. Regarding implementation, we have argued that the proposed method is scalable and preserves confidentiality. Finally, we have shown the results of simulation studies which demonstrate the behaviour of the proposed selling mechanism.

Chapter 4

Splitting Revenues

4.1 Introduction and Motivation

In every cooperative task, burdens and profits are naturally expected to be shared in a fair way among all the involved actors. Besides, the means of sharing, whichever costs or revenues, is as well a means to influence the actors' behaviours. These two general statements, are valid in many contexts, and are arguable motivations for a revenue sharing method tailored to our particular context. Indeed, NSPs will collaborate only if the revenue share obtained from such collaboration is attractive from their point of view. Regarding the revenue sharing persuasive power, consider the following simple example to highlight its importance. For instance, if the NSP responsible from bringing more clients to the alliance is compensated adequately, all NSPs would dedicate efforts to attract clients towards the alliance, which in the end would result in higher revenues for the alliance. The aspects discussed in this chapter are motivated by both a fair and a right-incentive provider revenue sharing method, for our particular context. This particular context, determined by the overlay NSPs alliances and the bandwidth allocation method introduced in Chapter 3, makes the revenue sharing a challenge, since existing methods can not be applied directly and must be carefully revised. The main reason why these methods are not suitable for our case is because their good properties relay on specific properties of the revenue function, which in our problem do not hold. For instance, the so-called Shapley value, a well known revenue sharing method, does not guarantee that members have incentive to collaborate and remain in the alliance since our revenue function is not a convex function of the capacities. Moreover, the intersection of solutions that provide both incentive properties and fairness in the broad sense is not always clear. In this chapter we aim to find such intersection, while seeking as well implementable solutions.

In the sense commented above, this chapter provides the following contributions: formal representation of the problem and discussion of the desired properties, evaluation of existing methods which concludes that none of them are suitable for our problem, guidelines for a new method, and a solution proposal. The method is validated through simulation studies. Our studies are motivated by their application in the NSPs alliance context. However, it is worth highlighting that the conclusions derived in this chapter apply to much broader situations. Indeed, they are valid for every context where the resources are allocated solving a NUM-like problem.

This chapter is organized as follows. Section 4.2 introduces additional notation used throughout the chapter and states the desired properties for the revenue sharing mechanism. We then present related work, which is split into two sections. Section 4.3 reviews the most common sharing rules used in the economics field, and argue on why they are not useful for our problem, while Section 4.4 comments on related work in the networking field. In Section 4.5 we present a new method, which provides a solution that guarantees stability and efficiency in economic terms. Simulation results that demonstrate the correct behaviour of the proposed method are shown in Section 4.5.1.C and Section 4.5.2.C along with some implementation considerations. Finally, a summary of the chapter is presented in Section 4.6. Further simulation results that support the conclusions of this chapter are provided in Appendix B. This chapter is based on the results published in [29].

4.2 Problem Description

4.2.1 Definitions and Notations

The notation needed to represent the NSPs alliance, and the services sold on top of it has already been introduced in Chapter 3. We now introduce extra notations in order to represent the revenue sharing problem. The set of nodes N is here as well referred to as the grand coalition and sub-groups of nodes receive the name of sub-coalitions or sub-alliances. The revenue function associates to each sub-coalition $\mathcal{Q} \subseteq N$ with capacities $c^{\mathcal{Q}}$ and utility functions $U = \{U_s\}_{s \in S}$, a real value $V(\mathcal{Q}, c^{\mathcal{Q}}, U)$, where $c^{\mathcal{Q}}$ is the capacities vector restricted to sub-coalition $\mathcal{Q}$, that is:

$$c_n^{\mathcal{Q}} = c_n \text{ if } n \in \mathcal{Q} \; ; \; c_n^{\mathcal{Q}} = 0 \text{ otherwise.} \tag{4.1}$$

In our framework, services are sold according to the NUM problem introduced in Chapter 3. For clarity's sake and ease on the intuition of the results we shall only consider capacity constraints, instead of considering in addition end-to-end QoS constraints. The revenue function V is thus given by the solution of Problem (4.1), which states that services are sold (i.e. bandwidth is allocated) in such a way that the revenue of the alliance is maximized, while respecting the capacity constraints.

Problem 4.1

$$\max_a \sum_{s \in S} U_s(a_s)$$
$$s.t. \; Ra \leq c^{\mathcal{Q}}$$

We also accept the notation $V(\mathcal{Q})$ to indicate the total revenue of coalition $\mathcal{Q} \subseteq N$, where the capacities and utility functions are implicit or $V(\mathcal{Q}, c)$ when utility functions are implicit by context.

We define the contribution v_n of node $n \in N$ to the alliance as $v_n = V(N) - V(N \setminus \{n\})$ and we shall refer to the contributions vector v defined as $v = \{v_n\}_{n \in N}$. The total revenue is shared among all the nodes in N according to the sharing function $\Phi(N, c, U)$ which computes a revenue sharing vector $\{\Phi_n\}_{n \in N}$, where Φ_n is n's share. This function depends on the coalition N, the capacities c and the utility functions U. For convenience and brevity, we shall also use the shorter notation x to denote the revenue sharing vector, where $x \in \mathbb{R}^{|N|}$ is a column vector containing on each component x_n, $n \in N$, the revenue share of node n, when the values of N, c and U are implicit by context.

4.2.2 Desired Properties of the Revenue Sharing Mechanism

We shall now comment on the properties that are usually seek for a revenue sharing mechanism and highlight those pertinent for the NSP alliances scenario. Generally speaking, two main objectives motivate the properties. First, to provide with fair and sensible allocations. Second, to provide to the NSPs the right incentives to remain in the alliance and contribute to it. The properties discussed below, are usually discussed in cost/revenue sharing problems, with slightly different definitions (see for instance [47], [72], [112]). We select from them the ones that we believe are of more relevance to our problem, argument on why they are relevant and formally define them.

First, the mechanism should distribute all the alliance's revenue among its members, that is what we call efficiency, which is defined as

Efficiency. $\Phi(N, c, U) = x$ is efficient if

$$\sum_{n:n \in N} x_n = V(N). \tag{4.2}$$

In order to assure the sustainability of the alliance, the mechanism should not provide incentives to any sub group of NPSs to break the grand coalition. That is to say, no sub-coalition should

have economic incentives to form a smaller coalition outside the alliance, since this would lead to instabilities in the alliance. This is the so-called stability property.

Stability. $\Phi(N, c, U) = x$ is stable if

$$\sum_{n:n \in \mathcal{Q}} x_n \geq V(\mathcal{Q}), \ \forall \mathcal{Q} \subseteq N. \tag{4.3}$$

The stability definition requires the shares to be such that, for every possible sub-coalition, the sum of the shares of the nodes belonging to that sub-coalition are as least as large as the revenue that sub-coalition would perceive. This property is also usually referred as the *stand alone* property. Please note that this definition also implies that the revenue perceived by each node $n \in N$ in the coalition is not less than the revenue it could achieve alone, i.e. $x_n \geq V(\{n\})$, $\forall n \in N$. The set of points that verify Equation (4.3) constitutes the so-called *core set* in the context of Coalitional Game Theory. The reader is referred to [140] for more details on the core concept and coalitional game theory.

We shall now focus on the incentives provided by the revenue sharing mechanism. Regarding resources, the mechanism should provide the right incentives to the nodes to increase their resources towards the coalition. In our model, these resources are considered in the capacity. We formally define this property as follows.

N-resource-monotonicity. $\Phi(N, c, U)$ is N-resource-monotonic if given c and $\hat{c}$ two vectors of capacities, such that $\hat{c}_k \geq c_k$ and $\hat{c}_n = c_n$ $\forall n \neq k$ then $\Phi_n(N, \hat{c}, U) \geq \Phi_n(N, c, U)$ $\forall n \in N$.

N-resource monotonicity provides incentives to a NSP to increase its capacity, while NSPs not increasing their capacity have no incentive to discourage such increase, which makes it a very compelling property.

The aim is to have a rule that satisfy the property for any given alliance, which is intuitively a very hard task. As a consequence, we shall also admit a less restrictive property, which we simply call resource-monotonicity, and only ask for a non-decrease on the revenue of that NSP increasing its capacity. More formally:

Resource-Monotonicity. Given c and $\hat{c}$ two vectors of nodes capacities, such that $\hat{c}_n = c_n$ $\forall n \in N \setminus \{n\}$ and $\hat{c}_n \geq c_n$, Φ is resource-monotonic if $\Phi_n(N, \hat{c}, U) \geq \Phi_n(N, c, U)$.

The resource-monotonicity property means that if an NSP increases its capacity then its revenue will as well increase or remain the same. This property is usually referred as to resource incentive, or monotonicity in the resources, in the context of coalitional game theory. We shall herein simply refer to it as *Monotonicity*.

Regarding incentives, the monotonicity property still provides incentives to a NSP to increase its capacity towards the alliance. In the context of coalitional game theory, N-resource-monotonicity is usually asked so as to guarantee that no player would have incentives to block some other player's action towards the increase of the revenue of the whole alliance. In particular, if we ask simply for Monotonicity and not for N-Resource-Monotonicity, it could happen that some NSPs in the alliance have no interest in the capacity increase of other NSPs. However, even if the capacity increase of one NSP could decrease some other's NSP revenue, we claim that this situation could act itself as an incentive to NSPs to remain competitive in terms of their resource contribution towards the alliance.

Beyond resources, sharing criteria could be provided with respect to the utility functions. In the context of industrial operations, for instance, an arguable desirable objective is to reward for efficiency, that is if two process are carried out in a same machine, and the marginal revenue/cost of one is greater/smaller than the one of the other process, revenue/cost shares should respect this order [155]. In our context, we could extend this concept by considering the marginal revenue produced by a node with respect to its capacity when utility functions change. We capture this in

the property that we name Revenue-monotonicity. More formally, let $U^{(1)}$ and $U^{(2)}$ be two vectors of revenue functions. The revenue-monotonicity property is defined as follows.

Revenue-monotonicity.   A revenue sharing rule Φ is revenue-monotonic if given two different vectors of utility functions $U^{(1)}$ and $U^{(2)}$ then for any $n \in N$ verifying $\frac{\partial V(N,c,U^{(1)})}{\partial c_n} \geq \frac{\partial V(N,c,U^{(2)})}{\partial c_n}$ the revenue shares verify $\Phi_n(N, c, U^{(1)}) \geq \Phi_n(N, c, U^{(2)})$.

The intuition behind this property is that if the marginal revenue due to n is greater with utility functions $U^{(1)}$ than with utility functions $U^{(2)}$, then the revenue share with one and other vector of utility functions should respect this order. Intuitively, this rule provides the right incentives in our case. Indeed, if the utility functions increase, then the revenue of the whole alliance increases, or remains the same. However, in our problem the revenue function is not differentiable along the whole values of c, so the applicability of this property is limited. The alliance as a whole has a clear incentive in the increase of the utility functions, since this increases the total revenue. However, the pertinence of this property as an incentive to nodes might be questionable, since utility functions are something exogenous in the model, thus the NSPs' influence on the utility functions can be argued to be limited, or at least not captured by our model. For these reasons we shall not consider this property as mandatory, though it is interesting to evaluate it when possible.

Finally, another interesting property related to monotonicity is the one that evaluates the influence of new members entering the alliance, usually referred as to population monotonicity.

Population-monotonicity.   Given capacity vector c, a revenue sharing rule Φ is population monotonic if $\forall S, T \subseteq N$ such that $S \subset T$, $\Phi_n(T, c^T, U) \geq \Phi_n(S, c^S, U) \ \forall n \in S$.

A population-monotonic revenue sharing rule guarantees that the entrance of a new NSP to the alliance does not reduce the revenue of each of the NSPs already there. However, we focus on the study of fixed alliances, and not on the dynamics of how to build them. The alliance could be set in place for reasons further to the ones captured by the revenue function, such as business agreements and geographical coverage. Thus, we shall not consider this property as mandatory for a sharing rule. It is rather a property that should be checked whether it is verified or not given a particular alliance.

We now focus on fairness. We want the mechanism to be fair in the sharing. There is not a general consensus in the literature regarding the notion of fairness. Moreover, the properties enumerated before can as well be interpreted as fairness. Indeed, the stability property states that every sub-coalition will get an aggregate share of at least the sub-coalition's revenue. The monotonicity property, besides providing incentives to increase capacity, can as well be interpreted as a fairness one. Indeed, if one NSP makes an effort to improve its capacity towards the alliance, then it deserves to be rewarded adequately.

In addition to the properties stated so far, we propose some common in the literature and intuitive rules that should be fulfilled. We base our definitions on the contribution of each node, defined for node n as $v_n = V(N) - V(N \setminus \{n\})$. If $v_n \geq v_j$ then $x_n \geq x_j$, which is usually known as *order preserving*. If the previous inequalities are interpreted as strictly, that is if $v_n = v_j$ implies $x_n = x_j$ this has received the name of *equal treatment of equals* and is probably one of the oldest stated fairness rules, which was in particular a legacy of Aristotle [38]. In addition we propose $v_n = 0$ then $x_n = 0$, which we shall call *no free-riders* and can be seen as a particular case of the so-called dummy property in the context of coalitional game theory. All in all, we define fairness as follows.

Fairness. Φ is fair if it is order preserving, guarantees equal treatment of equals and no free-riders.

The choice of the contribution vector to evaluate fairness is important. Some classical rules of revenue or cost sharing propose fairness criteria based on the resources or costs of each agent. However, in our case, we base the fair principle in the contribution of each NSP in terms of revenue

rather than in terms of capacity, since what is important in the alliance is not only the capacity provided by each NSP but as well its position in the topology. Vector v is a measure that takes into account both relevant components, namely the capacity and the position in the topology.

4.3 State-of-the-Art Sharing Methods

We now present existing revenue sharing techniques, which have been proposed in the field of economics. A detailed review can be found in [72]. We also comment on why these techniques are not suitable for our problem.

4.3.1 The Shapley Value

The Shapley value, proposed by Lloyd Shapley in 1953 [138], is probably the most well known technique to perform revenue sharing in an association or coalition. It has been widely used in the literature for its good properties, which we shall review in the following. The Shapley value provides a closed-form expression to compute the share of each agent in a cooperative context. Intuitively, it can be interpreted as computing the average contribution of each agent to the coalition. Let us formalize this in what follows.

With our notations, the Shapley value for player $n \in N$ is defined as:

$$x_n^{sh} = \frac{1}{|N|!} \sum_{\mathcal{Q} \subseteq N \setminus \{n\}} |\mathcal{Q}|!(|N| - |\mathcal{Q}| - 1)! \left[V(\mathcal{Q} \cup \{n\}) - V(\mathcal{Q}) \right]. \tag{4.4}$$

From Equation (4.4) we can see that indeed agent n's share depends on $[V(\mathcal{Q} \cup \{n\}) - V(\mathcal{Q})]$, n's contribution to sub-coalition $\mathcal{Q} \subseteq N$. To see the intuition behind the formula for node n, imagine that what we are computing is n's contribution to all possible sub-coalition $\mathcal{Q} \subseteq N$, and averaging over all the different sequences according to which the grand coalition can be formed. More precisely, suppose the grand coalition is built up from the empty set adding nodes uniformly at random. When the turn comes to add node n, compute its contribution to the sub-coalition formed by all previously added nodes, let us call that previous sub-coaliton as $\mathcal{Q}$. The contribution of node n to $\mathcal{Q}$ is the same regardless the order on which $\mathcal{Q}$ is built. We thus multiply n's contribution to $\mathcal{Q}$ by all the possible ways on which $\mathcal{Q}$ can be formed, that is by $|\mathcal{Q}|!$. Analogously, we multiply that result by the $(|N| - |\mathcal{Q}| - 1)!$ possible ways of choosing the remaining of the nodes. Finally, we sum over all possible sub-coalitions $\mathcal{Q}$, and take the average by dividing by the number of possible orderings of all nodes, that is $|N|!$.

We present now a concrete, simple example of the computation of the Shapley value in the case we are studying. Consider Topology A in Fig. 4.1a, capacities are equal to 1 unit for all nodes and services of 1 unit of bandwidth are sold. Consider as well that the utility functions are such that $V(N) = V(\{2,3\}) = 5$, $V(\{1,3\}) = 2$.

The Shapley value for node 1 is given by:

$$\begin{aligned} x_1^{sh} &= \frac{1!1!}{3!} [V(\{1,2\}) - V(\{2\})] + \frac{1!1!}{3!} [V(\{1,3\}) - V(\{3\})] \\ &+ \frac{2!1!}{3!} [V(N) - V(\{2,3\})] \\ &= \frac{1}{6} 0 + \frac{1}{6} 2 + \frac{1}{3} 0 = \frac{1}{3}. \end{aligned}$$

Analogously, for node 2 we obtain $x_2^{sh} = \frac{11}{6}$, and for node 3 $x_3^{sh} = \frac{17}{6}$.

We now introduce some useful definitions that will characterize the Shapley value. A Pre-imputation is the set of payoff vectors such that the sum of all x_n is equal to $V(N)$. A Dummy player is a player whose contribution to the coalition is the same as the one he would achieve on his own. With these definitions the axioms of Symmetry (if n and j contribute the same to any coalition then $x_n = x_j$), Dummy player (if n is a dummy player then $x_n = V(\{n\})$) and Additivity

(a) Topology A.   **(b)** Topology B.
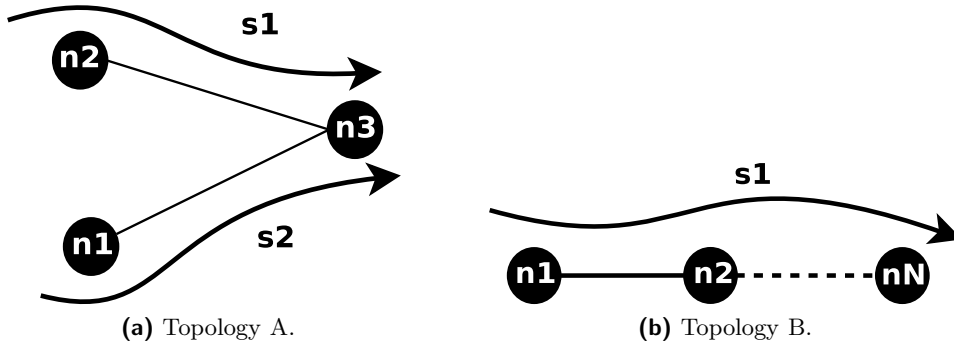
Figure 4.1: Example topologies.

$(\Phi_n(N, c, U^{(1)} + U^{(2)}) = \Phi_n(N, c, U^{(1)}) + \Phi_n(N, c, U^{(2)}) \; \forall n \in N)$ are introduced. The Shapley value is the only sharing rule verifying all three axioms, as proved by Shapley [138].

In addition to verifying the Symmetry, Dummy player, and Additivity properties stated above, the Shapley value is Efficient (it shares the total revenue), Fair according to its own definition of fairness, and Resource-monotonic. Fairness is defined in terms that for any two players $n, j \in N$, n's contribution to j is equal to j's contribution to n, that is $\Phi_n(N, c, U) - \Phi_n(N \setminus \{j\}, c, U) = \Phi_j(N, c, U) - \Phi_j(N \setminus \{n\}, c, U)$.

Recalling the previous example, we can readily verify that the solution obtained is Efficient, that is $x_1^{sh} + x_2^{sh} + x_3^{sh} = \frac{2 + 11 + 17}{6} = 5 = V(N)$. We can as well verify that the solution is Fair according to the previous definition. Take for instance nodes 1 and 2, we have $\Phi_1(N, c, U) - \Phi_1(N \setminus \{2\}, c, U) = \frac{1}{3} - 1 = -\frac{2}{3}$ and $\Phi_2(N, c, U) - \Phi_2(N \setminus \{1\}, c, U) = \frac{11}{6} - 2.5 = -\frac{2}{3}$.

In the following theorem we prove that the Shapley value applied to our problem provides Resource-monotonicity.

Theorem 4.1 *Incentive for improving capacities. Let (N, V, c) be a coalitional game where the set of nodes N are the players, c represents the equivalent capacities of the nodes in N and V is the revenue function defined by Problem (4.1). If $n \in N$ increases its capacity then its revenue share (i.e. Φ_n) will be not decreased. That is, letting $\hat{c}$ represent the capacities of the nodes where n's capacity is increased, $\Phi_n(N, \hat{c}, U) \geq \Phi_n(N, c, U)$, where $\Phi_n(N, c, U)$ is the Shapley value of node n given the game (N, U, c).*

Proof By definition of Shapley value $\Phi_n(N, \hat{c}, U) = \frac{1}{|N|!} \sum\limits_{\mathcal{Q} \subseteq N \setminus \{n\}} |\mathcal{Q}|!(|N| - |\mathcal{Q}| - 1)![V(\mathcal{Q} \cup \{n\}, \hat{c}) - V(\mathcal{Q}, \hat{c})]$, where $V(\mathcal{Q}, \hat{c})$ as defined above represents the worth function for sub-coalition $\mathcal{Q}$ when the capacities are given by $\hat{c}$.

$$\Phi_n(N, \hat{c}, U) = \frac{1}{|N|!} \sum_{\mathcal{Q} \subseteq N \setminus \{n\}} |\mathcal{Q}|!(|N| - |\mathcal{Q}| - 1)![V(\mathcal{Q} \cup \{n\}, \hat{c}) - V(\mathcal{Q}, c)],$$

holds since the revenue function of any coalition without node n is the same, regardless the capacity of node n. By subtracting n's revenue share with and without increasing its capacity we have:

$$\Phi_n(N, \hat{c}, U) - \Phi_n(N, c, U) = \frac{1}{|N|!} \sum_{\mathcal{Q} \subseteq N \setminus \{n\}} |\mathcal{Q}|!(|N| - |\mathcal{Q}| - 1)![V(\mathcal{Q} \cup \{n\}, \hat{c}) - V(\mathcal{Q} \cup \{n\}, c)].$$

We now determine if the inequality $V(\mathcal{Q} \cup \{n\}, \hat{c}) \geq V(\mathcal{Q} \cup \{n\}, c) \; \forall \mathcal{Q} \subseteq N \setminus \{n\}$ holds. Indeed, V is the solution to Problem (4.1), which is the maximization of a concave function with convex constraints. By increasing the capacity we relax such problem, thus doing so yields to greater or equal solutions, which concludes the proof. $\qquad\square$

In spite of fulfilling the aforementioned compelling properties, the Shapley value is not suitable for our problem, as it does not always provide stable solutions. That is to say, shares computed

through the Shapley value do not always fulfil inequalities given by Equation (4.3). Nonetheless, its great popularity in previous work is due to the fact that it is proven that it provides with stable solutions when the revenue function is a convex function of c (see e.g. [140]), which occurs in many cases. Moreover, when the revenue function is convex the Shapley value provides with stable, efficient and N-monotonic solutions (see e.g. [72]). Examples where the Shapley value is adopted as the solution concept to a sharing problem in the networking context are commented in Section 4.4.

As for our problem, the revenue function V is not a convex one and solutions through Shapley value can lie outside the *core*. This can be seen, for instance, in simple examples as the previous one. We have computed the revenue share in that example through Shapley vale, which renders $x^{sh} = (1/3, 11/6, 17/6)$. We shall show in Section 4.5 that the *core* in that example is $\{x = (0, 3 - \epsilon, 2 + \epsilon) : \epsilon \in \mathbb{R}, 0 \leq \epsilon \leq 3\}$. Hence x^{sh} does not belong to it.

4.3.2   The Proportional Share

One of the simplest way to perform the revenue sharing is to split revenues proportionally to some contribution measurement. In our case, as explained above, vector $v = \{v_n\}_{n \in N}$ where, we recall, $v_n = V(N) - V(N \setminus \{n\})$, quantifies this contribution. Using the definitions introduced in Section 4.2 we can write the proportional share as:

$$x_n^{pr} = \frac{v_n}{\sum_{n' \in N} v'_n} V(N). \tag{4.5}$$

For a concrete example, consider again Topology A in Fig. 4.1a and that the utility functions are such that $V(N) = V(\{2, 3\}) = 5$, $V(\{1, 3\}) = 2$. The proportional share for node 1 gives:

$$\begin{aligned} x_1^{pr} &= \frac{V(N) - V(\{2, 3\})}{\sum_{n \in N} V(N) - V(N \setminus \{n\})} V(N) \\ &= \frac{5 - 5}{(5 - 5) + (5 - 2) + (5 - 0)} 5 = 0. \end{aligned}$$

Analogously we obtain, $x_2^{pr} = \frac{15}{8}$ and $x_3^{pr} = \frac{25}{8}$.

The proportional share a priori seems to be a very attractive distribution rule. It fulfils the properties of Efficiency and Fairness and it is very simple to compute. However, it has the drawback that it does not always guarantee Stability, as we shall show later on this chapter.

4.3.3   The Aumann-Shapley Rule

The Aumann-Shapley Rule for cost sharing [34] was introduced by Shapley and Aumann in 1974, and can be applied analogously for a revenue sharing problem. The idea of this rule is to compute the revenue share of node $n \in N$ as its average marginal revenue along a certain path going from capacity equal to 0 to c_n. More precisely, the share for node $n \in N$ according to this rule is defined as:

$$x_n^{as} = \int_0^{c_n} \partial_n V(N, \frac{t}{c_n} c) dt = c_n \int_0^1 \partial_n V(N, tc) dt, \tag{4.6}$$

where the notation $\partial_n V(N, c)$ means the first order derivative of V at c with respect to c_n. Please note that in Equation (4.6) we have used the alternative notation for V where its dependency on the sub-coalition and the equivalent capacities is explicitly mentioned.

In first place, it must be noticed that the derivative of V with respect to c_n is not defined for all values of c_n. Indeed, consider a simple topology with only one service crossing several nodes, as Topology B shown in Fig 4.1b and consider that all nodes have the same capacity. Let $\hat{c}$ be that capacity. If a given node n increases its capacity, the other nodes will act as bottlenecks and the revenue will not change, while if n reduces its capacity then it will itself become the bottleneck and the revenue will decrease. Hence, the derivative of V takes different values at both sides of $\hat{c}$ and it

is not defined at $c_n = \hat{c}_n$. What is even more important, this rule does not fulfil the Monotonicity property, this is due to the characteristics of our revenue function. Furthermore, this rule applied to our problem could even provide incentives to reduce capacity, since nodes are incentivized to be the bottlenecks.

We now present a concrete example. Consider Topology B given by Fig. 4.1b and that the capacity of all nodes but node 1 are equal to a value c_{max}, and node 1's capacity is equal to c_{min}, with $c_{min} < c_{max}$. Consider that the utility function is a linear function of the admitted bandwidth, with slope U_1. Thus $V(N) = U_1 c_{min}$. In this case the Aumman-Shapley rule gives to node 1:

$$x_1^{as} = \int_0^{c_{min}} U_1 dt = U_1 c_{min}.$$

Analogously, we can compute the share for every node $n \in \{2 \dots N\}$. Since the derivative of V with respect to c_n is zero $\forall n \in \{2 \dots N\}$, as the revenue is determined by c_{min}, which is limiting the admitted bandwidth, $x_n^{as} = 0$.

From this simple example we can readily see that this rule applied to our scenario can provide the wrong incentives. Indeed, every node except node 1 is interested in decreasing its capacity so as to become the bottleneck, as this would give that node a non-null share.

4.3.4   The Friedman-Moulin Rule

This rule was proposed by Friedman and Moulin in 1999 [64]. We introduce the operator $\wedge$, which is defined for two vectors r and $q \in \mathbb{R}^{|N|}$ as $q \wedge r = \min(q_n, r_n)$ $n \in N$ and column vector e, which is of dimension $|N|$ and has all its components equal to one. This rule is similar to the Aumman-Shapley one, in terms that it integrates marginal revenues, but in this case the integration is done through a different path. According to the Friedman-Moulin rule, the share for node $n \in N$ is calculated as:

$$x_n^{fm} = \int_0^{c_n} \partial_n V(N, t \cdot e \wedge c) dt. \tag{4.7}$$

This rule can not be applied in our context since V is not derivable along the whole path, for the same reasons explained above.

4.4   Sharing Techniques Used in the Networking Field

The literature related to sharing rules is extremely vast. However, it has been mainly a domain of study of economists, from which we have reviewed the main results and solution concepts in the previous section. Approaches similar to ours, which we shall shortly present to be based on choosing a point from the stable and efficient set by optimizing some objective function, have been discussed before, see for instance [73,78]. However, the discussion there remains for general games, which not necessarily apply to our particular kind of game, where the revenue is determined by a NUM problem.

In this Section we aim, rather than to comment on sharing rules literature, to review how this issue has been addressed in the networking field, which solutions have been adopted from the economics field and tailored for particular networking problems, more closely related to our problem. In other words, we shall focus on applications rather than on tools.

As aforementioned, the Shapley value is probably the most widely used method for sharing costs or revenues in a collaborative context. So is the case in the networking field. For instance, the Shapley value has been used in [96], where the proposal is to change the Internet economics by business contracts whose payment is determined by the Shapley value. And also in [113], where the aim is to optimize the routing within an alliance of NSPs and the revenue is shared by means of Shapley value. More recently, it has also been used in [141] for splitting cost savings among several domains.

Yet another interesting usage of the Shapley value is proposed in [111], where a content delivery network takes advantage of a peer-to-peer architecture and peers are encourage to collaborate by having a reduction in their connectivity fees. These reductions are computed using the Shapley value. A fluid approximation is shown to apply when the number of peers becomes large, which makes it possible to easily compute the reductions.

It is only in contexts where stability is desired and where the Shapley value does not provide solutions in the stability set that different methods are used. A very interesting approach was proposed quite recently in [47] to share the costs of providing connections with guaranteed capacity. The problem treated there is quite different from ours, moreover their cost function is linear with the capacity. However, compelling properties are discussed, such as monotonicity, stability, which have nourished our discussion of desirable properties in Section 4.2.

In [112] a cost sharing rule for sharing the cost of a set of non-redundant services is proposed. In particular an application to share the costs of connectivity in a network among its users is presented. Such rule could be readily adapted so as to be applied to our revenue sharing problem. Indeed, the share of an NSP n would be composed by summing up *per-services* shares for those services passing through NSP n. Each *per-service* being defined by evenly splitting the revenue due to the service, among the NSPs in the route of such service. However, we have shown through an example that this rule applied to our problem does not provide Monotonicity.

This chapter has intentionally not discussed bargaining approaches. Bargaining processes model classical economic problems in which players negotiate in order to collectively choose an outcome, in situations where there is no consensus about which the best outcome is (see for instance [119]). Formalizing the output of this problem is not an easy task, since there is a plethora of actions that a player could take throughout the process, and a wide variety of ways to carry out the negotiation, involving different sequential order on which the players play and time-frames on which they take decisions or the game ends. In particular, according to the bargaining power, preferences and impatience of the different players, different outputs could be obtained. Bargaining situations lie in the core of the interests of game theory, discipline which has provided different formalizations of such problems, among which the solution concepts derived first by Nash in 1950 [115] and then by Rubinstein in 1982 [133] are seminal papers. Both of these solutions deal with two players problems, and generalizations to more players, even if they have been proposed (e.g. [85]), are not trivial.

In particular, in the NSPs alliance scenario, several players (i.e. NSPs) should negotiate to agree on an outcome (i.e. revenue share). In order to be able to unambiguously predict the output of a bargaining process, the preferences of the NSPs, over all possible shares should be determined. In [67] a revenue sharing method for NSPs that work jointly to provide and end-to-end service is proposed and is based on a bargaining problem. However, the bargaining power there is arbitrarily defined, granting some claimed pertinent properties. They propose to trade the optimal pricing by a distributed allocation and revenue sharing mechanism. That is to say, that the revenue of the alliance is not maximized, in exchange of allowing for a decentralized revenue sharing mechanism. The method is said to provide monotonicity, however no discussion about stability is presented. It is a very interesting proposal in line with our interest but however in a different scenario. In our scenario, the NSPs' bargaining powers are not clearly stated and would be very difficult to establish. Indeed, a NSP has interests in receiving at least a certain revenue, and so do the different sub-coalitions of NSPs, but knowing the preferences of each NSP's towards different revenue shares is, at least, not easy. On the other hand, in our framework, trust is assumed among NSPs and a central trusted entity can be assumed to be in place. Thus, a more straightforward solution seems to be relaying on this centralized trusted entity to compute the revenue sharing solution, so as to combine all NSPs interests in the best way possible.

4.5   The Proposed Method

Having seen that existing techniques are not suitable for our problem, we shall now propose a new method to perform the revenue sharing in our specific scenario. We focus on two properties:

Stability and Efficiency. Nevertheless, we shall present a flexible method which allows to include further properties. We first study the set of possible solutions and following we focus on how to choose a point belonging to that set. For clarity's sake, we consider this set of solutions in a simple scenario, which we call the one-shot scenario. In this scenario services are sold through what we call a service selling phase and revenue sharing is performed right afterwards. In the one-shot scenario we consider given utility functions, and its random nature is not considered. We shall later on move to a multi-period scenario, in which several service selling phases occur and revenue sharing is performed once over the revenues of all the periods. Each service selling phase occurs for independent realizations of the utility functions, and we shall discuss different approaches that work with statistics on the utilities.

4.5.1   One-Shot Scenario

We now study the problem assuming that service selling is performed and revenue sharing occurs right afterwards.

4.5.1.A   The Feasible Solutions Set

In order to have stability in the coalitions inequality (4.3) must hold, that is $\sum_{n:n\in\mathcal{Q}} x_n \geq V(\mathcal{Q}), \forall \mathcal{Q} \subseteq N$. Let us enumerate all the possible sub-coalitions $\mathcal{Q} \in N$ and index them using index $j = 1\dots 2^{|N|}$. We rewrite inequality (4.3) as a linear system as:

$$Qx \geq \hat{v}, \tag{4.8}$$

where $Q = \{Q_{j,n}\}$ is a $2^{|N|} \times |N|$ matrix that indicates which nodes belong to each sub-coalition (i.e. $Q_{j,n} = 1$ if node n belongs to sub-coalition j and 0 otherwise) and $\hat{v} = \{V(\mathcal{Q}_j)\}_{j=1\dots 2^{|N|}}$ is the vector that indicates in the j-th component the revenue of sub-coalition j.

We must consider at the same time the Efficiency property, which we write as the vector representation of Equation (4.2):

$$e^T x = V(N). \tag{4.9}$$

We refer to the set of points verifying Equation (4.8) and Equation (4.9) as to the feasible set. Depending on the alliance topology and the utility functions, the feasible set might determine a unique point, a non empty set included in $\mathbb{R}^{|N|}$, or an empty set. The following examples show two cases where the two latter situations occur.

An empty feasible set.   Consider Topology C shown in Fig. 4.2. The capacities of all nodes are equal to 1 unit. The three services illustrated on the mentioned figure are sold, each one of them is defined for 1 unit of bandwidth. Nodes' capacities are all equal to 1 unit, and the utility values for the one unit of bandwidth of each service are $U_1(1) = 5$, $U_2(1) = 4$ and $U_3(1) = 2$. Thus, the revenue of the sub-coalitions are $V(N) = V(\{1,2\}) = 5$, $V(\{2,3\}) = 4$ and $V(\{1,3\}) = 2$.

In order to achieve stability the total revenue (5 monetary units) must be split in such a way that every route receives at least what they would receive alone. It is not difficult to see that this is not possible at the same time for all routes, since the following inequalities must hold: $x_1 + x_2 \geq 5$, $x_1 + x_3 \geq 2$, $x_2 + x_3 \geq 4$ and $x_1 + x_2 + x_3 = 5$. Hence, the feasible set is empty.

It is interesting to remark that for different values of the utility functions, and the same topology, the feasible set could be non-empty.

A feasible region.   Consider now Topology A shown in Fig. 4.1a. The capacities are again equal to 1 unit for all nodes and we sell services of 1 unit of bandwidth. Utility functions are now $U_1(1) = 5$ and $U_2(1) = 2$. Thus, $V(N) = V(\{2,3\}) = 5$ and $V(\{1,3\}) = 2$. A feasible solution must fulfil $x_1 + x_3 \geq 2$, $x_2 + x_3 \geq 5$ and $x_1 + x_2 + x_3 = 5$. The vectors x that satisfy all equations are $\{x = (0, 3 - \epsilon, 2 + \epsilon) : \epsilon \in \mathbb{R}, 0 \leq \epsilon \leq 3\}$, which corresponds to a segment in $\mathbb{R}^2$.
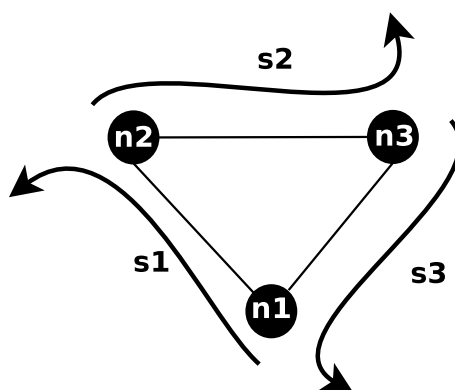
Figure 4.2: Topology C.

4.5.1.B   The Choice of a Point within the Feasible Solutions Set

We have seen in the previous subsection that configurations with no solution can exist, in this case we claim that the coalition for those utility functions should not exist as such, since there is no revenue sharing method that can make it stable. This decision could be made based on mean utility functions, as we shall comment on the Multi-shot scenario in the following section. Therefore, we focus our attention on the case where constraints (4.8) and (4.9) determine a region. In order to choose a point from such region we formulate the following Optimization Problem:

Problem 4.2

$$\min_x \; f(x)$$
$$s.t. \; Qx \geq \hat{v}, \; e^T x = V(N),$$

where $f(x)$ is a convex function. Please note that we can dispense with the restriction of non negative revenue shares, since it is already considered by the Stability property. Indeed, the constraints $Qx \geq \hat{v}$ include constraints of the form $x_n \geq V(\{n\})$, $\forall n \in N$ and by definition $V(\{n\})$ is non negative. Problem (4.2) constitutes a family of methods which can be tuned to cover additional properties by considering different objective functions. We now introduce different objective functions. We shall explore the properties they provide in the following subsection through simulations.

Projections.   One possible natural approach is to consider either the Shapley value or the Proportional share and project either of these vectors onto the feasible set. The method would inherit the good properties of the Shapley value or the Proportional share as appropriate, when the share is already in the feasible set, and otherwise it would return the closest value. In those cases the objective functions of Problem (4.2) would be $f(x) = ||x - x^{sh}||^2$ or $f(x) = ||x - x^{pr}||^2$, where the square of the norm is considered in order to have a quadratic programming optimization problem.

Another possibility is to project the contributions vector v onto the feasible set. Intuitively this would behave well as a sharing rule, since we are choosing the closest point to the contributions vector. Please note that the proportional share is actually a linear transformation of the contributions vector. However, the projection onto a polytope is not a linear transformation so the results of projecting contributions vector v and the Proportional share need not to be the same. In this case, the objective function takes the form $f(x) = ||x - v||^2$.

Equalization of the shares.   Yet another candidate to the objective function is the square of the Euclidean norm of the revenue share vector, that is $f(x) = ||x||^2$. Considering this function would intuitively provide with more even shares among the nodes. On the other hand, this function

does not keep any record of the nodes' contributions to the revenue, thus monotonicity and fairness are likely not to be fulfilled.

Weighted sums. Another intuitive candidate for the objective function is one that shares proportional to either the capacity or the contributions vector. We shall consider thus the sum of the shares weighted by either, the capacity or the contribution of each node. This comes to a linear objective function and in order to have a convex function we consider the opposite of the weighted sum, that is $f(x) = -c^T x$ or $f(x) = -v^T x$. Intuitively this criterion would give more share to highest weights.

The impact of the choice of the objective function is evaluated through simulations in the following subsection.

Regarding implementation aspects, the proposal is to have a central trusted entity computing the revenue shares. This entity must know the utility functions for each service and the topology of the alliance, at the NSP level. We shall comment more on this in Section 4.8, the Multi-period scenario.

4.5.1.C   Simulations

Simulations were performed with two different objectives, namely to further show the need of a new revenue sharing method, and to evaluate the proposed method. Illustrative results are shown in this chapter, while exhaustive simulative studies evaluating the proposed method's behaviour are shown in Appendix B. The simulations presented in this chapter were performed on a regular computer with a i5 processor of 2.67GHz and 3.6 GB of RAM memory. The optimization problems were solved using CPLEX through AMPL.

The need of the new revenue sharing method We shall consider the topologies in Fig. 4.3, where $c_n = 10$ for all nodes n and the amount of bandwidth of each service $\sigma_s = 5$ for all services s, all values being expressed in some coherent unit. Tabel 4.1a shows the utility in some monetary unit, say \$, for carrying 5 and 10 units of bandwidth, where the sum of the underlined values corresponds to the revenue of the alliance (that is, the sum is the solution to Problem (4.1)).

(a) Topology 1.   **(b)** Topology 2.

(c) Topology 3.
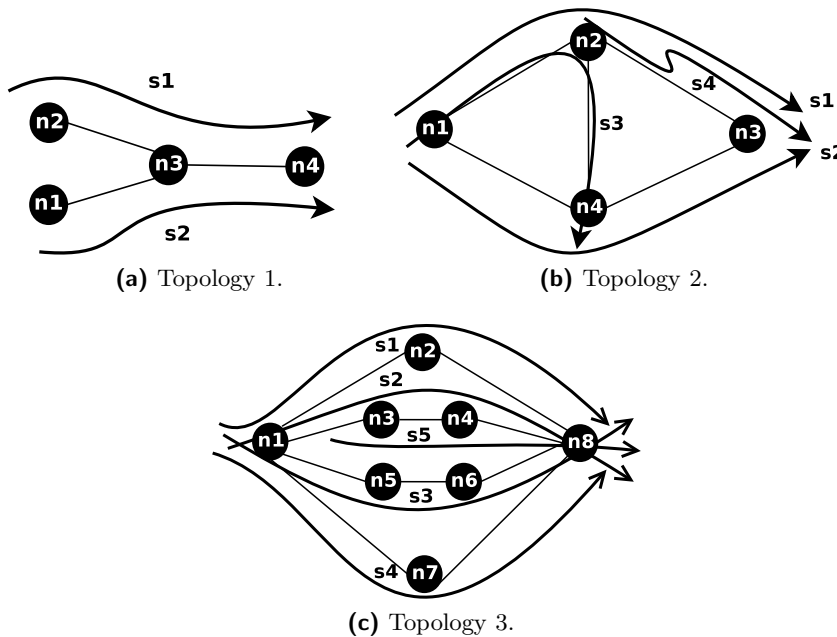
Figure 4.3: Topologies used throughout the simulation studies.

Service	Utility ($)					
	$U_s(5)$	$U_s(10)$	$U_s(5)$	$U_s(10)$	$U_s(5)$	$U_s(10)$
s1	1	2	5	9	$\underline{7}$	8
s2	6	$\underline{9}$	$\underline{7}$	11	2	3
s3	-	-	$\underline{11}$	16	5	8
s4	-	-	$\underline{12}$	18	5	8
s5	-	-	-	-	$\underline{6}$	11
	Topology 1		Topology 2		Topology 3	

(a) Utility values

Topology	Revenue Share (%)				
	$x_\perp^{sv}$	x^{sv}	$x_\perp^{pr}$	x^{pr}	v_n
1	0.284	0.278	0.280	0.280	7
	0	0.018	0	0	0
	0.358	0.352	0.360	0.360	9
	0.358	0.352	0.360	0.360	9
2	0.175	0.175	0.200	0.211	12
	0.353	0.353	0.344	0.333	19
	0.297	0.297	0.256	0.246	14
	0.175	0.175	0.200	0.211	12
3	0.130	0.183	0.077	0.105	2
	0.023	0.076	0.077	0.105	2
	0.039	0.106	0.039	0.053	1
	0.039	0.106	0.039	0.053	1
	0	0.014	0	0	0
	0	0.014	0	0	0
	0	0.036	0	0	0
	0.769	0.465	0.768	0.684	13

(b) Results using existing methods and its projections into the feasible set.

Table 4.1: Revenue sharing, one-shot scenario. Illustration of the need of a new sharing method.

Revenue shares were computed using the Shapley value, defined by Equation (4.4) and the proportional share, defined by Equation (4.5). These values were afterwards projected into the feasible set. That is to say, Problem (4.2) was solved setting $f(x) = ||x - x^{sh}||$ and $f(x) = ||x - x^{pr}||$. Results are shown in Table 4.1b (where notation $x_\perp^*$ stands for the projection of * into the feasible set), along with the value of the contribution v_n for each node. Topologies 2 and 3 constitute examples where the Proportional share does not lie into the feasible region, so do topologies 1 and 3 for the case of the Shapley value. We can easily verify this by noting that projections are different to the original vectors. Consequently, as aforementioned, such methods are not suitable for our problem.

The impact of the objective function We now explore through simulations the use of the different objective functions introduced above. By construction, all solutions verify the Efficiency and Stability properties. Hence, we are interested in evaluating their behaviour with respect to the Fairness and Monotonicity properties. We shall thus divide the simulation studies into those two cases.

Fairness evaluation. Consider topologies in Fig. 4.3 and the Utility functions shown in Table 4.2. Capacity nodes are equal to 10 and services bandwidth is 5, all values in a coherent unit. We compute the revenue share in these scenarios with the proposed method and the objective

Service	Utility ($)					
	$U_s(5)$	$U_s(10)$	$U_s(5)$	$U_s(10)$	$U_s(5)$	$U_s(10)$
s1	<u>5</u>	7	6	10	2	4
s2	<u>3</u>	6	<u>20</u>	27	2	3
s3			<u>18</u>	31	5	7
s4			<u>10</u>	16	6	8
s5					8	<u>16</u>
	Topology 1		Topology 2		Topology 4	

Table 4.2: Utility functions for simulations evaluating the Fairness property.

functions introduced above. Results are shown in Fig. 4.4, along with the Shapley value and the Proportional share, all values normalized.

Fig. 4.4 should be read as follows. On the right-hand side of the figure the Proportional share and the Shapley value are shown as reference values to ease the interpretation of the results. These values are normalized, thus the Proportional share coincides with the contributions vector. The Proportional shares are stacked up in descending order, bottom-up from the highest contribution to the lowest. This same order is respected in all bars and in the legend. On the left-hand side of the figure each bar corresponds to the stacked shares computed with the proposed method and with different objective functions. Shares are as well stacked up following the order imposed by the Proportional share.

Fig. 4.4a shows that considering the weighted sums as objective functions does not provide Fairness. Indeed, the *equals treatment of equals* property is not verified. We can readily see that, while the contributions of nodes 3 and 4 are the same, their shares differ significantly since node 3 is getting all the revenue and node 4 receives no revenue. Moreover, with a linear objective function Problem (4.2) does not necessarily have a unique solution, which is the case in this example. Indeed, since the criterion is linear, it can be proved that the minimum is found at a vertex of the feasible set. Any vertex minimizing $-\sum(x_1 + x_2 + x_3 + x_4)$ or $-\sum(x_1 + 2x_2 + 8x_3 + 8x_4)$ for the cases weighted by the capacities and by the contributions respectively, is a solution to the revenue sharing problem (i.e. a solution to Problem (4.2)). In this case, for instance, for the capacity weighted function, the vectors $(0, 0, 8, 0)$, $(0, 0, 0, 8)$, $(1, 2, 0, 5)$ and $(1, 2, 5, 0)$ are all examples of feasible solutions, of which Fig. 4.4a only shows the former. However, none of them gives the same share to node 3 and node 4, so fairness in not fulfilled with any of those solutions. We discard the use of linear functions since they neither provide with a unique solution nor they provide fairness.

Fig. 4.4b shows that function $f(x) = ||x||^2$ does not provide with Fairness either. In this case is the property *order preserving* that is not verified. Indeed, while, for instance, node 1's contribution is greater than node 4's contribution, they all receive the same share. We observe as well that, as commented above, considering this objective function tends to equalize the shares.

Fig. 4.4c shows that the *no free riders* property is verified by all the considered functions, while it is not respected by the Shapley value. Indeed, only nodes 8, 3 and 4 have positive contributions, while all nodes receive some non null share according to the Shapley value. The solutions computed through the proposed method with any of the considered objective functions respect the no free riders properties.

In addition, results with all three topologies show that projecting the Shapley value, the Proportional share, and the contributions vector, behave well with respect to our definition of Fairness. In all cases order preserving, no free riding and equal treatment of equals is fulfilled.

With respect to the projection of the contributions vector, it could be surprising that for nodes with strictly positive contribution, the rule may assign a null share, while Stability and Fairness are fulfilled. This is the case, for instance, for node 1 and node 2 in Topology 1, whose results are shown in Fig. 4.4a. However, take for instance node 1, if it would have a strictly positive revenue when acting alone, that is if $V(\{1\}) > 0$, a strictly positive share would be assured through the Stability property. In other words, even if its share is null, it would not get more revenue by acting
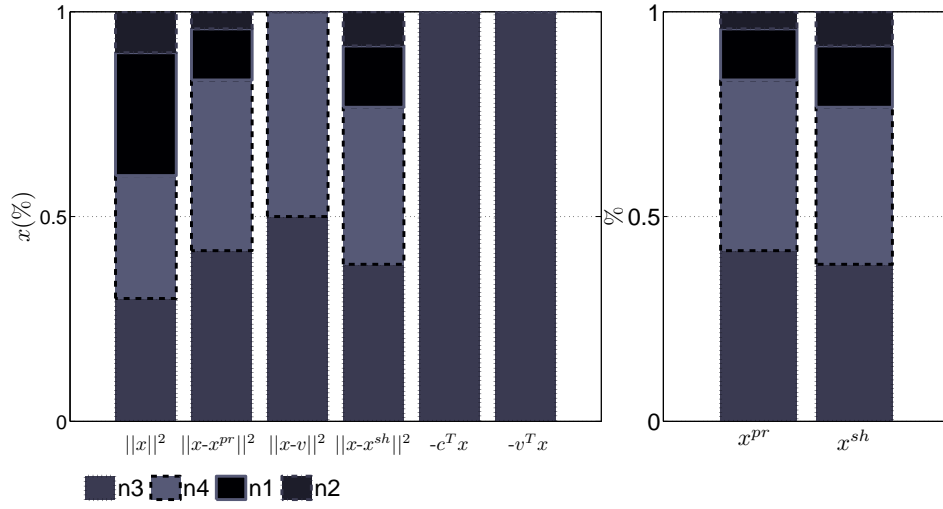
(a) Results for Topology 1.
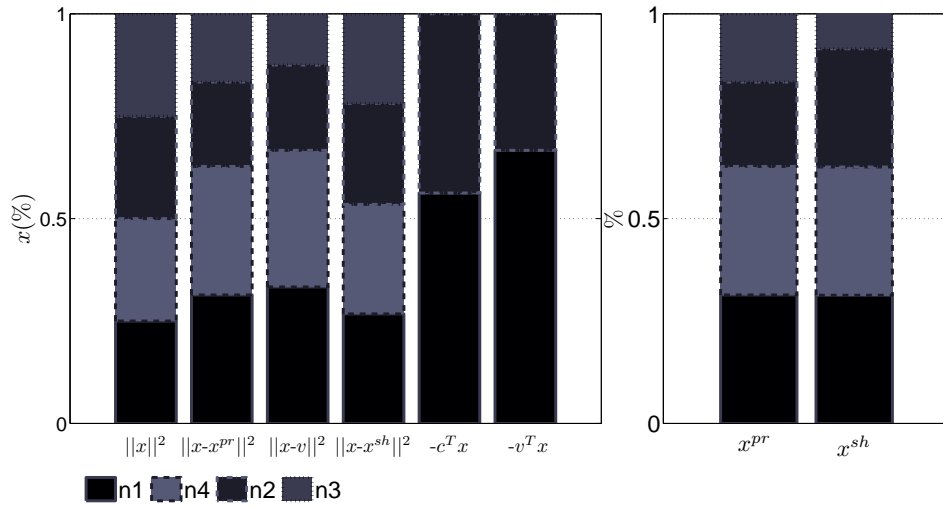
(b) Results for Topology 2.
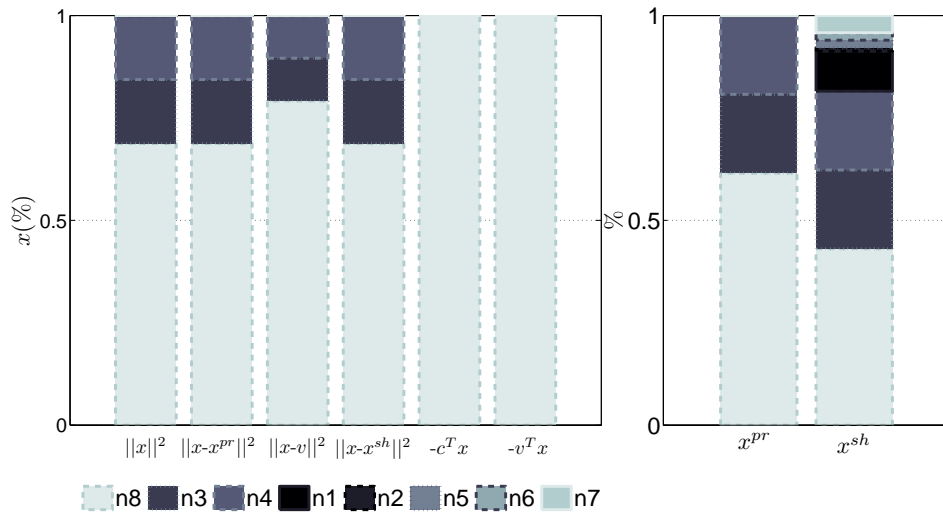
(c) Results for Topology 3.

Figure 4.4: Revenue sharing with the proposed method and different objective functions. Evaluation of the Fairness property.

	Utility ($)	
Service	$U_s(1)$	$U_s(2)$
s1	4	8
s2	4.4	5
	Topology A.	

Table 4.3: Utility functions for simulations evaluating the Monotonicity property.

	Property	Objective function $f(x)$
		$\|x\|^2$	$\|x - x^{sh}\|^2$	$\|x - x^{pr}\|^2$	$\|x - v\|^2$	$-v^T x$	$-c^T x$
Fairness	Stability	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
	Efficiency	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
	Eq. treat. of eq.	$\times$	$\sim$	$\sim$	$\sim$	$\times$	$\times$
	Order preserving	$\times$	$\sim$	$\sim$	$\sim$	$\times$	$\times$
	No free riders	$\sim$	$\sim$	$\sim$	$\sim$	$\sim$	$\sim$
	Monotonicity	$\times$	$\times$	$\times$	$\sim$	$\sim$	$\times$

Table 4.4: Summary of the properties provided by the proposed method according to the objective function. ($\checkmark$) fulfilment, ($\times$) no fulfilment, ($\sim$) no counter example found.

alone and being in the alliance is thus not detrimental to it.

As a consequence of the previous simulations we discard the use of weighted sums as objective functions, since as shown above they provide neither a unique solution nor Fairness. Following simulations focus on the objective functions that involve the projection of the Shapley value, the Proportional share, or the contributions vector. In particular we shall evaluate the results when the capacities of the nodes increase, one at a time. In other words, we evaluate the Monotonicity property as defined in Section 4.2.2.

Monotonicity evaluation. Consider Topology A shown in Fig. 4.1a where nodes capacities are equal to one unit. Services of one unit of bandwidth are sold. Consider the utility functions shown in Table 4.3. Results for the different objective functions when varying the capacity of node 3 are shown in Fig. 4.5, along with the normalized contribution vector. Please note that in this case shares are stacked up bottom-top starting from node 3, then node 1 and finally node 2. This allows to easily check the change in node 3's share when its capacity increases.

This simple example shows that $f(x) = \|x - x^{pr}\|^2$, $f(x) = \|x\|^2$ and $f(x) = \|x - x^{sh}\|^2$ do not provide monotonicity. Indeed, as long as node 3 is the bottleneck of both services, its share increases when its capacity increases. Once node 3 is no longer the bottleneck, the increase in its capacity produces, of course, an increase in the total revenue, but however a decrease in its own share. However, the projection of the contributions vector, that is when $f(x) = \|x - v\|^2$, passes this monotonicity test. Further simulative studies showing the fulfilment of the monotonicity property when using $f(x) = \|x - v\|^2$ are provided in Appendix B.

All in all, we have shown that the objective function of the proposed revenue sharing problem, that is of Problem (4.2), has a great influence in the properties fulfilled by the revenue shares obtained by the proposed method. Table 4.4 summarizes the evaluated functions and the obtained results. It is important to say that the results shown in this subsection and summarized in Table 4.4 are supported by further extensive simulative studies, some of them are presented in Appendix B. In particular, those properties that are marked as no counter example found, respond to those extensive simulative studies and not only the cases presented in this chapter.

Results allow to conclude that the projection of the contributions vector, that is considering $f(x) = \|x - v\|^2$, presents the desired properties. Indeed, Stability and Efficiency are verified by
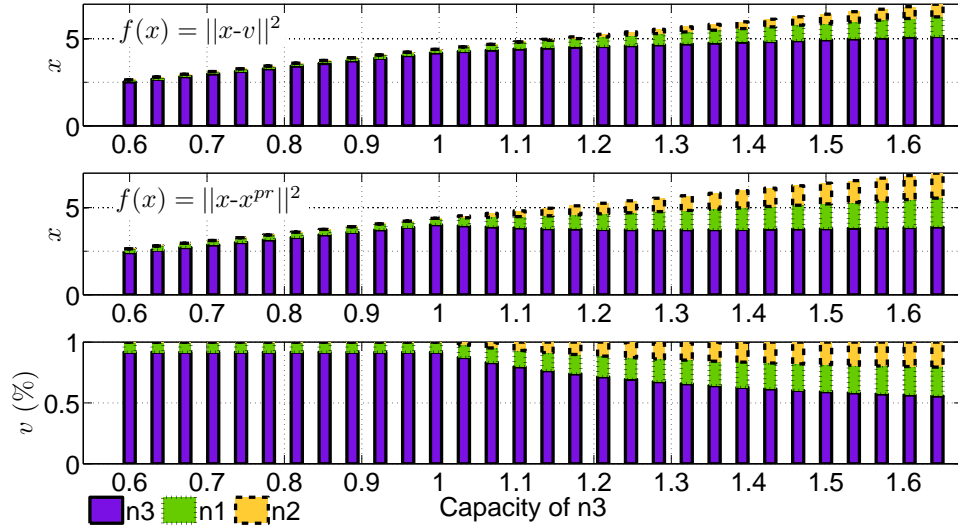
(a) Projection of the contributions vector and the proportional share, along with the normalized contributions vector.
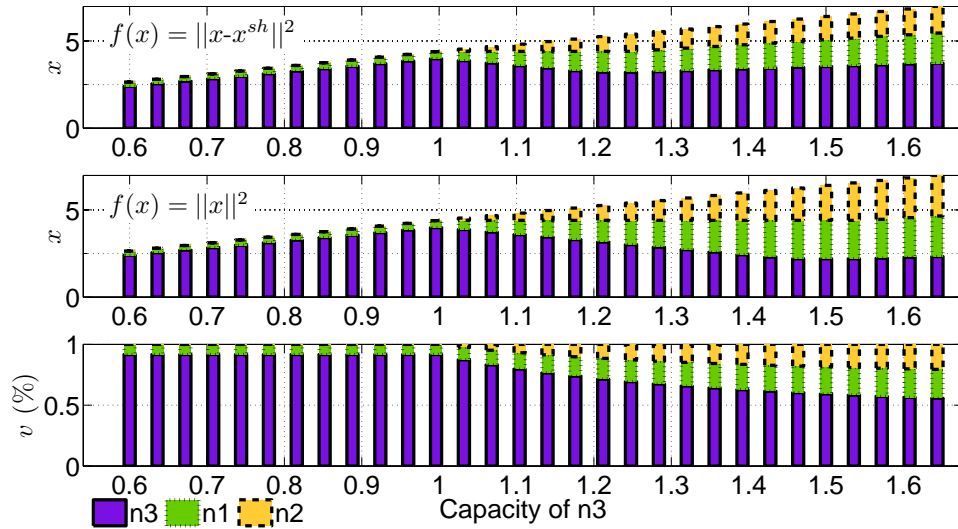
(b) Projection of the Shapley value and minimization of the Euclidian norm of the revenue share vector, along with the normalized contributions vector.

Figure 4.5: Revenue sharing with the proposed method and different objective functions. Evaluation of the monotonicity property.

construction, and Monotonicity and Fairness were verified in a variety of conducted simulative studies.

4.5.2 Multi-period Scenario

We shall now focus on the multi-period scenario, that is to say, when several phases of service selling occur. We shall assume that, provided that the revenue sharing rule assures the discussed properties, NSPs do not leave or enter the alliance during the multi-period. In other words, alliance formation and dissolution occurs in a much longer time-scale. However, the utility functions vary in a much shorter time-scale. A new period thus implies new utility functions, that is different

values for $U_s, s \in S$. This necessarily leads to a different feasible set. Finding on each period a valid revenue sharing vector would involve performing a great number of computations, besides to a great exchange of information among the domains and the central entity solving the revenue sharing. In other words, the multi-period case may pose the problem of scalability thus, we face the challenge of providing a scalable approach. One could naively propose as a solution to compute the revenue sharing once, and then simply keep the sharing proportion for the subsequent revenue sharing phases. However, if we were to use the same proportion for a new service selling instance, then the new revenue sharing vector can lie within the new feasible set or outside of it, which leads us to discard that option. In order to illustrate this situation we present a simple example. Consider Topology D in Fig.4.6. All nodes have capacity of 1 and all services have bandwidth equal to 1. Consider first that utility functions are $U^{(1)} = \{U_s^{(1)}\}$, $s = 1, 2, 3$ with $U_1^{(1)} = U_2^{(1)} = 1$ and $U_3^{(1)} = 5$. The optimal allocation is thus selling all the capacity to service 3 and the total revenue is $V(\{1, 2\}, c, U^{(1)}) = 5$. Besides, $V(\{1\}, c, U) = V(\{2\}, c, U) = 1$ and the contributions vector v is $v = (4, 4)$. The feasible set for this situation is illustrated in Fig. 4.7a and is indicated by the thick line. Computing the revenue shares with the proposed method and with the criterion of projecting the contributions vector v onto the feasible set, we obtain $x = (2.5, 2.5)$, which is indicated with a red cross in Fig. 4.7a. This corresponds to 50% of the revenue for each NSP. Consider now that in a subsequent service selling phase utility functions $U^{(2)}$ are such that $V(\{1, 2\}, c, U^{(1)}) = 7$, $V(\{1\}, c, U^{(2)}) = 1$ and $V(\{2\}, c, U^{(2)}) = 4$. The new feasible set is indicated with a thick line in Fig. 4.7b. If the same revenue share proportion, that is 50% for each NSP indicated by the dotted line in Fig. 4.7b, is kept for this phase, the resulting revenue share lays outside the new feasible set.

Figure 4.6: Topology D.
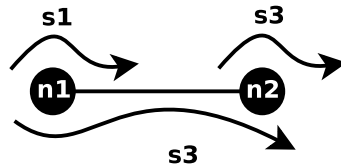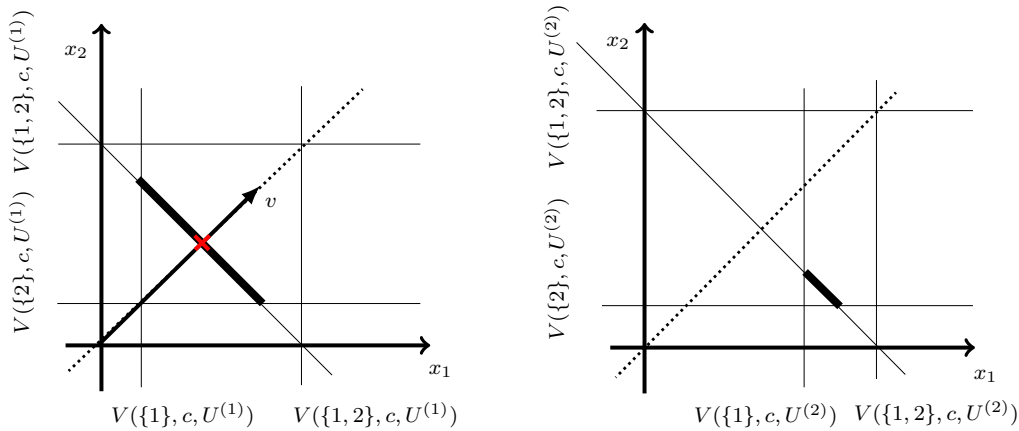

(a) Revenue share with utility functions $U^{(1)}$.   **(b)** The feasible set with utility functions $U^{(2)}$.

Figure 4.7: Example of two subsequent Revenue Sharing phases for Topology D with different utility functions.

All together, in order to provide a scalable solution, we are motivated to perform the revenue sharing on a longer time-scale than the service selling phase, and work with statistics of the utilities received during the several service selling phases considered for a given revenue sharing phase. In the following we shall discuss two different approaches to work with such statistics.

4.5.2.A Approach 1.

In order to model the multi-period situation, let us introduce the assumption that the utility functions are concave random functions. This is the case, for instance, if utility functions are built after the bids received for buying the services through a first price auction mechanism, as the one introduced in Chapter 3. Provided this, we can represent the utility functions of several service selling phases occurred during a certain period of time by their mean over that period of time. As usually, notation E represents the expectation of a random variable. We define the mean utility function as:

$$\overline{U}_s(a_s) = E\{U_s(a_s)\}, \tag{4.10}$$

which is still a non-decreasing concave function of $a_s, \forall s \in S$. Finally, we redefine the revenue function V by Problem 4.3, and call it $\mathcal{V}$.

Problem 4.3

$$\max_a \sum_{s \in S} \overline{U}_s(a_s)$$
$$s.t. Ra \leq c^{\mathcal{Q}}.$$

The procedure then continues as in the one-shot scenario, solving the revenue sharing problem, that is solving Problem 4.2 to chose one point within the feasible set. The difference with the one-shot case is that now in order to compute the feasible set, that is to say to compute $\hat{v}$, function $\mathcal{V}$ instead of V is used.

This approach mechanism makes it possible to perform the computation only once in a while (e.g. monthly). In addition, the amount of information exchanged is also kept small, since the only information that has to be transmitted to the central entity on each revenue sharing phase is the mean of the utilities over that period, that can span many service selling phases. However, a pertinent question is whether this approach still guarantees the desired properties. In particular we shall discussed in the second approach the fulfilment of the stability property.

4.5.2.B Approach 2.

Usually, providers' decisions are based on mid-term or long term behaviours, mainly to keep network stability. Likewise, the interest of the providers to remain in the alliance would be based on its economic stability in the long term. That is, they would likely be interested in remaining in an alliance that is economically attractive in the long term. In order to consider such situation, we compute the long term feasible set, computed based on the expectation of the revenues of each sub-coaliton, and obtain the revenue sharing from such set. This is summarized on Problem (4.4).

Problem 4.4

$$\min_x \ f(x)$$
$$s.t. \ Qx \geq E\{\hat{v}\}, \ e^T x = E\{V(N)\}.$$

Please note that $E\{\hat{v}\}$ is obtained by computing the expectation of the output of the problem defining the revenue function, that is Problem (4.1), for each sub-coalition $\mathcal{Q} \subseteq N$.

The raised question reduces then to answering if the point chosen by Problem (4.3) lies within the feasible set of Problem (4.4) or not. Unfortunately this is not necessary true. Indeed, as shown in

[101], where relationships between stochastic non-linear programming problems are demonstrated, the following inequality applies:

$$E\{V(\mathcal{Q})\} \geq \mathcal{V}(\mathcal{Q}), \forall \mathcal{Q} \subseteq N, \tag{4.11}$$

which means that the feasible set of Approach 2 is contained in the one of Approach 1. However, we have no indication about the tightness of the bound, thus we shall evaluate the impact of using either of both approaches by simulation, in the following section.

Please note that Approach 1 has a lower computational complexity than Approach 2. Indeed, in Approach 1, $2^{|N|}$ NUM problems must be solved in order to determine the feasible solutions set. On the other hand, Approach 2 needs to solve, for each selling phase, $2^{|N|}$ NUM problems in order to determine the feasible solutions set. The following subsection presents simulation studies that compare the results obtained using both approaches and evaluates the computational time consumed by both approaches.

4.5.2.C   Multi-Period Simulations

We now compute the solution according to Approach 1 and Approach 2. In both cases, a number of 50 service selling phases were performed before a revenue sharing (RS) phase and the projection of the contributions vector v was used as criterion. Simulations were performed on a regular computer with a i5 processor of 2.67GHz and 3.6 GB of RAM memory.

Results for Topology 2 are shown in Fig. 4.8a. For this topology, on every RS phase the results obtained using both approaches are almost the same. Same thing occurs for all the simulations performed, in particular for the one over Topology 3, whose results for selected nodes are shown in Fig. 4.8b.

Further simulative studied are available in Appendix B, where the behaviour seen in this Chapter are as well observed. In particular, for all the considered topologies and utility functions both approaches provide very similar results.

We now evaluate the computation time consumed by each approach. We shall consider a simple topology with only one service defined, as Topology B illustrated in Fig. 4.1b, and linearly increase the number of nodes in the service's path. Since NSPs alliances are likely to have no more than 10 nodes, considering, for instance, that the average AS path in the Internet is of 4 ASs [55], we have increased the number of nodes up to 8 nodes, and evaluated if at that scale the computation time is affordable.

Results show that for both approaches the time consumed by the method increases exponentially with the number of nodes in the network. This is related to the Stability property, since for taking it into account we consider all sub-coalitions of nodes (i.e. $2^{|N|}$ cases). For a topology of 8 nodes, Approach 1 consumed on the average 2 ms while Approach 2 consumed 135 ms. However, Approach 2 is still feasible, moreover considering it is proposed to be performed off-line and in a long time-scale. Besides, the computation of the feasible set can be performed in a completely parallel fashion.

All in all, we can claim that Approach 2 provides with a solution that fulfils the sought properties with affordable computation time.

$$f(x) = ||x-v||^2$$
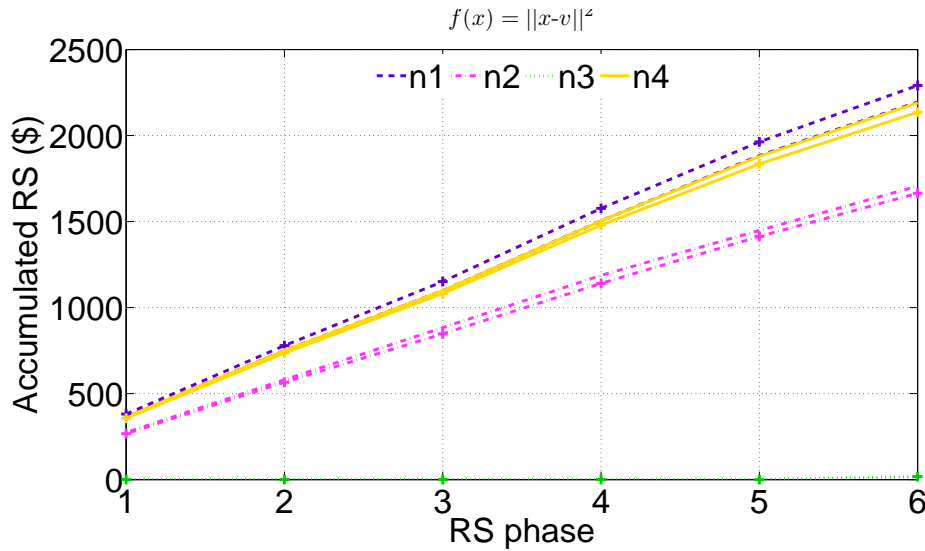
(a) Results for Topology 2.

$$f(x) = ||x-v||^2$$
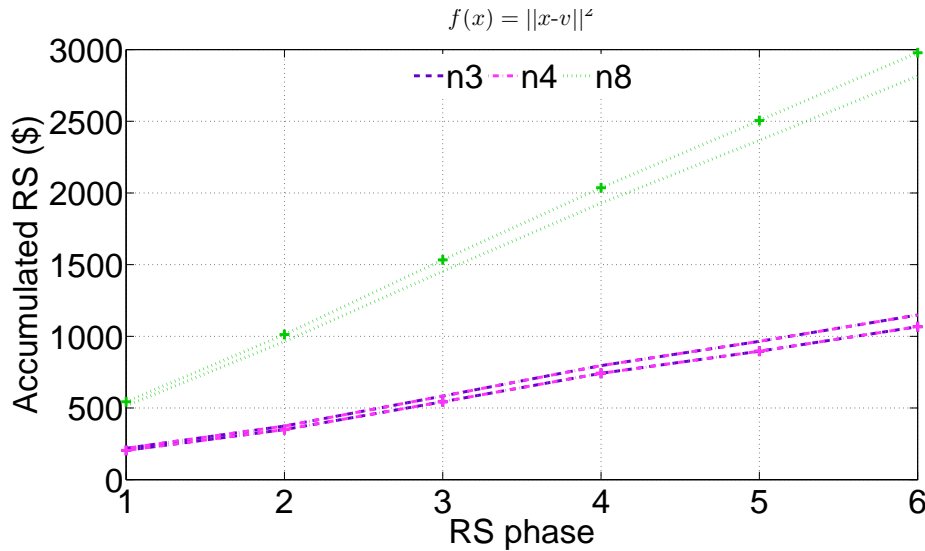
(b) Results for Topology 3.

Figure 4.8: Comparison of the accumulated revenue share for each NSP when using Approach 1 (-) and Approach 2 (+).

4.6 Summary

In this chapter we have addressed the problem of revenue sharing in the context of NSP alliances. We have focused on the case where the income of the alliance is determined by the output of a NUM problem. This particular scenario poses new challenges. Indeed, previous results for performing revenue sharing were found to be inappropriate applied to this case.

The desired properties for the revenue sharing in an NSP alliance have been formally stated and a new method has been proposed. This method is conceived for providing economic stability and efficiency to the alliance and it is flexible enough to be adapted to fulfil additional properties. The method is based on solving optimization problems and considers statistics on the income. In addition, implementation concerns have been discussed and scalability has been resolved through a centralized long-scale revenue sharing phase. Two different approaches were discussed for working with statistics of the income. The method's proper behaviour has been evaluated through extensive

simulation studies showing both to fulfil further properties as incentives to the NSPs to increase
their resources and fairness, and to run in affordable time.

Chapter 5

Feedback from Network Monitoring to the Business Plane: A Pricing Scheme based on First-price Auctions with Reimbursement

5.1 Introduction

As we have mentioned in Chapter 1, an expansion in service types and quality levels is expected in the near future [21]. Tele-presence, tele-medicine, online gaming and teleconferencing are a few examples of future enhanced services. In order to provide these services, in addition to intrinsic networking requirements such as scalability, confidentiality and technical aspects, market implications and customers' behaviour must be taken into account. Therefore, holistic and interdisciplinary approaches are needed. Such approaches have enriched the Networking and Internet Economics research fields over the past few decades, as recent research results show. The proposals range from interdomain Quality of Service (QoS) path composition and Service Level Agreement negotiations, such as in [37, 122], to higher layers issues, such as modelling user reactions to changes in Internet pricing [117] or net-neutrality analysis, see for instance [25]. The enlargement of the service offer also aims to create new market opportunities and target different kinds of user profiles. The real space for this new market has been identified as an issue to be studied, along with how users are expected to react to it. In this regard, quality of experience (QoE) and its influence on willingness to pay has gained importance and has put the end user back in stage [128, 136].

In this context, traditional flat rates, where end users pay a single fee for Internet access regardless of usage, have to be revisited, not in order to eliminate them, but rather to identify enhanced services where special pricing (per-service, per-amount-of-bandwidth, per-level-of-quality, etc.) could be needed. Moreover, the mere existence of services with enhanced quality presupposes differentiated pricing, since otherwise every user would choose the highest level of quality, which is sustainable from neither a technical nor an economic point of view. In this regard, several pricing schemes for enhanced network services have been proposed (see [53] for a survey), including some based on QoS [126, 156].

The justification of new pricing schemes for enhanced services is quite unquestionable from the point of view of the network service providers. But would buyers accept differentiated pricing? Our intuition is that they would be ready to pay more for services that are assured without question to be delivered in high quality.

On the other hand, in today's networks, failures, while less and less frequent, still do occur, producing a negative effect on buyers' willingness to pay. Moreover, several studies have shown experimentally that user satisfaction has a positive effect on willingness to pay (e.g. [70]). A failure in this context could account for a QoS threshold violation, such as a bound on the delay, a jitter value, or even a service interruption. Intuition also says that, while potential failures have a negative impact on willingness to pay, reimbursement should have a positive one.

It is in this context that we propose a pricing scheme for assured-quality service selling where buyers can make bids to obtain a quality-assured bit pipe, henceforth referred as to the service, and can be reimbursed if ultimately the service fails. We shall show that this reimbursement scheme, under the main assumption of symmetric buyers with private values, incentivizes buyers not to decrease their willingness to pay due to possible failures, which in the end results in an increase in the expected seller's revenue. Moreover, we show that reimbursing 100% overcomes problems like the *market for lemons* and *moral hazard*, which we show would arise when rational buyers are uncertain about service performance.

In particular, a first-price sealed auction mechanism is proposed for selling services. Auctions make it possible both to find the market price of services that are not yet widely deployed, since services' market price is revealed as part of the mechanism, and to have guidelines to model willingness to pay.

In our framework, the seller could be, for example, an alliance of domains who sells a pipe for transit traffic with guaranteed quality. Buyers could be, for instance, other domains who need to buy transit, or content providers who need bandwidth with quality guarantees in order to properly deliver their enhanced services. Hereafter, we shall simply refer to sellers and buyers.

The proposed pricing scheme assumes the existence of a monitoring infrastructure, which would be the one triggering the reimbursement process. The research and industry community has somehow agreed that for QoS provisioning a monitoring infrastructure is essential. Examples of this are found in recent projects (e.g. [4]), in recent standardisation activities (e.g. [18]), and a wide variety of conducted research, for instance [40, 137]. In this sense, our pricing scheme proposes using the existence of the monitoring infrastructure at the business plane, providing an economic justification for monitoring.

The remainder of this chapter is organized as follows. In Section 5.2 we review the related literature. In Section 5.3 we clearly state the model, assumptions and definitions required. In Section 5.4 we study buyers' willingness to pay in the context of first-price auctions, which is given by so-called best bidding strategies. In Section 5.5, we study the problem from the seller's standpoint in order to derive the best percentage of reimbursement. In particular we present the pricing game modelled through a Stackelberg game. Finally, a summary of the chapter is provided in Section 5.6. This chapter is based on [28].

5.2   Related Work

In Chapter 3 we have addressed the bandwidth allocation problem and presented bandwidth auctions as mechanisms that make it possible to determine the price of the services in sale. In particular the use of first-price auctions mechanisms was proposed. A review of related work was presented and supporting reasons for the choice of this particular mechanism were given in Section 3.5.2. In this section we shall focus on work related to the specific topics of this chapter: reimbursement schemes and pricing games. Economic literature provides a vast yet fertile field for mechanisms based on first-price auctions. Equilibrium or best bidding strategies have been studied under different assumptions. The basic bidding model was introduced in [149], where results for equilibrium bidding strategies with independent private values were shown for valuations drawn from a uniform distribution. More detail was later provided in [132]. Further results relaxing some assumptions were derived in [92, 130] and instructive and complete summaries can be found in [84] and [109]. However, none of these results consider either failures or reimbursements.

Moreover, regarding reimbursement policies, the literature closely related to our scenario is rather limited. Nevertheless, some related proposals have been made. Perhaps the closest work is that one proposed by Tuffin *et al.* in [145]. In such work, a simple pricing model for communication networks is presented in which reimbursement occurs if a certain delay threshold is exceeded. Prices are fixed by the seller such that for a given amount of reimbursement, his or her own revenue is maximized. The authors model demand such that it is proportional to the probability that the utility exceeds a given cost. This cost is a function of the price paid, the cost of waiting and the negative cost in case the delay threshold is exceeded, which corresponds to a reimbursement.

A certain shape for the utility's probability distribution function is assumed in order to draw conclusions and perform simulations. The authors show through simulations that this mechanism increases the seller's revenue compared to the case with no reimbursement. The idea behind this method is the same as ours, though buyers' side is modelled in a very different way, since in that work the price is fixed by the seller, while in our work the price is determined through the auction mechanism.

Yet another approach is proposed in [46], where second-price auctions are used for buying one unit of a computing resource. Winning buyers pay and with a certain probability will indeed need the service. If winners do not use the service, they receive a refund of a percentage of the payment. The authors propose a simulative approach to determine the best refund strategy, and define a correlation between the valuation of the object and the probability of not using it. They conclude that different correlations result in different percentages of optimal reimbursement. The scenario is quite different to ours, but the logic behind the mechanism is very similar to our proposal. However, in this chapter we focus on analytical results, both for computing the dependency of the willingness to pay on the probability of failure and reimbursement, and to compute the seller's revenue and optimal percentage of reimbursement. In addition, we focus on first-price auctions.

Other reimbursement schemes have also been studied in another, non-networking context. A particular case is services that have the peculiarity that they can be returned afterwards and will still possess some value for the seller (see for instance [102]). This is not the case in the network scenario, in which the service no longer has any value to the seller once it has been used. Another similar widely studied case is that one of assurances, which as well differs significantly from our context.

With respect to the study of the optimum percentage of reimbursement, we shall model the pricing game through a Stackelberg game, which are very suitable for modelling pricing situations, where the network or the seller typically acts as *leader* and the users or buyers act as *followers*. This kind of game has been widely used in the literature to design revenue-maximizing network policies. For example, [39,139] where an Internet packet-pricing scheme is proposed for monopolistic service providers and large numbers of users, or [26], where a pricing scheme for differentiated services is proposed, or yet [144], where a user loyalty model to Internet service providers is proposed and applied in a game-theoretical framework in order to derive optimal Internet access pricing strategies. In addition, this kind of hierarchical game has also been studied for pricing along with power control in wireless networks, see for instance [24,61,153] and for spectrum sharing in such networks [152].

5.3 The model

Let us begin by describing our working scenario and introducing the notations, definitions and assumptions needed to model it. We are studying a situation where quality-assured services are sold over an interdomain network. Such services could be, for instance, video on demand, a VPN service interconnecting two remote sites or a network game. In all cases, the service can be abstracted to a certain amount of bandwidth guaranteed between two sites through an overlay network, and with certain quality parameters associated with it. We shall call this abstraction an *object*. The quality parameters associated with the object could be given, for instance, by values of the delay, the jitter, the percentage of packet failures, the percentage of service availability, etc.

Objects are sold via a first-price sealed auction mechanism. The following assumptions are made regarding this mechanism. We first assume a single-object case, that is to say that M bidders or buyers compete to buy one object. We then move to the case of multi-object, single-unit demand. In other words, M bidders compete to buy K identical objects, and each bidder is interested in buying one single unit of such objects. Each bidder i assigns a valuation X_i to the object and we assume that the X_is are independently and identically distributed according to a common distribution function F. This is the so-called symmetric model, since all bidders' valuations are distributed according to the same distribution function. At the moment of bidding, bidder i knows the realization x_i of his or her valuation but does not know the valuation attached to the object by

other bidders, and this knowledge would not affect his or her own valuation, which is the so-called private values model.

Conversely, we assume that the service has no value to the bidder if it fails. Please note that actually bidders could attach a negative value to the service when it fails, rather than a null one. This would be the case, for instance, if the failure causes losses to the buyer's business. This could be easily modelled by considering a negative deterministic valuation in case of failure. For clarity's sake, we shall not consider this artefact in the model, though doing so would not change the methodology applied to address the problem.

Bidders are assumed to be risk-neutral, as they seek to maximize their expected profits. Bidder i's bid is denoted by b_i and it is obtained according to a bidding strategy called β_i. That is to say, bidder i's bid is determined as $b_i = \beta_i(x_i)$. Finally, we assume a discriminatory payment rule, which means that the winning bidder pays his bid. We shall generally simplify notation and refer to x as the realization of the valuation of any given bidder.

The service has an associated probability of failure, denoted by θ in our framework. If indeed the service fails, money is given back. The amount of money returned is proportional to what has been paid for the object and the coefficient of proportionality is represented by $q \geq 0$, which could *a priori* be greater than one. We shall as well refer to q as a percentage of reimbursement.

The percentage of reimbursement associated with the object is always announced to the bidders before they announce their bids, and it is the same value for all bidders. Bidders have their own estimations regarding the probability of failure of a service, which will have an impact on the value of the bid they submit. This estimation could, *a priori*, be based on service performance perception. Buyers could even perform their own measurements on historical observations in order to estimate the probability of failure, or they could infer it from the percentage of reimbursement announced. But we shall address this issue later on. For the moment let us denote the probability of failure assumed by the bidders at the moment of placing their bid as $\tilde{\theta}$, which is not necessarily equal to θ.

5.4   The Optimal Bidding Strategy

In order to determine how the willingness to pay for a service is affected by the probability of failure $\tilde{\theta}$ and the percentage of reimbursement q, we study the optimal bidding strategy under such conditions, under the assumptions of Section 5.3, first for a single object on sale and then for multiple objects.

5.4.1   The Single-Object Case

The single-object case models the situation in which the total available capacity, along with quality guarantees, is to be allocated to one single client, i.e. to the winning bidder. We shall show that in this case and under the assumptions of Section 5.3, a symmetric equilibrium exists, that is to say an equilibrium where all bidders adopt the same best strategy. Theorem (5.1) formally states this along with the mathematical expression for the best bidding strategy.

Theorem 5.1 *The Symmetric Equilibrium, Single Object Case. Given a set of M symmetric bidders whose valuations X_i, $i = 1 \dots M$ are identically and independently distributed (i.i.d.) from a probability distribution $F(x)$, the bidding strategy that maximizes each bidder's payoff in a first-price sealed auction mechanism for a single object which is assumed to fail with probability $\tilde{\theta}$ and for which a percentage q of the amount paid is given back if it actually fails, is the same for all bidders and is given by:*

$$\beta(x) = \frac{1 - \tilde{\theta}}{1 - q\tilde{\theta}} E[Y_{M-1}^{(1)} | Y_{M-1}^{(1)} \leq x], \ (\tilde{\theta}, q) \in D, \tag{5.1}$$

where $Y_{M-1}^{(1)}$ is a random variable defined as the maximum over $M - 1$ i.i.d. random values from distribution F and $D = \{\tilde{\theta} \in [0, 1), q \geq 0 : q\tilde{\theta} < 1\}$.

Proof Let us first assume that a symmetric equilibrium exists, meaning that all bidders follow the same strategy, $\beta_i = \beta$, $i = 1 \ldots M$. Any bidder's payoff $\tilde{P}$ can thus be expressed as a function of his or her bid b as in Equation (5.2), where $\beta(x) = b$ and $\mathbb{1}_e$ is equal to 1 if event e occurs, and 0 otherwise.

$$\tilde{P} = \mathbb{1}_{win}(x\mathbb{1}_{not\,failure} - b(1 - q\mathbb{1}_{failure})), \tag{5.2}$$

Now let G be the cumulative distribution function of the maximum valuation over $M - 1$ valuations i.i.d. according to F, which we shall denote as $Y_{M-1}^{(1)}$. The notation in $Y_{M-1}^{(1)}$ means that we are selecting the highest value, indicated by superscript 1, among a sample of size $M - 1$, indicated by subscript $(M - 1)$. Please note that $Y_{M-1}^{(1)}$ is then the $(M - 1) - th$ order statistics of a sample of $(M - 1)$ i.i.d. values according to F. The expectation of a bidder i's payoff can be expressed as:

$$E\{\tilde{P}|X_i = x_i\} = G(\beta^{-1}(b_i))(x_i(1 - \tilde{\theta}) - b_i(1 - q\tilde{\theta})). \tag{5.3}$$

In Equation (5.3) we have used the fact that the probability of winning the auction is

$$Pwin(b_i) = \mathbb{P}(b_i > \max_{j \neq i} b_j) = \mathbb{P}(\beta(x_i) > \max_{j \neq i} \beta(X_j))$$
$$= \mathbb{P}(x_i > \max_{j \neq i} X_j) = G(x_i), \tag{5.4}$$

where symmetric equilibrium is assumed, and in the last equality, the assumption is made that β is a strictly increasing function of x. Please note that in the reasoning above we have used $\tilde{\theta}$ and not θ, since we are looking at the problem from the buyer's point of view. Since the previous reasoning is valid for any bidder, in what follows subscript notation is avoided.

Finding the bidding strategy β that maximizes Equation (5.3) reduces to setting its derivative with respect to b equal to zero and imposing $b = \beta(x)$. The derivative of the expected payoff with respect to b is shown in Equation (5.5), where we have introduced the notation $g(x) = G'(x)$ and where we have applied the well-known formula for the derivative of the inverse function.

$$\frac{g(\beta^{-1}(b))}{\beta'(\beta^{-1}(b))}(x(1 - \tilde{\theta}) - b(1 - q\tilde{\theta})) - G(\beta^{-1}(b))(1 - q\tilde{\theta}) = 0, \tag{5.5}$$

Under the assumption of symmetric equilibrium $\beta^{-1}(b) = x$ holds. Applying this equality to Equation (5.5) we obtain:

$$xg(x)(1 - \tilde{\theta}) - g(x)\beta(x)(1 - q\tilde{\theta}) - G(x)\beta'(x)(1 - q\tilde{\theta}) = 0 \tag{5.6}$$

The study must be divided into two cases, namely $q\tilde{\theta} < 1$ and $q\tilde{\theta} \geq 1$. We assume as well that $\beta(0) = 0$.

$q\tilde{\theta} < 1$. First consider $\tilde{\theta} \neq 1$. In this case Equation (5.6) can be rewritten as:

$$\beta'(x) + \beta(x)\frac{g(x)}{G(x)} - x\frac{g(x)}{G(x)}\frac{1 - \tilde{\theta}}{1 - q\tilde{\theta}} = 0 \tag{5.7}$$

whose solution is

$$\beta(x) = -e^{-S(x)}\int_0^x e^{S(z)}T(z)dz, \tag{5.8}$$

where:

$$S(x) = \int_0^x \frac{g(z)}{G(z)}dz = \log G(x) \text{ and } T(x) = -x\frac{g(x)}{G(x)}\frac{1 - \tilde{\theta}}{1 - q\tilde{\theta}}.$$

Hence, operating we obtain:

$$\beta(x) = \frac{1}{G(x)} \int_0^x zg(z)dz\frac{1-\tilde{\theta}}{1-q\tilde{\theta}}$$

$$= E[Y_{M-1}^{(1)}|Y_{M-1}^{(1)} \leq x]\frac{1-\tilde{\theta}}{1-q\tilde{\theta}}, \tag{5.9}$$

where the last equality comes directly from the definition of conditional expectation. Consider now $\tilde{\theta} = 1$. Equation (5.5) reduces to

$$\frac{g(\beta^{-1}(b))}{\beta'(\beta^{-1}(b))}(-b(1-q\tilde{\theta})) - G(\beta^{-1}(b))(1-q\tilde{\theta}) = 0, \tag{5.10}$$

which results in

$$\frac{-g(x)}{G(x)} = \frac{\beta'(x)}{\beta(x)}. \tag{5.11}$$

Integrating Equation (5.11) on both sides we obtain

$$\beta(x) = \frac{\kappa}{G(x)}, \tag{5.12}$$

where κ is a real constant of integration.

In order to verify whether the assumption of symmetric bidding functions holds, we suppose, without loss of generality, that all bidders but one bid with the same optimal bidding function found above. We shall check if it is also optimal for the remaining bidder to bid according to this function.

Consider first the case of $\tilde{\theta} \neq 1$. Bidder 1's expected payoff ($\tilde{P}_i$) if he or she bids $\beta(z)$ when his or her value is actually x is:

$$\tilde{P}_i(\beta(z), x) = G(z)(x(1-\tilde{\theta}) - \beta(z)(1-q\tilde{\theta})). \tag{5.13}$$

Hence, the difference with the bidder's expected payoff if bidding $\beta(x)$ is:

$$\tilde{P}_i(\beta(z), x) - \tilde{P}_i(\beta(x), x) =$$
$$G(z)(x(1-\tilde{\theta}) - \beta(z)(1-q\tilde{\theta}))-$$
$$G(x)(x(1-\tilde{\theta}) - \beta(x)(1-q\tilde{\theta}))$$
$$= (1-\tilde{\theta})x(G(z)-$$
$$- G(x)) + (1-\tilde{\theta})\int_z^x vg(v)dv$$
$$= (1-\tilde{\theta})x(G(z)-G(x)) + (1-\tilde{\theta})\left[G(v)v \mid_z^x - \int_z^x G(v)dv\right]$$
$$= (1-\tilde{\theta})\left[G(z)(x-z) + \int_x^z G(v)dv\right], \tag{5.14}$$

where we have used integration by parts. It remains to be determined if Equation (5.14) is negative for any value of z. If it were to be negative, then bidder 1's expected payoff by bidding something different from $\beta(x)$ would not be greater than the payoff obtained if he were to bid $\beta(x)$.

We graphically show that indeed, Equation (5.14) is strictly negative $\forall z \neq x$. Consider Fig. 5.1, where two different cases are distinguished. When $z < x$, as illustrated in Fig. 5.1a, the shaded
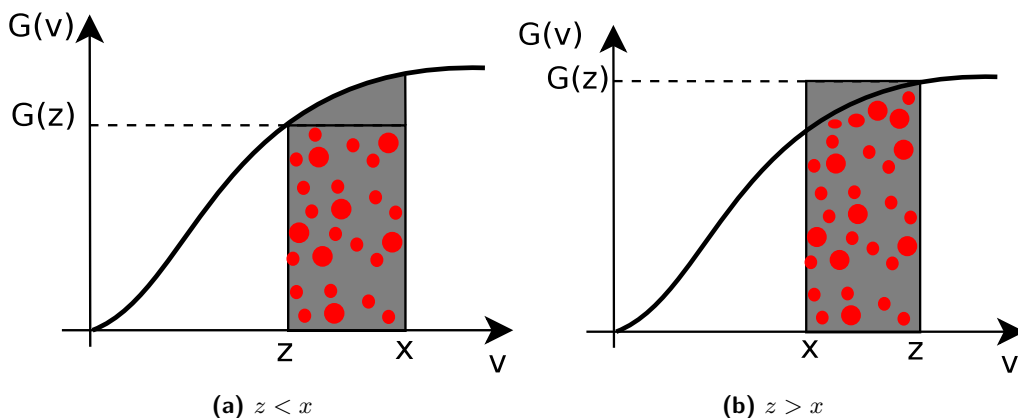
(a) $z < x$ **(b)** $z > x$

Figure 5.1: Illustration of the components of Equation (5.14).

area corresponds to $-\int_x^z G(v)dv$ and the dotted area corresponds to $G(z)(x - z)$. The shaded area is greater than the dotted area for any G (we recall that G is a distribution function, thus it is always non-decreasing). Hence, Equation (5.14) is negative for any $z < x$.

Likewise, when $z > x$, which corresponds to the situation in Fig. 5.1b, the dotted area corresponds to the term $\int_x^z G(v)dv$ and the shaded area to the opposite of $G(z)(x - z)$. The shaded area is always greater than the dotted one, thus Equation (5.14) is again strictly negative. All together, we obtain that Equation (5.14) is negative for any value of $z \neq x$, which means that there is no other bid leading to a greater payoff than bidding $\beta(x)$. We conclude that the assumption of symmetric equilibrium holds for the case $(\tilde{\theta}, q) \in D$.

Consider now the case $q < 1$ and $\tilde{\theta} = 1$. β is given by Equation (5.12), where κ is such that $\beta(0) = 0$. But since $G(0) = 0$, it follows that there is no value of κ, either than the trivial $\kappa = 0$, verifying $\beta(0) = 0$. We have obtained a contradiction, hence we conclude that the assumption of a symmetric equilibrium is not possible, and no equilibrium exists for this case.

$q\tilde{\theta} \geq 1$. In this case we notice that the expected payoff, given in Equation (5.3), is always increasing with b, thus there is no equilibrium. $\square$

5.4.2 Multi-Object, Single-Demand Case

The multi-object scenario corresponds to the case where the available capacity is split into several identical bandwidth chunks, each with certain quality guarantees, and each to be assigned to a different client. Single demand means that each client is interested in exactly one of these chunks or services. Let us assume that the auction mechanism is launched for selling K objects. The only thing that changes with respect to the single-object case is the probability of winning the auction, which in this case corresponds to the probability of being among the K highest bids.

Theorem 5.2 *The Symmetric Equilibrium, Multiple-Object Single-Demand Case. Consider a set of M symmetric bidders whose valuations X_i, $i = 1 \ldots M$ are i.i.d. from a probability distribution $F(x)$. Consider a mechanism implemented through a first-price sealed auction for selling K identical objects, which are assumed to fail with a probability $\tilde{\theta}$ and for which a percentage q of the amount paid is given back if the service actually fails. Consider that each bidder is interested in a single object.*

In such conditions, the bidding strategy that maximizes each bidder's payoff is the same for all the bidders and is given by:

$$\beta(x) = E[Y_{M-1}^{(K)}|Y_{M-1}^{(K)} \leq x]\frac{1 - \tilde{\theta}}{1 - q\tilde{\theta}}, \tilde{\theta}, q \in D, \tag{5.15}$$

where $Y_{M-1}^{(K)}$ is a random variable defined as the K-th highest value among $(M-1)$ i.i.d. values according to F, and where $D = \{\tilde{\theta} \in [0,1), q \geq 0 : q\tilde{\theta} < 1\}$.

Proof Since β is assumed to be an increasing function of x_i, and we assume a symmetric equilibrium, i.e. $b_i = \beta(x_i) \, \forall \, i$, the probability of winning for bidder i is equal to the probability that his valuation x_i is among the K highest valuations.

Considering this probability of winning the auction the proof is analogous to the single object case. Hence, the symmetric equilibrium for the optimal bidding strategy is readily generalized to the multi-object scenario and is given by Equation (5.15). □

5.4.3 Bidding Behaviour Remarks

The best bidding strategy deserves a closer look. First, it is interesting to note that it follows some intuition. Indeed, according to Theorem (5.1) and Theorem (5.2), β increases when the percentage of reimbursement does so, and when the percentage of reimbursement is less than 100%, β decreases when the probability of failure increases. This means that buyers decrease their bids when they assume that services fail frequently unless the percentage of reimbursement is greater than or equal to 100%, which is quite intuitive. It means as well that for the same level of failures, the higher the percentage of reimbursement, the higher the bid, which is also consistent with intuition.

This behaviour is shown in Fig. 5.2 for different values of $(\tilde{\theta}, q) \in D$, where α, defined as $\alpha = \frac{1-\tilde{\theta}}{1-q\tilde{\theta}}$, is the multiplying factor on the best bidding strategy.

Let us finalize this section with two illustrative examples.

Buyers with uniformly distributed valuations. Consider a situation where there are M buyers whose valuations are uniformly distributed on $[0,1]$ and there is a single service for sale. In this case $F(x) = x$, $G(x) = x^{M-1}$ and the optimal bidding strategy is given by Equation (5.16), which is valid for $(\tilde{\theta}, q) \in D$.

$$\beta(x) = \frac{1-\tilde{\theta}}{1-q\tilde{\theta}} \frac{M-1}{M} x. \tag{5.16}$$
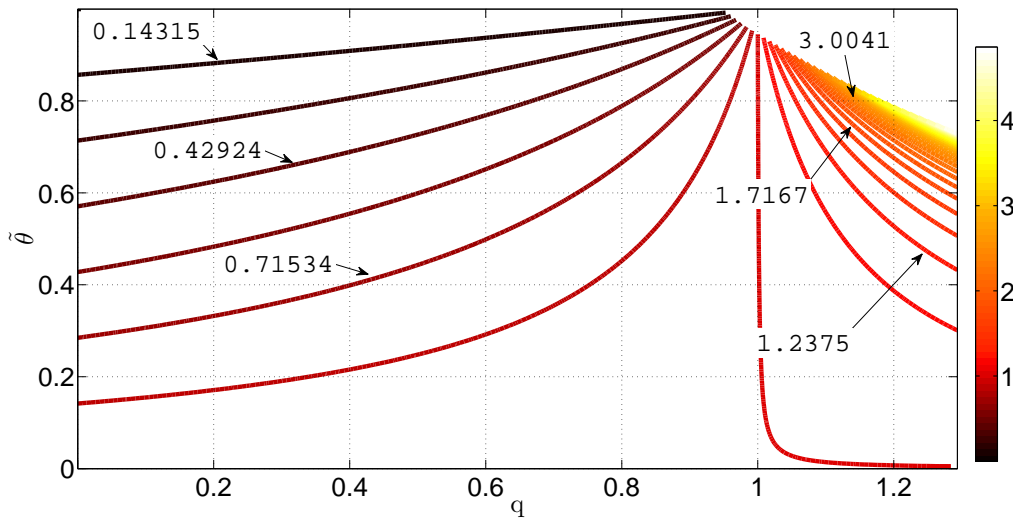

Figure 5.2: Contour Lines of function α, the multiplying factor of the best bidding strategy, for different assumed probabilities of failure $\tilde{\theta}$ and percentage of reimbursement q.

It is worth noting that for different values of q and $\tilde{\theta}$, bidders either shade their valuations, announce their true valuation, or overbid. We say that they shade their valuations when the bid is smaller than the valuation, as occurs in the case with no reimbursements and no failures. Conversely, we say that bidders overbid when their bids are greater than their valuations. In contrast, in cases with no failures, bidders always shade their bids.

Two buyers with exponentially distributed valuations.   Suppose that we have two bidders whose valuations are i.i.d. from an exponential distribution of parameter λ. The best bidding strategy can be directly obtained by considering $F(x) = 1 - exp(-\lambda x)$ and the result of Theorem (5.1).

We first observe that $\beta(x)$ can be rewritten as

$$\beta(x) = \frac{1 - \tilde{\theta}}{1 - q\tilde{\theta}} \left(x - \int_0^x \frac{G(z)}{G(x)} dz \right) \tag{5.17}$$

Indeed, let us, again, for convenience note $\alpha = \frac{1 - \tilde{\theta}}{1 - q\theta}$.

$$\begin{aligned}
\beta(x) &= \alpha \frac{1}{G(x)} \int_0^x zg(z)dz \\
&= \frac{\alpha}{G(x)} \left(G(x)x - \int_0^x zg(z)dz \right) \\
&= \alpha \left(x - \int_0^x \frac{G(z)}{G(x)} dz \right),
\end{aligned} \tag{5.18}$$

where the second equality is obtained by integrating by parts.

Since we have only two bidders, $G(x) = F(x)$. Thus

$$\begin{aligned}
\beta(x) &= \alpha \left(x - \int_0^x \frac{1 - exp(\lambda z)}{1 - exp(\lambda x)} dz \right) \\
&= \frac{\alpha}{\lambda} - \frac{\alpha x \cdot exp(-\lambda x)}{1 - exp(-\lambda x)}
\end{aligned} \tag{5.19}$$

5.5   Expected Seller's Revenue

We shall now study the problem from the seller's standpoint, with the ultimate objective of finding the optimal value of the percentage of reimbursement q. For these purposes, we shall study the seller's outcome. Since there are some uncertainties at each transaction, namely whether the service will fail or not, as well as the intrinsic uncertainty of the selling price due to the auction mechanism in place, we shall model the seller's outcome through his or her expected revenue. We recall that there are K units of the same object for sale. There are $M \geq K$ bidders, who participate in a first-price auction to obtain one object. Let us order their bids as:

$$b^{(1)} \geq b^{(2)} \geq \cdots \geq b^{(M)}, \tag{5.20}$$

which are obtained as $b^{(i)} = \beta(x^{(i)})$, where $x^{(i)}$ represents the ordered bidders' valuation and β is given by Theorem (5.2).

The services are allocated to the K highest bids; thus the money paid can be expressed as a

function of K and the valuations $\mathbf{x} = (x^{(1)}, \ldots, x^{(M)})$ as:

$$
\begin{aligned}
I(K, \mathbf{x}) &= \sum_{i=1}^{K} b^{(i)} = \sum_{i=1}^{K} \beta(x^{(i)}) \\
&= \sum_{i=1}^{K} E[Y_{M-1}^{(K)} | Y_{M-1}^{(K)} \le x^{(i)}] \frac{1 - \tilde{\theta}}{1 - q\tilde{\theta}}.
\end{aligned}
\tag{5.21}
$$

In order to simplify the notations henceforth let us introduce function $u(K, \mathbf{x})$ defined as:

$$
u(K, \mathbf{x}) = \sum_{i=1}^{K} E[Y_{M-1}^{(K)} | Y_{M-1}^{(K)} \le x^{(i)}].
\tag{5.22}
$$

According to our proposed pricing scheme, if failures take place the seller will give money back. Let us assume that failures occur for all service at the same time. This model accounts for failures given by an equipment failure or congestion for example. Then the seller's revenue given that the bidders' valuations are $\mathbf{x} = (x^{(1)}, \ldots, x^{(M)})$, is:

$$
\begin{aligned}
&R(K, \mathbf{x}) \\
&= I(K, \mathbf{x}) \mathbb{1}_{no\,failure} + (1 - q) I(K, \mathbf{x}) \mathbb{1}_{failure}
\end{aligned}
\tag{5.23}
$$

Since valuations and failure events are independent, the mean seller's revenue given the bidders' valuations $\mathbf{x}$ is:

$$
\begin{aligned}
&\bar{R}(K, \mathbf{x}) \\
&= I(K, \mathbf{x})(1 - q\theta) = u(K, \mathbf{x}) \frac{1 - \tilde{\theta}}{1 - q\tilde{\theta}} (1 - q\theta).
\end{aligned}
\tag{5.24}
$$

Please note that if we were to assume that failures do not occur for all objects at the same time, it is easy to check that we would also obtain Equation (5.24) for the expected seller's revenue.

Finally, the *ex ante* expected seller's revenue, that is the seller's expected revenue taking into account the randomness of valuation's vector $\mathbf{X}$, is obtained as

$$
E\{\bar{R}(K, \mathbf{X})\} = E\{u(K, \mathbf{X})\} \frac{1 - \tilde{\theta}}{1 - q\tilde{\theta}} (1 - q\theta),
\tag{5.25}
$$

where the expectation is over the valuations X_i, and $\mathbf{X} = (X^{(1)}, \ldots, X^{(M)})$ is the vector of valuations X_i, $i \in \{1, \ldots, M\}$ sorted in non-increasing order.

Hence, the seller's expected revenue can be tuned through the value of q. However, the value of $\tilde{\theta}$, which also influences the seller's expected revenue, is determined by the buyers. We shall divide the following study into three parts, each of which makes different assumptions about the buyers' behaviour.

5.5.1   Complete Information

We shall first assume that buyers have complete information about the services' performance. The information can be obtained, for instance, through a monitoring infrastructure available for buyers' consultation, or from knowledge obtained through previous observations. In our model, this situation is translated into $\tilde{\theta} = \theta$. We shall refer to this scenario as complete information because it supposes that seller and buyers have the same information regarding the probability of failures.

Hence, the expected seller's revenue becomes, regardless the value of q, equal to:

$$E\{\bar{R}(K, \mathbf{X})\} = E\{u(K, \mathbf{X})\}(1 - \theta). \tag{5.26}$$

According to Equation (5.26), the seller has incentives to keep the probability of failure low, which is quite intuitive.

5.5.2 Asymmetric Information with Naive Buyers

We shall now consider the situation where buyers have no means of determining the probability of failure on their own. We refer to this situation as asymmetric with respect to the information, since the seller has more knowledge about service performance than buyers do. In this situation, the seller can announce the probability of failure, along with the percentage of reimbursement. We shall assume that buyers will take the value of the probability of failure announced by the seller as granted. We have called buyers in this situation as *naive*, risking to be using a too strong characterization. This should be rather interpreted as opposed to the rational buyers case, which we shall introduce following in Subsection 5.5.3, and the term naive should only be interpreted as buyers being seller's announcements takers

It can be readily derived from the seller's expected revenue in Equation (5.25), that this expected revenue increases with q when $\tilde{\theta}$ is greater than θ.

Indeed, let us define $B_\theta : D \to \mathbb{R}^+$ as:

$$B_\theta(q, \tilde{\theta}) = \frac{1 - \tilde{\theta}}{1 - q\tilde{\theta}}(1 - q\theta), \tag{5.27}$$

where $D = \{\tilde{\theta} \in [0, 1), q \geq 0 : q\tilde{\theta} < 1\}$.

According to Equation (5.25), the behaviour of the seller's revenue is driven by Equation (5.27), which is shown in Fig. 5.3. The seller could take advantage of this behaviour by announcing a probability of failure higher than the real one, and setting reimbursement at a value greater than 100%. In other words, negative marketing could be used, with a higher probability of failure being announced than the real one, with the goal of fooling naive buyers for the seller's benefit.
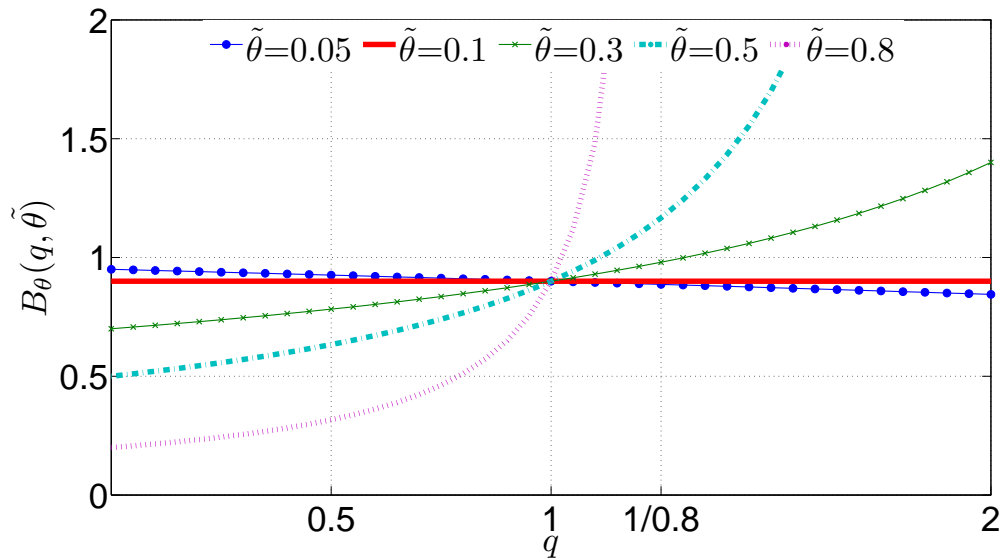

Figure 5.3: Variation of the seller's expected revenue as a function of reimbursement q for a real probability of failure $\theta = 0.1$ and for different values of probability of failure assumed by the buyers.

However, the negative marketing policy could be disadvantageous for the seller as well, for at least two reasons. First, if in the end buyers disregard sellers announcement with respect to the probability of service and assume some other value convenient for them, seller's revenue could diminish, to a value even lower than the revenue obtained when reimbursing 100%, as illustrated by Fig. 5.3. Second, the seller could be judged by buyers as untrustworthy, which could lead to losses not captured in our model. This situation demonstrates again that the seller's revenue depends on buyers' behaviour. In the following subsection we shall anticipate this behaviour in the case of asymmetric information by supposing that buyers act rationally.

5.5.3 Asymmetric Information with Rational Buyers

Let us now consider the case where buyers are uncertain about the probability of failure of the service they wish to buy and where they act rationally, seeking to maximize their payoffs. The seller's ultimate objective is still to set the value of q such that his or her revenue is maximized. We shall show that these two objectives, namely maximizing seller's and buyers' payoff, are conflicting, thus rather than finding an optimum we shall look for an equilibrium setting. Let us formalize this in what follows.

We recall that the dynamics of the proposed pricing mechanism implies that the seller announces a percentage of reimbursement q for a service which fails with probability θ. After the announcement, buyers bid to obtain this service, assuming that the probability of failure is $\tilde{\theta}$, *a priori* not necessarily equal to θ, and being aware of the value of q. The dynamics of service selling naturally impose an order: the seller announces a value of q and the buyers follow. Each side of the market, seller and buyers, take an action seeking to maximize their own utilities.

This kind of interaction is conveniently modelled by Stackelberg games [150], introduced by von Stackelberg in 1934. In a two sided Stackelberg game there is a *leader* that plays first, in our case the seller, and the *follower*, in our case the buyers, who plays next knowing the leader's move. We have already introduced the seller's utility, that is the seller's expected revenue in Equation (5.25). Let us now introduce the utility of the buyers, that is the bidder's expected payoff.

5.5.3.A Bidders' Expected Payoff

We now derive the buyer's expected payoff considering their payoff when bidding according to the best bidding strategy, given by Theorem (5.2). Following the same reasoning as for deriving the best bidding strategy in Section 5.4 a bidder's payoff is:

$$P = \mathbb{1}_{win}(x\mathbb{1}_{not\,failure} - \beta(x)(1 - q\mathbb{1}_{failure})), \tag{5.28}$$

where β is expressed in Equation (5.15) and the *failure* event refers to the real event of failure.

Computing expectations over the event of winning or not and the event of real failures, and replacing β by its definition we obtain the following expression for the real expected payoff of each bidder:

$$
\begin{aligned}
E\{P|X=x\} &= G(x) \\
&\cdot \left[x(1-\theta) - E[Y_{M-1}^{(K)}|Y_{M-1}^{(K)} \leq x]\frac{1-\tilde{\theta}}{1-q\tilde{\theta}}(1-q\theta) \right].
\end{aligned}
\tag{5.29}
$$

In Equation (5.29) we have considered a given realization of X. Let us now consider the expected payoff prior to having knowledge of this realization, by computing the so called *ex ante* expected payoff as:

$$E\{P\} = E\{E\{P|X = x\}\} = E\{G(X)X\} \cdot (1 - \theta)$$
$$- E\{\int_0^X vg(v)dv\}\frac{1 - \tilde{\theta}}{1 - q\tilde{\theta}}(1 - q\theta), \qquad (5.30)$$

where the expectation is over the valuations.

5.5.3.B The Pricing Game

It can be readily noticed from the expected seller's revenue and buyers' payoff in Equations (5.25) and (5.30) respectively, which we have reproduced in Table 5.1 for convenience, that seller's objective and buyers' objective comes to respectively maximizing and minimizing $B_\theta(q, \tilde{\theta})$. They thus have opposing objectives, so the optimal reimbursement value would be an equilibrium in the following pricing game.

Best Bidding Strategy	$\beta(x) = E[Y_{M-1}^{(K)}|Y_{M-1}^{(K)} \leq x]\frac{1-\tilde{\theta}}{1-q\tilde{\theta}}$
Seller's Expected Revenue	$E\{\bar{R}\} = E\{u(K, \mathbf{X})\}B_\theta(q, \tilde{\theta})$
Buyers' Expected Payoff	$E\{P\} = E\{G(X)X\} \cdot (1 - \theta)$ $-E\{\int_0^X vg(v)dv\}B_\theta(q, \tilde{\theta}).$
$B_\theta(q, \tilde{\theta}) = \frac{1-\theta}{1-q\tilde{\theta}}(1 - q\theta), (q, \tilde{\theta}) \in \{\tilde{\theta} \in [0, 1), q \geq 0 : q\tilde{\theta} < 1\}$	

Table 5.1: Summary of derived expressions.

Problem 5.1 *The Pricing Game is a zero-sum static Stackelberg game where:*

- *The leader is the seller and the buyers are followers*
- *The leader's set of available actions is $\{q : q \in \mathbb{R}^+\}$*
- *The follower's set of available actions is the set $\{\tilde{\theta} \in [0, 1) : 0 \leq \tilde{\theta} < \frac{1}{q}\}$*
- *The leader's utility is $B_\theta(q, \tilde{\theta})$ and the follower's utility is $-B_\theta(q, \tilde{\theta})$*

In Problem (5.1) we have assumed that all buyers would play the same $\tilde{\theta}$. This comes directly from the fact that buyers are symmetric.

We shall solve Problem (5.1) through the so-called backward induction method, that is to say that the maximization is solved first at the follower's level and this result is in turn used to solve the equilibrium at the leader's level. Let us formalize this solution in the following Theorem.

Theorem 5.3 *The game formulated in Problem (5.1) has as a solution the set $\{(q, \tilde{\theta}) \in \mathbb{R} \times \mathbb{R} : q = 1, 0 \leq \tilde{\theta} < 1\}$.*

Proof Note that Problem (5.1) can be reformulated as the following bi-level optimization problem.

$$\max_q B_\theta(q, \tilde{\theta}) \qquad (5.31)$$
$$\text{s.t. } q \geq 0, \tilde{\theta} \in \underset{\tilde{\theta}' \in [0,1):q\tilde{\theta}'<1}{\operatorname{argmin}} B_\theta(q, \tilde{\theta}')$$

In order to solve Problem (5.31), the backward induction method is applied. Hence, we first solve the second level optimization. As usual, in order to solve

$$\min_{\tilde{\theta} \in [0,1)} B_\theta(q, \tilde{\theta}) \qquad (5.32)$$
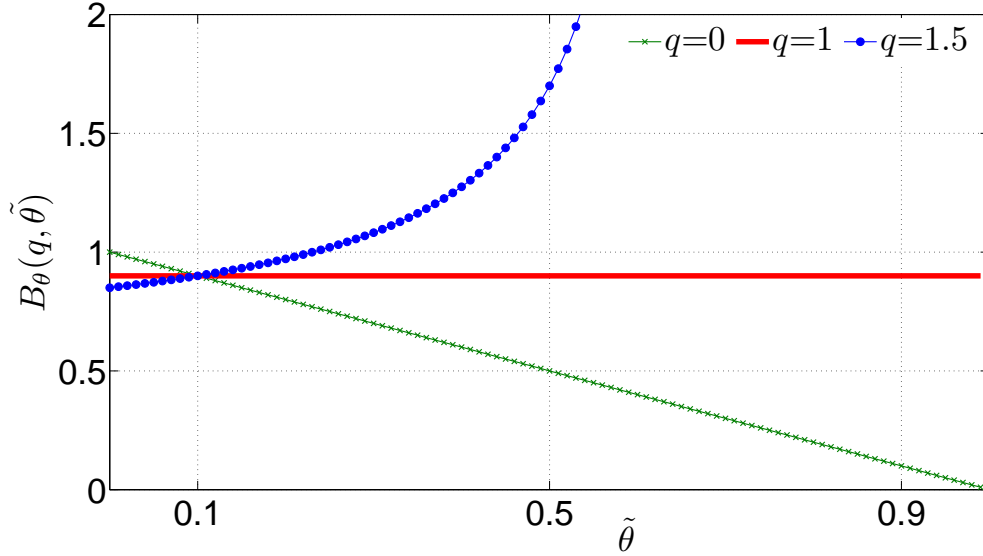$$\text{s.t. } 0 \leq \tilde{\theta} < 1/q, q \geq 0$$

Figure 5.4: Variation of B_θ as a function of $\tilde\theta$ for a real probability of failure $\theta = 0.1$ and for different values of reimbursement. Rational buyers select $\tilde\theta$ such that it minimizes B_θ.

the minimum is found at the values of $\tilde\theta$ where the first derivative of $B_\theta(q, \tilde\theta)$ with respect to $\tilde\theta$ is equal to zero or at the border of B_θ's domain. The first derivative of B_θ with respect to $\tilde\theta$ is

$$\frac{\partial B_\theta(q, \tilde\theta)}{\partial \tilde\theta} = \frac{q-1}{(1-q\tilde\theta)^2}(1-q\theta), \tag{5.33}$$

and there is no value of $\tilde\theta$ that renders it equal to zero. Hence, given that B_θ is a continuous function, the minimum, or infimum, must be reached at the border of its domain. Three cases must be distinguished, namely:

- $0 \le q < 1$: The infimum is attained at $\tilde\theta = 1$
- $q = 1$: Function B_θ is constant for all $\tilde\theta \in [0, 1)$
- $1 < q$: The infimum is attained at $\tilde\theta = 0$ and it is a minimum

This behaviour is shown in Fig. 5.4, where B_θ is plotted as a function of $\tilde\theta$ for different values of q.

Finally, the solution to the second level optimization is incorporated to Problem (5.31) obtaining the following equivalent problem

$$\max_q \{B_\theta(q, 1-\epsilon)\mathbb{1}_{q<1} + B_\theta(1, \tilde\theta)\mathbb{1}_{q=1} + B_\theta(q, 0)\mathbb{1}_{1<q}\}, \tag{5.34}$$

which, evaluating B_θ, can be expressed as

$$\max_q \left\{ \frac{\epsilon}{1-q(1-\epsilon)}(1-q\theta)\mathbb{1}_{0\le q<1} + (1-q\theta)\mathbb{1}_{1\le q} \right\} \tag{5.35}$$

and where ϵ is an arbitrarily small positive real number. It is easy to see that the solution to the seller's problem is attained at $q = 1$, which concludes the proof. Fig. 5.5 shows the behaviour of B_θ for the different cases considered in Problem (5.35), which illustrates this result. $\qquad\square$
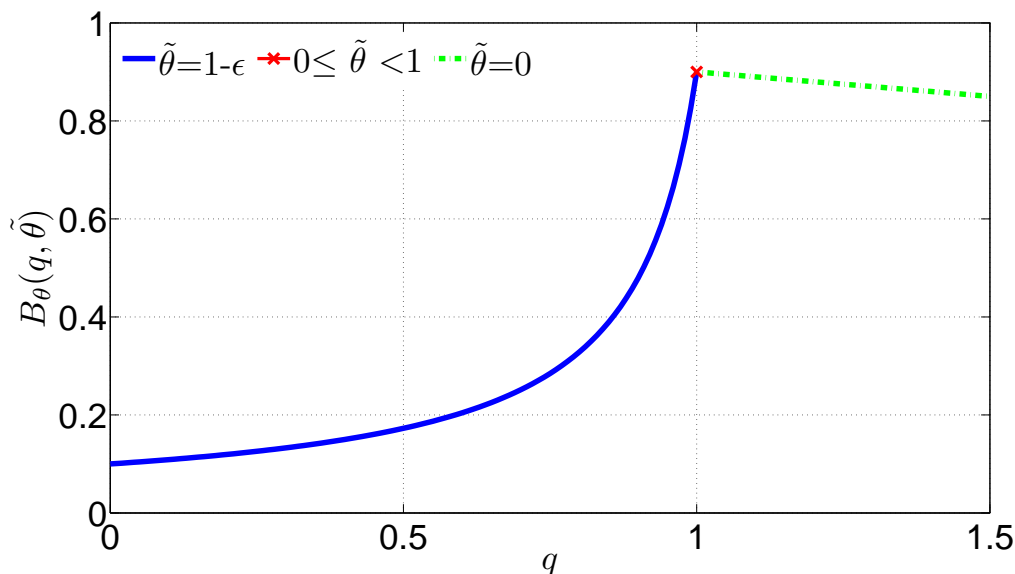
Figure 5.5: Variation of the seller's expected revenue as a function of reimbursement q for a real probability of failure $\theta = 0.1$ and a probability of failure estimated by rational buyer's $\tilde{\theta}$ equal to their best response for each q. The seller selects q such that it maximizes B_θ.

5.5.3.C Remarks and Interpretations

Interesting interpretations can be derived from the analytical results obtained above.

First, let us highlight the intuition behind the obtained results. If the seller announces a rather small percentage of reimbursement, buyers will, to some extent, tend to believe that the service fails a lot, and estimate the probability of failure $\tilde{\theta}$ close to 1. This is the so-called *market for lemons* phenomenon, introduced by Akerlof in 1970 [23]. The *market for lemons* states that when buyers are uncertain about the quality of the goods to buy, the market for high quality goods is reduced until it disappears. Indeed, this is what happens according to the theoretical analysis presented above: buyers assume that quality is very bad, which causes the value of the bids to approach zero.

Conversely, if the seller announces a high percentage of reimbursement, greater than 100%, buyers would intuitively assume that failures are not frequent, and thus estimate the probability of failure $\tilde{\theta}$ close to 0. We obtain here the so-called *moral hazard* behaviour. That is to say, that the buyers take a risk, by considering $\tilde{\theta}$ small (theoretically equal to zero), because if a failure were to occur it would be the seller who would bear the cost, through a high reimbursement. This behaviour is observed in many contexts where one of the players taking a decision is not the one bearing the responsibility for this decision. See, for instance, [104] for details on this phenomenon.

All in all, a reimbursement of 100% overcomes the problems that arise when there is asymmetric information. In addition, setting $q = 1$ provides the following three properties, worth highlighting.

Credibility. When the percentage of reimbursement is 100%, and this value is announced to the buyers along with a given probability of failure, then buyers can safely trust the announced probability of failure. Indeed, according to the expression of the seller's expected revenue shown in Equation (5.25) and illustrated through Fig. 5.3, when q is set to 1, the seller's expected revenue is constant. The seller thus has no incentives to announce a misleading value for the probability of failure in order to take advantage of naive buyers.

Insensitivity to the buyers' network performance assumption. At the setting $q = 1$, the seller's expected revenue is insensitive to the probability of failure assumed by the buyers. This can be directly seen setting q equal to 1 in Equation (5.25), which thus renders $E\{\bar{R}\} = E\{u(K, \mathbf{X})\}(1-\theta)$,

which is constant for any value of $\tilde{\theta}$. In particular, the seller's expected revenue is the same that he would obtain in the complete information case.

The analogous interpretation from the buyer's standpoint is translated into the following statement.

No value of information. At the setting $q = 1$, knowing the real probability of failure has no value to the buyer from the point of view of his or her payoff. Each buyer expected payoff is the same as when having complete information. This is readily derived from the buyers' expected payoff in Equation (5.30), which shows that when setting $q = 1$, the buyer's expected payoff is not affected by the assumed probability of failure $\tilde{\theta}$. Of course this knowledge could be valuable for the buyers for further reasons not captured in the model.

5.6   Summary

In this chapter we have proposed a pricing scheme where Assured-Quality Services over data networks are sold via first-price auctions and where in case of failures buyers are reimbursed a certain percentage of what they have paid to obtain the service. The percentage of reimbursement is announced by the seller before the service is sold. Under these conditions and with certain symmetry assumptions among buyers, we have analytically derived the best bidding strategy, which presents an intuitive behaviour. Indeed, for the same level of assumed probability of failure, the higher the percentage of reimbursement, the higher the bid. In addition, for any given percentage of reimbursement lower than 100%, the higher the assumed probability of failure, the lower the bid.

We have modelled the pricing problem, for the case where there is asymmetric information about the service and buyers are rational, as a Stackelberg game and shown that there is an equilibrium setting that maximizes seller's revenue, which is given by a reimbursement equal to 100%. In such equilibrium, the *market for lemons* effect and the *moral hazard* one are overcome and seller's expected and buyers' expected payoff are the same as when having complete information.

In particular our results show that, under the same level of failures, reimbursing 100% provides more revenue in expectation than no reimbursement at all. It must be noted that, since the model assumes the existence of a monitoring infrastructure, this output should be compared to the cost of the monitoring infrastructure in order to conclude if the overall balance is positive.

Chapter 6

The Proposed Pricing Scheme when Relaxing the Symmetry Assumption

6.1 Introduction

In Chapter 5 we have proposed a pricing scheme and derived our analysis of it on the assumption of symmetry among buyers. Under this assumption, the valuations that buyers attach to the service on sale are equally modelled for all of them. However, in real scenarios not all buyers would necessarily value the service in the same way. Moreover, the valuation they attach to a service could be determined by many firm-dependent factors as, for instance, the buyers' business model, cultural characteristics or geographic localisation, among others.

In addition, in Chapter 5 a single type of service is on sale and over a single route. Our ultimate objective is to be able to apply the pricing scheme introduced in Chapter 5 to the allocation mechanism introduced in Chapter 3. That is to say, to allocate bandwidth solving a NUM problem, such that the revenue of the alliance is maximized and the end-to-end QoS constraints are fulfilled, and where utility functions are obtained through first-price auctions with reimbursement. Failures in this case would account, for instance, for service disruption due to equipment failure, busy servers, etc. We shall refer to this situation as to the network case. The problem altogether is extremely complex, as we shall show, mainly because analytical results for the best bidding strategies or buyers' expected payoffs are not available. We shall thus keep some assumptions of Chapter 5, namely that services are sold over the same route and for the same amount of bandwidth, and provide guidelines on how the problem could be addressed when these assumptions are as well relaxed.

Thus, in this chapter we address the pricing scheme proposed in Chapter 5, where the assumption of symmetric buyers and single service type is relaxed. These relaxations render already the problem rather complex. The main difficult of relaxing the symmetric bidders assumption stems from the fact that the differential equation solved in Chapter 5 for finding the best bidding strategy becomes a system of differential equations, which in the general case does not have a closed-form solution. Moreover, in the network case, for instance, even the expected payoff of each buyer may not have a closed-form expression. However, the existence of the optimum bidding strategies was proven in similar, though simpler, cases by Lebrun in 1999 [90], Bajari in 2001 [36] and, Maskin and Riley in 2003 [108].

This chapter is the sequel to Chapter 5, and presents a simulative approach to draw conclusions about the optimum percentage of reimbursement when buyers are asymmetric and services with different probability of failure and percentage of reimbursement are offered. Its structure is as follows. In Section 6.2 we formalize the model. Section 6.3 studies the best bidding strategies in this asymmetric situation. That section first addresses a simpler case, that one of asymmetric bidders but only one single unit of a single service on sale. This case, though without reimbursement, has been studied in the literature, and we provide there a brief review of works proving the existence and uniqueness of the best bidding strategies. We then present same properties of the

best bidding strategies that hold as well when failures and reimbursement are in place. The section continues with the proposal of a numerical method to obtain an approximation of the best bidding strategies, where we allow for asymmetric bidders, multiple services with different probability of failure and different percentage of reimbursement. This method is used in Section 6.4 to compute the buyers expected payoff and the seller's expected revenue. Section 6.4 also states the Stackelberg reimbursement game, which allows to determine the optimal percentage of reimbursement for each service. Simulation results are presented in Section 6.5 for two particular case studies. A third case study is presented in Appendix C. Finally, a summary of the chapter is presented in Section 6.6.

6.2   The augmented model

Let us slightly update the model introduced in Chapter 5 so as to adapt it to the asymmetric case. Indeed, we shall allow for buyers with different characteristics and for services with different characteristics.

Let S be the set of the offered services. Each service $s \in S$ is characterized by a probability of failure θ_s and a percentage of reimbursement q_s. In order to be able to provide some analytical results we shall consider that all services are delivered over the same route of domains, and provide the same amount of bandwidth. For each service s there is a set $\mathcal{M}_s$ of buyers interested in that service. The total number of buyers is $M = |\bigcup_{s \in S} \mathcal{M}_s|$. Each buyer $i \in \mathcal{M}_s$ bids for the service s assuming a probability of failure $\tilde{\theta}_{i,s}$. As in Chapter 5, the real probability of failure θ_s is not necessary equal to the probability of failure assumed by a buyer $\tilde{\theta}_{i,s}$ for that service. We shall assume that $\tilde{\theta}_{i,s}, \theta_s \in [0,1)$. Herein, we shall note $q = \{q_s, s \in S\}$, $\tilde{\theta} = \{\tilde{\theta}_{i,s}, i \in \mathcal{M}_s, s \in S\}$ and subscript $-i$ shall be used to refer to all buyers in M but i. Each buyer i bids for getting one unit of service and attaches to that service s a positive valuation $X_{i,s}$. We assume that the $X_{i,s}$s are independently and identically distributed according to the distribution functions $F_{i,s}$, $i = 1 \ldots M$. A realization of the random variable $X_{i,s}$ is denoted as $x_{i,s}$. We shall generally simplify notation and refer to x as the realization of the valuation of any given bidder for a given service, when the context is clear.

Bidder i's bid is denoted as $b_{i,s}$. We shall assume that there is enough capacity to sell K any services and that the K highest bids win the auction. Thus, all buyers in the set M compete for getting one of the K units. Please note that the assumption that all services provide the same bandwidth allow us to select the winning buyers only based on the bid and not solving a NUM-like problem. Again as in Chapter 5, we assume a discriminatory payment rule, which means that the winning buyers pay their bids.

As in Chapter 5, bidders are assumed to be risk-neutral, as they seek to maximize their expected payoffs, which are linear functions of their valuation and bid. More precisely, given a valuation $x_{i,s}$ and a bid $b_{i,s}$ bidder i's payoff is $x_{i,s}(1 - \tilde{\theta}_{i,s}) - b_{i,s}(1 - q_s \tilde{\theta}_{i,s})$ if he or she wins the auction and 0 otherwise. Bidder i's expected payoff is thus,

$$\tilde{P}_i(\beta_{i,s}(X_{i,s})|X_{i,s} = x_{i,s}) = Pwin(b_{i,s})[x_{i,s}(1 - \tilde{\theta}_{i,s}) - b_{i,s}(1 - q_s \tilde{\theta}_{i,s})], \qquad (6.1)$$

where notation $Pwin(b_{i,s})$ refers to i's probability of winning when his or her bid is $b_{i,s}$ and where the other buyers bidding functions are implicitly β_{-i}, with $\beta_{-i}(X_{-i}) = \{\beta_{j,s}(X_{j,s})\}_{j \neq i}$.

Bidder i's bid is computed as a function of his or her valuation $x_{i,s}$ according to the best bidding strategy $\beta_{i,s}$. The best bidding strategy is for each bidder $i = 1 \ldots M$, the function $b_{i,s} = \beta_{i,s}(x), \beta_{i,s} : [0, x_i^{max}] \to \mathbb{R}_+$ that maximizes i's expected payoff, where the expected payoff is given by Equation (6.1). Please note that each $\beta_{i,s}$ is defined over a different support $[0, x_i^{max}]$ for each i, which corresponds to $F_{i,s}$'s support, where $x_i^{max} \in \mathbb{R}^+$.

The best biding strategies are thus, the Bayesian-Nash equilibrium of the game with incomplete information defined by the model introduced above. More precisely, they constitute a Nash equilibrium of the game with incomplete information where the buyers are the players, their expected payoffs given by Equation (6.1) are their utilities, the sets $[0, x_i^{max}]$ are the signals, types,

or valuations space, the functions $F_{i,s}$ define the probability distribution over the valuations, the set of actions is $\mathbb{R}_+$ and $\beta_{i,s}$ are the strategies. We discuss the existence of such equilibria in the following section. Throughout this chapter when we talk about equilibrium strategies they should be understood in that way.

6.3 Best Bidding Strategies

First-price auctions have been widely studied, and used in the networking field as reviewed in Chapter 3. However, when bidding behaviour is studied bidders are usually assumed to be symmetric, which is the same simplifying assumption on which results in Chapter 5 are derived. When this assumption is dropped, the symmetric equilibrium does not hold, which means that in order to find the best bidding strategies, instead of solving a first order differential equation as in Chapter 5, a system of ordinary differential equations must be solved. As pointed out in Section 6.1, this is a complex task, all the more so since the analytical expression of the expected payoff may not be available when we consider multiple services and asymmetric buyers. Next subsection provides more detail on this and formally formulates the problem.

6.3.1 Problem Formulation

We now formally formulate the problem of finding the best bidding strategies under the relaxed assumptions. We shall first introduce the single-object, single-service-type case, and then introduce the multiple-object, multiple-service-type case. Always with asymmetric bidders.

6.3.1.A The Single-Object and Single-Service-Type Case

In order to illustrate how hard the problem we are facing is, we shall first formulate it for the simplified scenario of a single object and a single type of service, and show that even in this case a closed-form expression for the best bidding strategies may not exist. In our model, single-object means $K = 1$, and since there is a single service type, we can save subscripts in s.

The probability of winning the auction for any given buyer i that submits a bid b can be expressed as:

$$Pwin(b) = \prod_{j \neq i} F_j(\beta_j^{-1}(b)) = \prod_{j \neq i} F_j(\phi_j(b)), \tag{6.2}$$

where $\phi_j(b)$ is defined as the inverse of β_j.

The expected payoff for buyer i is thus as follows:

$$\tilde{P}_i(\beta_i(X_i)|X_i = x_i) = \prod_{j \neq i} F_j(\phi_j(b))[x_i(1 - \tilde{\theta}_i) - b(1 - q\tilde{\theta}_i)]. \tag{6.3}$$

Let us, for convenience, define $\alpha_i = \frac{1 - \tilde{\theta}_i}{1 - q\tilde{\theta}_i}$. Imposing the first order condition to find a maximum of the expected payoff $\tilde{P}_i$ with respect to b for each bidder, and combining those equations we obtain:

$$\phi_i'(b) = \frac{F_i(\phi_i(b))}{(M-1)f_i(\phi_i(b))} \left\{ \frac{2 - M}{\phi_i(b)\alpha_i - b} + \sum_{k=1, k \neq i}^{M} \frac{1}{\phi_k(b)\alpha_k - b} \right\}, \quad i = 1 \ldots M \tag{6.4}$$

Equations (6.4) constitute a system of first order ordinary differential equations (ODE). Even though the solution to such system can not be expressed as a closed-form in most of the cases, its existence and uniqueness has been proved, under mild conditions of the distribution functions F_i, for the case with no reimbursement (i.e. $\alpha_k = 1$) and numerical methods for solving it have been proposed before. Subsections 6.3.2.A and 6.3.3.A respectively review those works.

6.3.1.B   The Multiple-Object, Multiple-Service-Type Case

Our scenario is more general than the single-object, single-service-type case, and consequently more complicated. If we consider K units over one path, the probability of winning can not be expressed as simply as in the previous case.

The problem is still to find the function $\beta_{i,s}$, $i = 1 \ldots M$ such that it maximizes each bidder's expected payoff, given by Equation (6.1), and where the probability of winning does not necessarily present an analytical expression.

We now present two asymmetric scenarios that remain simple enough in order to be solved analytically.

Proposition 6.1 *Best bidding strategies for two bidders with uniform valuations and two different services. Consider two bidders whose valuations are uniformly distributed on $[0, x^{max}]$. Bidder one bids for obtaining a service which he or she estimates fails with probability $\tilde{\theta}_{1,1}$ and for which a reimbursement equal to q_1 is announced. Bidder 2 bids for a service whose characteristics are $\tilde{\theta}_{2,2} \neq \tilde{\theta}_{1,1}$ and q_2. Only one of the services can be allocated. Then, the best bidding strategies are given by:*

$$\beta_i(x) = \frac{1 - \sqrt{1 - \alpha_i^2 k_i x}}{\alpha_i k_i x}, \ i = 1, 2, \tag{6.5}$$

where

$$\alpha_i = \frac{1 - \tilde{\theta}_{i,s}}{1 - q_s \tilde{\theta}_{i,s}}, \ (i, s) = (1, 1), (2, 2), \tag{6.6}$$

$$k_i = \frac{\alpha_j^2 - \alpha_i^2}{(x^{max} \alpha_i \alpha_j)^2}, \ i, j = 1, 2, \ i \neq j. \tag{6.7}$$

Proof Please see Appendix C. $\square$

Analogously, the scenario presented in the following preposition can be derived:

Proposition 6.2 *Best bidding strategies for two bidders with uniform valuations on two different intervals and one service. Consider two bidders whose valuations are uniformly distributed on $[0, x_i^{max}]$, $i = 1, 2$. Consider there is one service on sale that fails with probability $\tilde{\theta}$ and for which reimbursement is q. The best bidding functions are given by:*

$$\beta_i(x) = \frac{1 - \sqrt{1 - \alpha^2 k_i x}}{\alpha k_i x}, \ i = 1, 2, \tag{6.8}$$

where

$$\alpha = \frac{1 - \tilde{\theta}}{1 - q\tilde{\theta}}, \tag{6.9}$$

$$k_i = \frac{1}{(x_1^{max})^2} - \frac{1}{(x_j^{max})^2}, \ i, j = 1, 2, \ i \neq j. \tag{6.10}$$

Proof The proof is completely analogous to the one of Proposition (6.1). $\square$

As we shall see in the following subsection, the existence of a solution to the system of differential equations for particular cases has been proved, however in the case with multiple-object, multiple-service-type, with or without reimbursement, the proof remains open. We shall thus adopt, in Subsection 6.3.3, a simulative approach in order to explore this case. But first let us review the works related to the existence on the equilibrium and show some properties the best bidding strategies fulfil in the reimbursement case.

6.3.2   Characterization of the Equilibrium Strategies

For the single-object scenario, and when no reimbursement policy is in place, the literature studying optimal bidding strategies is rich. In particular, much work has focused on the study of the existence and uniqueness of the best bidding strategies, which we review in the following subsection.

6.3.2.A   Related Work: Existence and Uniqueness of the Equilibrium Strategies for the Single-Object no Reimbursement Case

The existence and uniqueness of an equilibrium bidding strategy on sealed single-unit first-price auctions with M bidders was proved independently by Lebrun [89,90], Bajari [36] and, Maskin and Riley [108]. The model in all these studies is very similar, they study the sealed single-unit first-price auctions with independent private values. Bidders have private information, i.e. their valuation or types, which are known only to each of them and are drawn from different distribution functions, which accounts for the asymmetry. All of them ask for the distribution functions to be defined over the same domain.

In [89,90], it is also shown that the bidding functions are strictly increasing functions of the valuation and that the maximum bid for all the bidders is given by a value that is the same for all of them. More specifically, they prove that the equilibrium is the solution of a system of first order ordinary differential equations with border conditions, and these border conditions are values of the bids at the upper extreme of the valuations' domain.

Bajari proves, in [36], the existence and uniqueness of the equilibrium by assuming that the inverse of the bidding functions are continuously differentiable. Maskin and Riley have studied the existence in [106] and the uniqueness in [108], proving that the inverse functions are indeed differentiable. In [108], the distribution functions are not initially required to be defined over the same domain. However, this assumption is introduced when deriving the uniqueness of the equilibrium, though the authors claim that their results can be readily extended if that assumption is relaxed. In the proofs of Lebrun and Bajari the bidders are modelled as risk-neutral, and thus they try to maximize a linear discontinuous payoff function (the same model we have assumed). In the work of Maskin and Reily they adopt a broader model and buyers are allowed to have payoffs other than linear. However, a number of hypothesis are required over these payoffs. One of such hypothesis asks that the payoffs are the same for the same valuation. In our case this is not necessarily true, since bidders' payoffs depend as well on the percentage of reimbursement q_s of the service and the probability of failure $\tilde{\theta}_{i,s}$ assumed by the bidder. The consequence of this is that, unless for the two bidders case, the maximum bid is not necessarily the same for all bidders.

We shall not attempt to extend the proofs of existence and uniqueness of the equilibrium to our reimbursement case, which are quite involved. However, we shall state some theoretical properties that the equilibrium, if it exists, fulfils in our scenario and we later on explore its existence through simulations.

6.3.2.B   Properties of the Best Bidding Strategies for the Reimbursement Case

We have mentioned it above, in the asymmetric case the best bidding strategies might not have a closed-form expression. There are, though, some properties about their behaviour that can be proved provided some assumptions hold.

When there is no reimbursement, the bidding strategies of asymmetric bidders under mild conditions of their valuations' distributions functions, have been proved to be strictly increasing [90,108]. Also, if the upper end of their domain is the same, then the bidding strategies at this end take the same value for all the buyers. The previous properties under some conditions, still hold in the scenario with reimbursement. Besides, we have proved, in Chapter 5, that for symmetric buyers symmetric equilibrium holds. This is also true for those symmetric bidders within an asymmetric group. The following lemmas formalize these statements.

Lemma 6.1 Non-decreasing best bidding strategies in the single-object single-service-type case. At

equilibrium, the strategies are non decreasing functions of the valuations.

Proof The proof can be readily adapted from that one provided in [107] for the single-object and no reimbursement case. We shall save subscript s since there is only one service on sale. Let b_i be the equilibrium bid of bidder i for a given realization of his or her valuation x_i. That is, $b_i = \beta_i(x_i)$. For convenience let us call the payoff of bidder i if he or she wins the auction as u_i, that is:

$$u_i(b_i, x_i) = x_i(1 - \tilde{\theta}) - b_i(1 - q\tilde{\theta}). \tag{6.11}$$

Let us call $h_i(b_{-i})$ to the joint density distribution of the bids of every buyer but i.

Since b_i is the equilibrium bid, i's expected payoff when bidding b_i must be greater than or equal to i's expected payoff when bidding any other bid, say $\hat{b}_i$. Thus,

$$\int_{b_{-i} \leqq b_i} u_i(b_i, x_i) h_i(b_{-i}) db_{-i} \geq \int_{b_{-i} \leqq \hat{b}_i} u_i(\hat{b}_i, x_i) h_i(b_{-i}) db_{-i}. \tag{6.12}$$

Notation $\int_{b_{-i} \leqq b_i} \cdot db_{-i}$ means the multiple integral of $\cdot$ with respect to b_i, $\forall i \neq j$, where the domain of integration should be interpreted as $0 \leq b_j \leq b_i$, $\forall j \neq i$.

Let now $\hat{b}_i < b_i$. To conclude the proof it suffices to prove that

$$\int_{b_{-i} \leqq b_i} \frac{\partial u_i(b_i, x_i)}{\partial x_i} h_i(b_{-i}) db_{-i} > \int_{b_{-i} \leqq \hat{b}_i} \frac{\partial u_i(\hat{b}_i, x_i)}{\partial x_i} h_i(b_{-i}) db_{-i}, \tag{6.13}$$

since if Inequality (6.13) holds, we can integrate both sides of it with respect to x_i from x_i to x_i', with $x_i' > x_i$, and sum up the result of the left-hand term with the left-hand term of Inequality (6.12) and do the same with the right-hand sides, to find that Inequality (6.12) holds strictly for $x_i' > x_i$, that is:

$$\int_{b_{-i} \leqq b_i} u_i(b_i, x_i') h_i(b_{-i}) db_{-i} > \int_{b_{-i} \leqq \hat{b}_i} u_i(\hat{b}_i, x_i') h_i(b_{-i}) db_{-i}. \tag{6.14}$$

Let us now prove that Equation (6.13) holds. The left-hand term in Inequality (6.13) is equal to:

$$\int_{\hat{b}_i \leqq b_{-i} \leqq b_i} \frac{\partial u_i(b_i, x_i)}{\partial x_i} h_i(b_{-i}) db_{-i} + \int_{b_{-i} \leqq \hat{b}_i} \frac{\partial u_i(b_i, x_i)}{\partial x_i} h_i(b_{-i}) db_{-i}. \tag{6.15}$$

Since b_i is a best response, $\hat{b}_i$ must win with strictly lower probability than b_i. In addition, by definition of u_i, $\frac{\partial u_i(b_i, x_i)}{\partial x_i} > 0$ (we recall that $\tilde{\theta}_i \in [0, 1)$). These two facts allow us to conclude that the first term in Expression (6.15) is strictly positive.

On the other hand, since $\frac{\partial^2 u_i(b, x_i)}{\partial x_i b_i}$ is positive and we have assumed $\hat{b}_i < b_i$, the second term in Expression (6.15) is at least as large as the right-hand of Inequality (6.13).

All in all, we have that Inequality (6.13) holds and thus Inequality (6.14) holds. Let b_i' be the best bidding strategy for bidder i when his or her valuation is x_i'. Thus i's expected payoff when bidding b_i' must verify

$$\int_{b_{-i} \leqq b_i'} u_i(b', x_i') h_i(b_{-i}) db_{-i} \geq \int_{b_{-i} \leqq b_i} u_i(b_i, x_i') h_i(b_{-i}) bx_{-i} \tag{6.16}$$

$$> \int_{b_{-i} \leqq \hat{b}_i} u_i(\hat{b}_i, x_i') h_i(b_{-i}) db_{-i}, \forall \hat{b}_i \leq b_i,$$

where the first inequality comes from the fact that b_i' is the equilibrium bid for valuation x_i' and the second from Inequality (6.14).

We want to prove that $b_i' \geq b_i$. Suppose by contradiction that $b_i' < b_i$. Then we can evaluate Inequality (6.14) at b_i' obtaining:

$$\int_{b_{-i} \leqq b_i'} u_i(b', x_i') h_i(b_{-i}) dx_{-i} > \int_{b_{-i} \leqq b'} u_i(b_i', x_i') h_i(b_{-i}) db_{-i}, \tag{6.17}$$

a contradiction, which concludes the proof. $\qquad\square$

The previous lemma is also valid for the multiple-object, multiple-service-type case. Indeed, redefine $h_i(b_{-i})$ as the joint density function of random variable $B_{-i,M-1}^{(K)}$, which indicates the K-th highest bid over the $M-1$ bids from all bidders but bidder i. Please not that we have used the same notation as in Chapter 5 and $B_{-i,M-1}^{(K)}$ is also the $(M-K)$-th order statistics over $(M-1)$ bids drawn from the bids distributions of buyers other than i. Redefine as well the domain of integration as $b_{-i,M-1}^{(K)} \leqq b_i$. The proof then follows as in the single-object, single-service-type case.

The following lemma states a border condition for the system of differential equations given by Equations (6.4), when bidders have the same upper valuation.

Lemma 6.2 Symmetric highest bid when bidders have the same upper valuation. If all bidders have the same upper maximum valuation, that is $x_{i,s}^{max} = x^{max}$ and there is a single-object and single-service type, that is, $\tilde{\theta}_{i,s} = \tilde{\theta}$, $q_s = q$, $i = 1 \ldots M$, then $\beta_i(x_i^{max}) = b^{max}$, $i = 1 \ldots M$.

Proof The proof can be readily extended from the case presented in [108], where no failures neither reimbursement are considered. Indeed, if all bidders have the same upper valuation, and if there is only one services both assumptions required in [108] hold, and the proof follows directly. We shall reproduce it here.

Let us define the following quantities

$$p_j(b) = log F_j(\phi_j(b)) \tag{6.18}$$

$$c(b,x) = -log(x(1 - \tilde{\theta}) - b(1 - q\tilde{\theta})) \tag{6.19}$$

Finally we define the logarithmic payoff as:

$$e_j(b,x) = \sum_{i \neq j} p_j(b) - c(b,x). \tag{6.20}$$

It can be readily seen that the best bidding strategies, those maximizing the expected payoff given by Equation (6.1), are as well maximizers of the logarithmic payoff given by Equation (6.20).

Let $b_j^{max} = \beta_j(x^{max})$. Without loss of generality we order the buyers' bids as

$$b_1^{max} \geq \ldots \geq b_M^{max}. \tag{6.21}$$

Then we can evaluate buyer 1's logarithmic expected payoff if he or she were to bid b_M^{max} obtaining:

$$e_1(b_M^{max}, x^{max}) = \sum_{j=2}^{M} p_j(b_M^{max}) - c(b_M^{max}, x^{max}) \leq \sum_{j=2}^{M} p_j(b_1^{max}) - c(b_1^{max}, x^{max}) \tag{6.22}$$
$$= -c(b_1^{max}) = e_1(b_1^{max}, x^{max}),$$

where the first equality comes form the definition of $e_i(\cdot)$ in Equation (6.20), the inequality comes from the fact that if b_1^{max} is the optimum strategy, then the payoff when bidding anything different

than b_1^{max} must be smaller. We have also used in the last equality the fact that $p_j(b_1^{max}) = log F_j(x^{max}) = 0$.

Let us now assume that b_1^{max} is strictly greater than b_M^{max}. In this case, $p_1(b_M^{max}) < 0$, since $\phi_1(b_M^{max})$ is strictly lower than x^{max}, and $p_M(b_M^{max}) = 0$. We then apply the definition of the logarithmic expected payoff to M and combine it with these facts and with the result of Equation (6.22) to obtain:

$$
\begin{aligned}
e_M(b_M^{max}, x^{max}) &= \sum_{j=1}^{M-1} p_j(b_M^{max}) - c(b_M^{max}, x^{max}) \\
&= p_1(b_M^{max}) + \sum_{j=2}^{M} p_j(b_M^{max}) - c(b_M^{max}, x^{max}) \qquad (6.23) \\
&< -c(b_1^{max}, \bar{x}) = e_M(b_1^{max}, x^{max}),
\end{aligned}
$$

where in the last equality we have used, again, that $p_j(b_1^{max}) = log F_j(x^{max}) = 0$. Then, b_M^{max} is not the best strategy for M when his or her valuation is x^{max}, a contradiction. $\square$

In particular, the previous lemma is generalized for every two-bidder single-object auction.

Lemma 6.3 Symmetric highest bid for the two-bidders single-object multiple-service-type auction. In a first-price two-bidders single-object first price auction $x_1^{max} = x_2^{max} = x^{max}$.

Proof Suppose without loss of generality that in equilibrium bidder 1's highest bid b_1^{max} is greater than bidder 2's highest bid b_2^{max}, then bidder 1 would for sure win the object if bidding any value slightly greater than b_2^{max}, and would have a greater payoff by decreasing his bid up to slightly above b_2^{max}. Thus, bidding b_1^{max} is not an optimum, a contradiction. $\square$

Lemma 6.4 Symmetric Equilibrium for symmetric buyers. Buyers whose valuations are drawn from the same distribution and who bid for obtaining the same service, of which they estimate the same probability of failure have the same best bidding strategy.

Proof The proof is immediate. Indeed, those buyers whose valuations are drawn from the same distribution and bid for obtaining the same service, of which they estimate the same probability of failure, are undistinguishable and would solve the same expected payoff, thus the same equilibrium strategy. $\square$

6.3.3   Computation of the Equilibrium

We now focus on the computation of the best bidding strategies when no closed-form is available. We shall first review existing methods and then propose an algorithm tailored for our particular scenario.

6.3.3.A   Related Work: Numerical Methods for Computing the Best Bidding Strategies without Reimbursement

Several works have focused on proposing numerical methods to find the best bidding strategies, though all of them correspond to the case without failures and without reimbursement. We following review such proposals with the idea in mind to evaluate if they are suitable to be adapted to the reimbursement multiple-object multiple service type case.

Methods for Computing Nash Equilibria in Bayesian Games by solving the underlying ODE system   As aforementioned, single-unit first-price auctions with asymmetric bidders have,

under some conditions on their distribution functions, a unique Bayesian Nash Equilibrium. This equilibrium is the solution to a system of first order differential equations (ODE) with boundary conditions, which in the general case does not have a closed form solution. Different methods were proposed to obtain numerical approximations of such solution.

A widely used numerical method is the so-called backward shooting method. The system of differential equations, which we have introduced in previous subsection in Equation (6.4), along with the border conditions poses two main problems. One is that it does not behave well near the origin, since the condition $\phi_i(0) = 0$ renders an indetermination. Indeed we get $\frac{0}{0}$ in the right-hand of Equation (6.4). Second, a border condition is given as well at the upper extreme point of the domain, but the point at which it is attained, i.e. the value of the bid at that upper extreme point, is unknown. The backward-shooting method assumes a point at which that border condition is attained and checks back to see if the condition at the origin is fulfilled. This method was proposed by Marshall in [103] and was used, for instance, in [65, 92, 106]. However, recently in [62] the backward-shooting method was shown to be unstable for the case of large number of buyers (for instance more than 20 buyers). In [62] an alternative, novel method is proposed, which is based on a boundary-value method and is said to overcome the backward-shooting methods instability shortcomings. In this method the system of differential equations is transformed so as to obtain one that can be solved by standard methods as the fix-point method or the Newton method, similarly to the first algorithm in [36], on which we shall comment below.

Still for the case of sealed first-price single-unit auctions, a sound summary of three numerical algorithms to approximate the inverse of the bidding strategies is presented by Bajari in [36]. The first algorithm proposed in that work is based on a proposal by Maskin and generalized by Riley and Li, in an unpublished work at that moment, later appeared in [92]. The algorithm consists on guessing an initial condition to the system of differential equations, and then using standard techniques to solve the system. The problems pointed out in the previous paragraph about the bad behaviour of the system at the origin are not encountered by them since they add the hypothesis that the valuations distributions functions are bounded away from 0 in the lower extreme of their common domain. The method was shown by the authors to present slow convergence. The second algorithm, which had been used previously in [35], starts by assuming that bidders bid their valuation and iteratively adapt the inverse of the bidding strategies according to best responses, which are computed sequentially by all bidders. The method was shown to perform well with the distribution functions considered by the authors. The idea behind this algorithm is very similar to the algorithm we shall finally propose for our problem, which we shall introduce in the following subsection. However, since this second algorithm is considered in a much simpler scenario than ours (single-object, single-service-type and no reimbursement case) at each iteration analytical expressions are available to compute the updates and this is not the case in our general scenario. The third algorithm approximates the inverse of the bidding functions by parametric functions and searches the parameters such that the system of differential equations, along with its border conditions, is approximately satisfied. The parametric function can be, for instance, a high order polynomial. The algorithm is said to obtain fast results with good initial values.

The methods described above assume that the probability of winning has a known expression, provided that buyers' valuations distribution functions are known. However, in more general configurations, as the multple-object multiple-service-type case, this is not the necessarily true. That is to say, even if valuations distributions functions are known, the probability of winning remains unknown. Nevertheless, the probability of winning, can be approximated, for instance by Monte-Carlo techniques as shall be presented later on in this chapter, and as proposed by the following method.

Bajari's third algorithm is closely related to the concept of Constrained Strategic Equilibrium (CSE) introduced by Armantier et al. [32]. In the CSE method, the bidding strategies are restricted to a set with certain properties and the equilibrium is searched restricted to that set. The authors show that a sequence of CSE solutions approximates the equilibrium in a Bayesian game (it is not restricted to first-price auctions) and they provide an algorithm to compute the solution. In order to implement this mechanism, the idea is to parametrize the proposed bidding strategies and solve a system with respect to the parameters instead of with respect to the bidding strategy.

The algorithm is not restricted to single-unit first-price auctions but can be applied as well, for instance, to multi-unit first-price auctions. This makes of this algorithm an attractive one to be adopted for our scenario, though its implementation is rather involved.

Learning Techniques   A different approach is to make the players or agents (the buyers in our framework) of the Bayesian game defined by the auction mechanism to learn how to play their best bid. The literature regarding learning in multi-agent systems is wide. However, very few works were found about learning in Bayesian games. A good review is available in [48].

In [71] a learning approach is proposed to find the equilibrium bidding strategy in first-price discrete auctions (auctions with discrete bid space). They model the problem as a repeated game in strategic form. They repeat the first-price auction assuming, at the beginning of each repetition, a sampled set of valuations. They show that after sufficient long time the players' bids are in the equilibrium of the one-shot auction. Although this constitutes a very interesting proposal the model is quite different from ours', for instance, they consider a discrete bid space, though we shall not adopt their approach.

A Q-learning technique is used in [148] to learn in a double-auction scenario. However, the choice of the technique and the simulations performed are not clearly presented, and no information about convergence is provided. In addition, a stochastic modelling is used in [120] for an adaptive agent to increase his or her payoff in a continuous double auction scenario.

Other techniques   Finally, an algorithm to compute the best bidding strategies where no assumptions are made about initial knowledge of the probability of winning was proposed by Holenstein, reported in [69]. However, no publications were found. The algorithm estimates the best bidding strategy for each player, by iteratively adapting it, such that the expected payoff of each player is maximized, in a way similar to Bajari's second algorithm introduced above. This algorithm is flexible, easy to implement, and assumes no previous knowledge about probabilities of winning. These are the main reasons why we shall adopt this approach and tailor it to our particular scenario. In particular, we need to adapt it to the reimbursement case and the multi-unit auction. Additional improvements, such as a different approach to update bidding strategies, have been done and are presented in detail in the following subsection.

6.3.3.B   The Proposed Algorithm for Computing the Best Bidding Strategies

We propose an algorithm that makes it possible to find an approximation of the best bidding strategies in the multiple-object, multiple-service-type with reimbursement case. Roughly speaking, the algorithm is based on a centralized, iterative procedure, which imitates the real bidding process. It relies on the Monte Carlo method to compute approximations of the probability of winning the auction, and on the Simulated Annealing mechanism to optimize the involved payoffs.

6.3.3.B.1   Description   For the sake of clarity let us first assume that the probability of winning the auction is a known function of the submitted bid, we shall relax this assumption later on.

The main logic of the algorithm is described by the diagram in Fig. 6.1, where $\hat{\beta}_i$ and $\hat{\beta}_{-i}$ represent the approximate bidding functions (and not particular bid values). For each buyer i, in turn, a valuation x is sampled and the best response to x, considering that other buyers' bidding functions are $\hat{\beta}_{-i}$, is computed. In more detail, this best response is computed as the value b_i that maximizes buyer i's expected payoff with respect to his or her bid for valuation x. The bidding functions of the other buyers are assumed fixed at their current approximation $\hat{\beta}_{-i}$, which allows to compute the probability of wining and thus the expected payoff. For the obtained value of b, the bidding function of the current buyer, that is $\hat{\beta}_i$, is updated in a way that we shall explain later on. This is repeated in turn for all buyers, and until convergence of the strategies. In order to be as general as possible, each bidding function is approximated by a piecewise linear function.
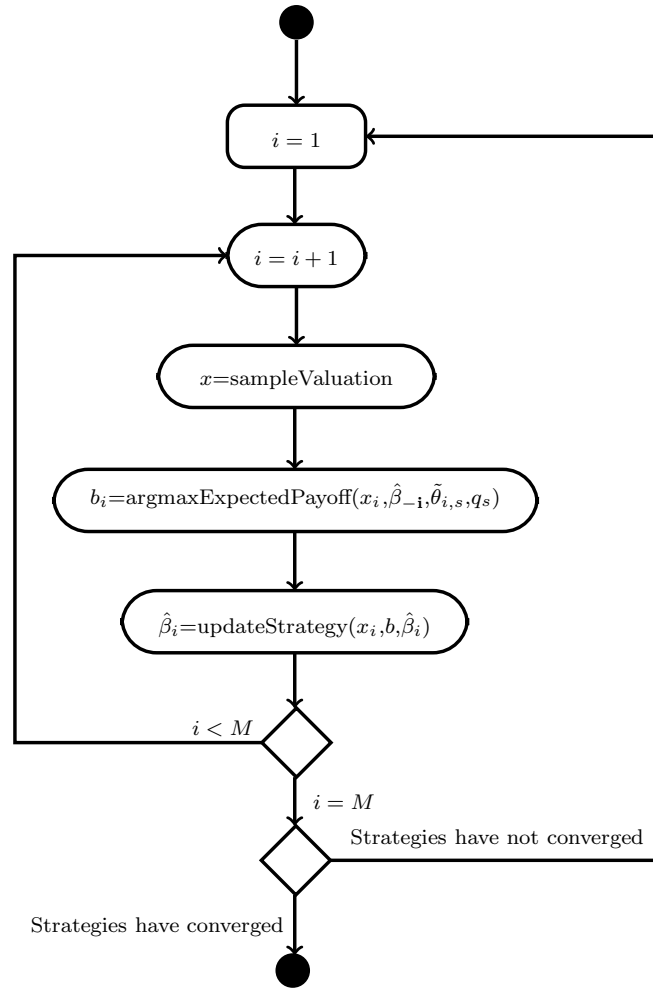
Figure 6.1: Schematic representation of the algorithm

We now further explain each of the steps of the algorithm.

Maximization of the Bidder's Expected Payoff. In order to maximize each buyer's expected payoff an iteration as well takes place, which is based on the Simulated Annealing method. The Simulated Annealing method is a probabilistic method designed to find the global optimum of a function that may possess several optima. It works by emulating the physical process of a solid that is slowly cooled till it is frozen, which occurs at a level of minimum energy. The method was originally proposed by [82] and [147] for finding global minimum, but it can be readily adapted to find a maximum.

A priori, any well-known optimization method could be used. However, since the function to optimize may have local maxima, we propose to use a Simulated Annealing approach, whose convergence may be slower in some cases compared to a classical gradient method, but it does not get stuck at local maxima.

The pseudo-code of the Simulated Annealing procedure adapted to our maximization problem can be seen in Pseudo-code 6.1. At each step of the iteration the optimum candidate is perturbed and its performance evaluated. As the temperature T gets colder, the acceptance rate of new worse solutions becomes smaller.

The routine `expected_payoff` computes a mean on the payoff for valuation x_i, bid b_i and assuming that other bidders bid according to the current estimation of their bidding strategies $\hat{\beta}_{-i}$ and considering the randomness on other bidders' valuations.

Pseudo-code 6.1: Maximization of bidder i's expected payoff via Simulated Annealing.

```
input: x_i,β̂_-i
b_i = initialize_bid()
T=initialize_temperature()
ep=expected_payoff(x_i,b_i,β̂_-i)
while not converged
    b̃_i = b_i+N(0,σ) %randomly perturb the bid
    new_ep=expected_payoff(x_i,b̃_i,β̂_-i)
    sample Z from uniform[0,1] distribution
    ratio=min(1,exp((new_ep-ep)/T))
    if ratio > Z
        b_i=b̃_i
        ep=new_ep
    T=cool(T)  % T decreases exponentially
    evaluate convergence
output: b_i
```

Assessing the Probability of Winning the Auction. We shall now relax the assumption that the probability of winning the auction is known. In that case, one can estimate the probability of winning via the Monte Carlo method. To consider these cases, we slightly modify the algorithm, and in the main loop, shown in Fig. 6.1, we first estimate buyer i's expected payoff as a function of his or her bid b_i, and for a given valuation x_i and strategies of the other buyers $\hat{\beta}_{-i}$. This function is then used at every iteration of the Simulated Annealing maximization, shown in Pseudo-code 6.1. The expected payoff depends on the probability of winning which is computed using the Monte Carlo method as we explain in the following paragraph.

To allow the algorithm to be suitable to the multiple-object and multiple-service-type case, we propose to use a classical Monte Carlo simulation method to evaluate the probability of winning as a function of the bids. That is, for a value b_i, the bid of bidder i, we run a number of MC independent auctions and determine at each auction if bidder i has won or not. Finally, we take the frequency of the winning events over all the events.

Since our working assumptions are that all services have the same bandwidth and are offered over the same route, there is no need on solving the allocation problem at each experiment. Indeed, the probability of winning can be found by computing the probability density function of the K-th highest bid over $M - 1$ i.i.d. draws from the corresponding distributions.

$$Pwin(b_i) \simeq \frac{1}{MC} \sum_{k=1}^{MC} \mathbb{1}_{b_i \geq b_{k,M-1}^{(K)}}, \tag{6.24}$$

where $b_{k,M-1}^{(K)}$ is the K-th highest bid off all bidders but i's bid, computed each bid as $b_j = \hat{\beta}_j(x_j^{\{k\}})$, $\forall j \neq i$ and $x_j^{\{k\}}$ a sample drawn from j's distribution $F_{j,s}$ in the k-th repetition of the Monte Carlo method. In order to obtain a better approximation of the probability of winning a density kernel approximation can be used, for instance with a Gaussian kernel. In that case, the density probability function of random variable $B_{k,M-1}^{(K)}$, which we call $k(b)$, can be approximated

as in Equation (6.25).

$$k(b) \simeq \frac{1}{MC} \sum_{k=1}^{MC} \frac{1}{\sqrt{2\pi h}} \exp \left(\frac{- \left(b - b_{k,M-1}^{(K)} \right)^2}{2h^2} \right).$$ (6.25)

In both cases, as MC becomes larger the accuracy of the estimator becomes better.

Updating the Bidding Strategy. Finally, let us describe in more detail the way the bidding strategies are approximated and updated at the end of each iteration of the main loop. We assume that the algorithm starts with no knowledge about the bidding strategies except their support, which is the valuations' distribution function support. In order to model this, we consider piecewise linear functions. Each bidding function's support is partitioned into a given number of subintervals, over which the bidding function is approximated by a linear function. We inicialize the bidding strategies assuming the bid is equal to the valuation, that is $\hat{\beta}_i(x) = x$. At each iteration t, for each buyer i a new pair (x_i^t, b_i^t) is obtained, with the sample valuation x_i^t belonging to a certain subinterval ρ. The linear function at subinterval ρ is updated to the linear function that best fits the latest samples obtained for that subinterval.

More formally, let $\Delta_i = \{\Delta_{i,1} \ldots \Delta_{i,\rho_i^{max}}\}$ denote the partition of buyer i's bidding function support, where ρ_i^{max} is the number of subintervals into which i's support is partitioned. Then, the bidding function for buyer i at iteration t can be expressed as in Equation (6.26), where $\alpha_{i,\rho}^t$ and $\gamma_{i,\rho}^t$ are the coefficients of the linear function that best fits the latest T samples obtained for buyer i at subinterval $\Delta_{i,\rho}$, $\rho \in \{1 \ldots \rho_i^{max}\}$ during previous iterations. The least squares method is used for computing those coefficients.

$$\hat{\beta}_{i,s}^t(x) = \sum_{\rho=1}^{\rho_i^{max}} \left(\alpha_{i,\rho}^t x + \gamma_{i,\rho}^t \right) \cdot \mathbb{1}_{\Delta_{i,\rho}(x)}$$ (6.26)

At the end of each iteration of the main loop, $\alpha_{i,\rho}^t$ and $\gamma_{i,\rho}^t$ are updated.

Evaluating the Strategies Convergence. In order to evaluate the strategies' convergence we compare for each subinterval the linear function obtained for that subinterval when considering the previous T samples and the one obtained when considering the actual sample together with the previous $T-1$ samples. We assess the difference between both approximations at the extreme points of the subinterval and if the maximum of that difference is smaller than a bound, we assume that the approximation for the corresponding subinterval has converged.

6.3.3.B.2 Validation Convergence and validation of the designed and implemented algorithm has been successfully carried out through extensive simulative studies. Those cases where analytical results are available provide us with a benchmark against which the accuracy of the developed algorithm can be measured. These cases are not only limited to symmetric buyers but also include some asymmetric cases with 2 different kinds of buyers as the case stated in Proposition (6.1). In addition, when no analytical solution is available, theoretical properties of the approximate solution can still be verified. Namely, we can verify those properties stated in Section 6.3.2.B. We now present different scenarios in order to validate the algorithm.

We first focus on symmetric scenarios. As explained in Chapter 5, in a symmetric scenario buyers are assumed to be symmetric, which leads to a symmetric equilibrium, in which all buyers' bidding functions are equal and can be explicitly computed.

Scenario 1 considers 4 symmetric bidders whose valuations are drawn from the uniform distribution on the interval $[0, 1]$ and one service on sale. In Fig. 6.2 algorithm's output is plotted along with the exact solution for this scenario and different amounts of reimbursement and probabilities of failure. In all the cases we observe an accurate result.
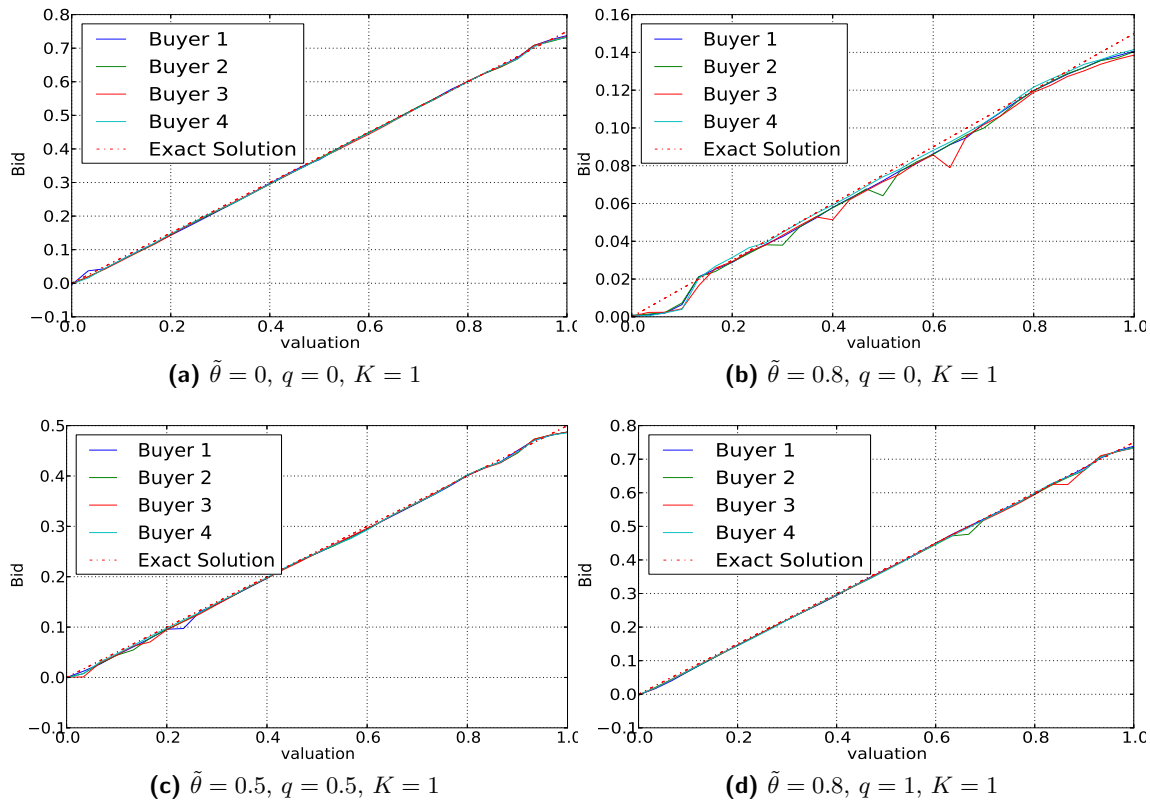
(a) $\tilde{\theta} = 0$, $q = 0$, $K = 1$

(b) $\tilde{\theta} = 0.8$, $q = 0$, $K = 1$

(c) $\tilde{\theta} = 0.5$, $q = 0.5$, $K = 1$

(d) $\tilde{\theta} = 0.8$, $q = 1$, $K = 1$

Figure 6.2: Simulated and exact best bidding strategies for Scenario 1: 4 bidders, valuations uniformly distributed on $[0, 1]$, 1 service on sale (i.e. K=1).

Scenario 2 considers as well symmetric bidders whose valuations are drawn from a uniform distribution on the interval $[0, 1]$, but in this case 2 services are on sale (i.e. K=2) and the number of bidders is either 3 or 6. Fig. 6.3 shows the algorithm's output along with the exact solution, for the two number of buyers considered, different percentages of reimbursement and probabilities of failure. For all cases the simulated results are good approximations of the exact functions.

Scenario 3 shows another symmetric situation, now for two buyers whose valuations are independently drawn from an exponential distribution and one service on sale. Results of the algorithm along with the exact expression for the bidding functions are shown in Fig. 6.4 for different values of the distribution's parameter λ, different amounts of reimbursements and probabilities of failure.

Scenario 4 considers 2 buyers that compete for one service, the valuations of both of them are drawn from a uniform distribution, for one of them over the interval $[0, 1]$ and for the other one over the interval $[0, 0.5]$. Thus, Scenario 4 is an asymmetric case, but in which the exact solution can still be derived (see Proposition (6.2)). A comparison of the exact formula and the algorithm's output is shown in Fig. 6.5.
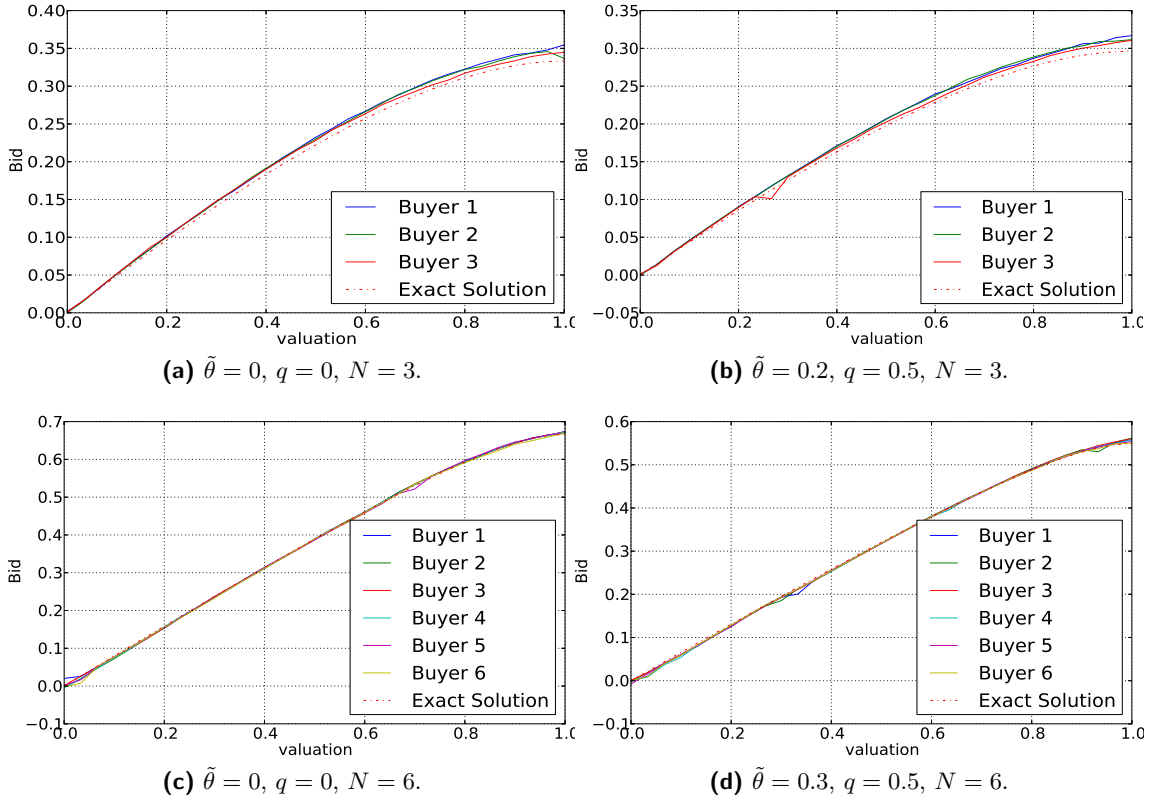
Figure 6.3: Simulated and exact best bidding strategies for Scenario 2: different number of bidders, valuations uniformly distributed on $[0, 1]$, 2 services on sale.

Finally, let us present some scenarios where an analytical expression is not available. We shall use the results of the lemmas in Section 6.3.2.B as guidelines to validate the results. Scenario 5 considers two of these cases. Results are shown in Fig. 6.6, and show that indeed the bidding functions are non decreasing, and that for symmetric bidders the obtained bidding functions coincide.

The results shown throughout this subsection allow to validate the algorithm design and implementation in several scenarios. We now comment on the algorithm's performance regarding its computational cost.

6.3.3.B.3   Enhancements and Code Acceleration

As aforementioned, the procedure relies on several Monte Carlo iterations, which renders the computational cost very elevated. The error obtained when approximating via Monte Carlo methods decreases with the root of the number of iterations. Hence, there is a trade-off between accuracy and running time. Having an accurate result may rapidly turn performance into an issue.

In order to tackle these potential computational problems, we have relied on PYTHRAN [9], an open-source static compiler for the Python language that makes it possible to turn a high level interpreted language into a static lower-level language, namely from Python to C++. The performance boost for our application reaches from x50 speed-up when compared to the original version. As the translation is completely automated, the performance improvement comes without code maintainability drawback.

The results presented above to validate the algorithm's implementation were computed on a regular computer with the following specifications. Intel®Core™ i5 CPU M 480 @ 2.67GHz × 4
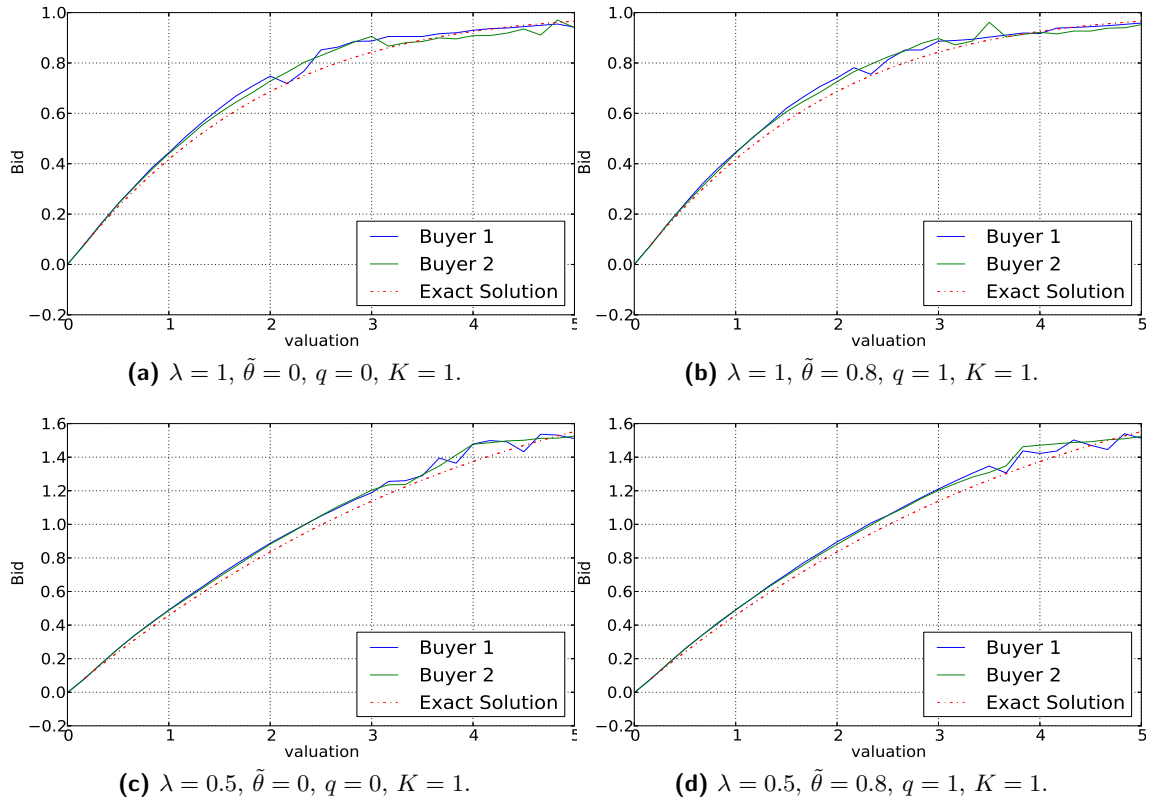
(a) $\lambda = 1$, $\tilde{\theta} = 0$, $q = 0$, $K = 1$.

(b) $\lambda = 1$, $\tilde{\theta} = 0.8$, $q = 1$, $K = 1$.

(c) $\lambda = 0.5$, $\tilde{\theta} = 0$, $q = 0$, $K = 1$.

(d) $\lambda = 0.5$, $\tilde{\theta} = 0.8$, $q = 1$, $K = 1$.

Figure 6.4: Simulated and exact best bidding strategies for Scenario 3: 2 bidders, valuations exponentially distributed, one service on sale.

(a) $\tilde{\theta} = 0$, $q = 0$.

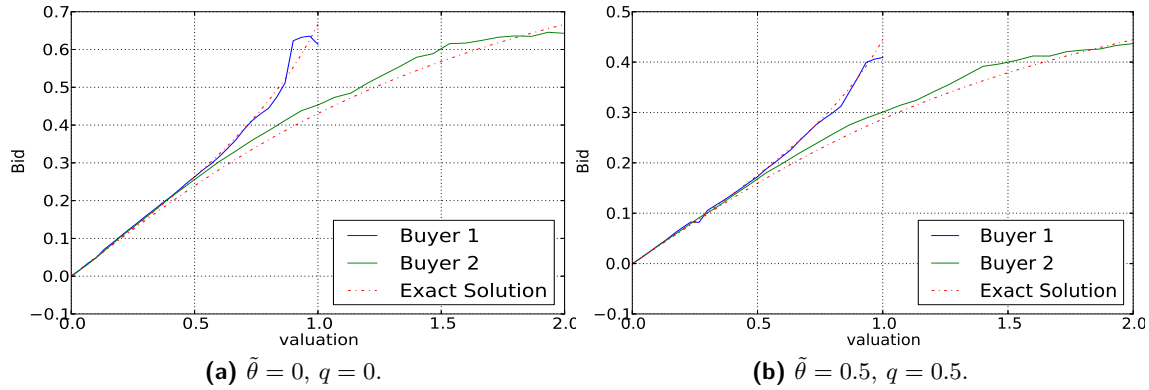**(b)** $\tilde{\theta} = 0.5$, $q = 0.5$.

Figure 6.5: Simulated and exact best bidding strategies for Scenario 4: 2 bidders, uniform valuations with different supports, one object.

with 3.5 GB of RAM memory running Ubuntu 12.04 Operative System. The average computation time for each one of the scenarios presented above are shown in Table 6.1, for both the Python implementation, and the C++ code. Results show that the speed-up is of at least 50x.

In order to further tackle the trade-off between the quality of the approximation through Monte Carlo and the computational cost, variance reduction techniques could be used, such as the importance sampling technique (see e.g. [134]). In that technique, samples are drawn from an auxiliary distribution which has to be carefully chosen in order to indeed result in a variance reduction of the estimator. However, the choice of this instrumental distribution and how to sample
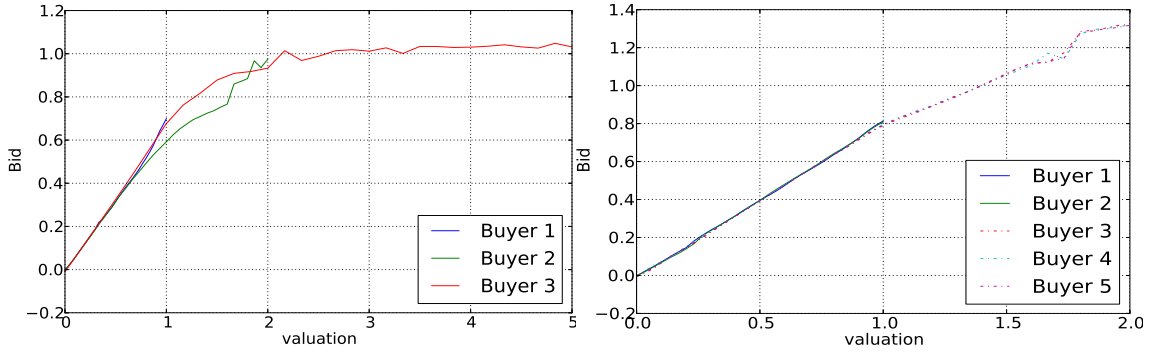
(a) Bidders 1 valuation $\sim U[0,1]$, $\tilde{\theta}_1 = 0.2$. Bidder **(b)** Bidders 1 and 2 with valuations $\sim U[0,1]$ and
2 valuation $\sim U[0,2]$, $\tilde{\theta}_2 = 0.9$. Bidder 3 valuation bidders 2 to 5 with valuations $\sim U[0,2]$, $\tilde{\theta} = 0$, $q = 0$,
$\sim exp(1)$, $\tilde{\theta}_3 = 0.9$. $q = 1$ and $K = 1$. $K = 1$.

Figure 6.6: Simulated best bidding strategies for Scenario 5: asymmetric general cases.

	Implementation		
Scenario	C++	Python	Ratio
1a.	8.28	425.60	51,40
1b.	6.93	414.91	59,87
1c.	8.58	421.64	49,14
1d.	8.59	422.70	49,21
2a.	6.43	354.14	55,10
2b.	6.30	344.10	54,62
2c.	13.03	681.79	52,32
2d.	13.33	654.40	49,10
3a.	10.14	733.90	72,38
3b.	8.65	700.00	80,92
3c.	16.27	1136.20	69,83
3d.	15.78	1067.90	67,67
4a.	8.87	777.18	87,62
4b.	3.98	232.04	58,30
5a.	25.83	1388.20	53,74
5b.	24.75	2379.10	96,13

Table 6.1: Average consumed time (seconds) to compute the best bidding strategies.

from it might not be easy from the theoretical point of view.

A graphical user interface was developed enriching the best bidding strategy computation
algorithm and providing a user-friendly tool, as shown in Fig. 6.7. In particular, in Fig. 6.7 the
approximate best bidding functions for two symmetric buyers whose valuations are uniformly
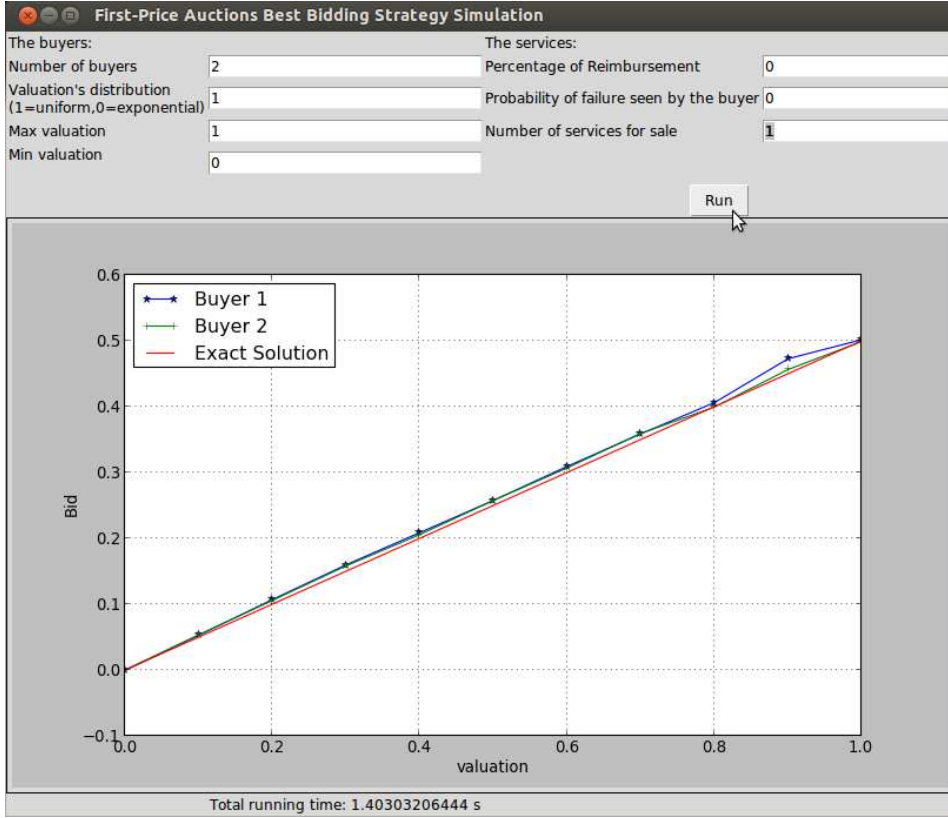distributed is shown along with the exact solution.

Figure 6.7: The First-Price auctions best bidding strategy simulator graphical interface.

6.4   Approximate Expected Seller's Revenue

In Chapter 5, with the aid of analytical results for the best bidding strategy, we have derived the expected seller's revenue considering three different scenarios, namely symmetric information, asymmetric information with naive buyers, and asymmetric information with rational buyers. For each of those different scenarios we have derived the optimal percentage of reimbursement q. In this chapter we focus on a case where no analytical results are available. We need thus to adapt the way we evaluate the expected seller's revenue to a simulative approach. For clarity's sake we shall focus our explanation on the asymmetric information with rational buyers scenario. That is, we recall, the scenario where buyers determine the estimated probability of failure $\tilde{\theta}_{i,s}$ such that their payoff is maximized. However, we shall see that from the results provided by this simulative approach, conclusions can be drawn for any of the three different scenarios.

Given valuation $x_{i,s}$ bidder i's expected payment is given by

$$
\begin{aligned}
p_{i,s} &= ProbWin \times AmountPaid \\
&= Pwin(\beta_i(x_{i,s}))\beta_i(x_{i,s}).
\end{aligned}
\tag{6.27}
$$

The *ex ante* payment is obtained taking the expectation of Equation (6.27), that is:

$$
E_{X_{i,s}}\{p_{i,s}(x_{i,s})\} = \int_0^{x_{i,s}^{max}} p_{i,s}(x_{i,s})f_{i,s}(x_{i,s})dx_{i,s}
\tag{6.28}
$$

The seller's *ex ante* expected revenue due to bidder i is $E_{x_{i,s}}\{p_{i,s}(x_{i,s})\}$ if there is no failure, and $E_{x_{i,s}}\{p_{i,s}(x_{i,s})\}(1 - q_s)$ if there is a failure. Failures occur with probability θ_s, thus the *ex*

ante expected revenue due to buyer i is $E_{x_{i,s}}\{p_{i,s}(x_{i,s})\}(1 - q_s\theta_s)$. Finally, the total seller's *ex ante* expected revenue is simply the sum of the revenue due to each buyer, that is:

$$\bar{R} = \sum_{i=1}^{M} E_{X_{i,s}}\{p_{i,s}(x_{i,s})\}(1 - q_s\theta_s) \tag{6.29}$$

Let us now compute the bidder's expected payoff, in an analogous way as we have performed it in Chapter 5. Bidder i's real mean payoff (that is considering the event of real failures) when his or her valuation is $x_{i,s}$ is:

$$P_{i,s}(x_{i,s}) = Pwin(\beta_{i,s}(x_{i,s}))[x_{i,s}(1 - \theta_s) - \beta_{i,s}(x_{i,s})(1 - q_s\theta_s)] \tag{6.30}$$

We obtain the *ex ante* expected payoff by computing the expectation of $P_{i,s}(x_{i,s})$ defined in Equation (6.30) with respect to random variable $x_{i,s}$ obtaining:

$$\bar{P}_{i,s} = E_{X_{i,s}}\{P_{i,s}(X_{i,s})\} = \int_{0}^{x_{i,s}^{max}} P_{i,s}(x_{i,s})f_{i,s}(x_{i,s})dx_{i,s} \tag{6.31}$$

Since in the asymmetric case we are considering that there are not necessarily analytical expressions for the probability of winning and for the best bidding strategies, we are not able to compute Equation (6.29) and Equation (6.31) analytically . We shall rather compute approximate values, using the Monte Carlo method. Let us call $\bar{R}^{MC}$ and $\bar{P}_{i,s}^{MC}$ the approximate *ex ante* expected revenue and *ex ante* expected payoff respectively. We compute those approximate expressions as:

$$\bar{R}^{MC} = \frac{1}{MC} \sum_{k=1}^{MC} \sum_{i=1}^{M} Pwin\left(\hat{\beta}_{i,s}(x_{i,s}^{\{k\}})\right) \hat{\beta}_{i,s}(x_{i,s}^{\{k\}})(1 - q_s\theta_s), \tag{6.32}$$

$$\bar{P}_i^{MC} = \frac{1}{MC} \sum_{k=1}^{MC} Pwin\left(\hat{\beta}_{i,s}(x_{i,s}^{\{k\}})\right) \left[x_{i,s}^{\{k\}}(1 - \theta_s) - \hat{\beta}_{i,s}(x_{i,s}^{\{k\}})(1 - q_s\theta_s)\right], \tag{6.33}$$

where MC is the number of iterations of the Monte Carlo method, Pwin is given by Equation (6.25), the best bidding strategies $\hat{\beta}_{i,s}$ are the approximate of $\beta_{i,s}$ obtained with the algorithm described in the previous section, and $x_{j,s}^{\{k\}}$ is the sample drawn from j's distribution $F_{j,s}$ in the k-th repetition of the Monte Carlo method.

For given values of $q, \theta, \tilde{\theta}$ we can thus assess $\bar{R}^{MC}$ and $\bar{P}_{i,s}^{MC}$ for each bidder by sampling MC valuations, each of them called $x_{i,s}^{\{k\}}$ and estimating $\hat{\beta}_{i,s}$ with the aid of the algorithm presented in the previous section.

The ultimate objective is, we recall, to find the optimal percentages of reimbursement q_s. For such purposes, we have introduced in Chapter 5 the pricing game. In few words, the pricing game in Chapter 5 is a Stackelberg game where the leader (i.e. the seller) proposes a value of q_s for each service and the followers (i.e. the buyers) choose the value of $\tilde{\theta}_{i,s}$ that maximizes their expected payoff for the value of q_s announced by the leader. The vector $\tilde{\theta}$ composed by the choices of $\tilde{\theta}_{i,s}$ of all the buyers constitutes a Nash Equilibrium of a game where the buyers are the players, $\tilde{\theta}_{i,s} \in [0, 1)$ is their space of actions and the *ex ante* expected payoffs are their utilities.

The backward induction method proposed in Chapter 5 to solve the pricing problem implies first solving an optimization problem to find the value of $\tilde{\theta}$ maximizer of the buyers' payoffs as a function of q, and with this value of $\tilde{\theta}$ going back to solve the seller's optimization problem to determine the optimal q. In order to adapt such methodology to simulations we define a grid for q_s $s \in S$ between 0 and a positive value q_s^{max}, and for each q_s a grid for $\tilde{\theta}_{i,s}$, $i = 1 \ldots M$ between 0 and $min\{1, 1/q_s\}$. Please note that following the indications of Chapter 5, we shall exclude value 1 for the probability of failure and limit it to $1/q_s$, which are necessary conditions for an equilibrium

as we have seen in Chapter 5. Let us denote the function that receives as an argument an interval and returns a discrete set of values within that interval as $Grid(\cdot)$.

We reformulate the pricing game for the asymmetric case and adapt it to simulations in Problem (6.1).

Problem 6.1 *The Pricing Game is a static Stackelberg game where:*

- *The leader is the seller and the buyers are followers*
- *The leader's set of available actions is $\{q = \{q_s\}_{s \in S} : q_s \in Grid([0, q_s^{max}])\}$*
- *Each follower i's set of available actions is the set $\{\tilde{\theta}_{i,s} \in Grid([0, min(1, \frac{1}{q_s}))\}$*
- *The leader's utility is $\bar{R}^{MC}$ and the followers' utilities are $\bar{P}_{i,s}^{MC}$*

For each $(M + |S|)$-tuple $(\tilde{\theta}_1 \dots \tilde{\theta}_M, q_1 \dots q_{|S|})$ the approximate seller's expected revenue $\bar{R}^{MC}$ and approximate buyers' expected payoff $\bar{P}_{i,s}^{MC}$ are computed. Finally, the solution to the Pricing game stated in Problem (6.1), is that one that provides the seller with the maximum expected revenue at the Nash equilibrium of the buyers. Please note that due to discretization, the procedure does not necessarily provide the optimal value of q but rather the maximum among the considered values. However, it allows us to gain intuition about whether the analytical results are still valid in asymmetric scenarios or not.

Computing the pure Nash equilibria can be a computational hard task when the number of bidders and types of services increase (see e.g. [58]). In consequence, for this stage, a high performance code is of paramount importance, all the more so because the best bidding strategies are to be updated for each $(M + |S|)$-tuple $(\tilde{\theta}_1 \dots \tilde{\theta}_M, q_1 \dots q_S)$ that has to be evaluated. The PYTHRAN tool, introduced in Subsection 6.3.3.B.3, allows us to obtain an overall high performance C++ code. In the following section we apply this simulative approach to two different case studies.

6.5   Case studies

We now evaluate our pricing scheme proposed in Chapter 5 and extended in the present chapter in two asymmetric cases. A third scenario is provided in Appendix C.
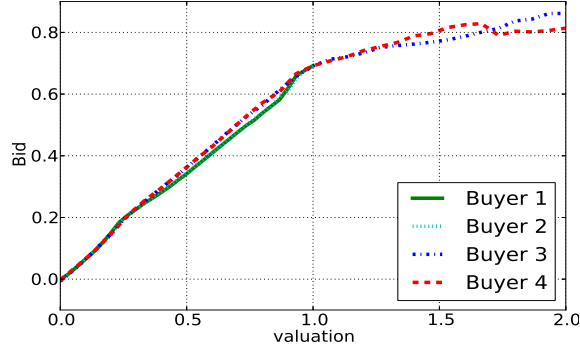
6.5.1   Non-Homogeneous Buyers Scenario

In this scenario we shall consider one type of service and buyers that value the service differently. This could account, for instance, for firms that have different business models. Since there is only one type of service on sale we drop sub-index s. The scenario is depicted in Fig. 6.8a. We shall consider two classes of buyers, say A and B, each one with 2 buyers. Buyers 1 and 2 belong to class A and have valuations which are distributed according to a uniform law on the interval $[0, 1]$. Buyers 3 and 4 belong to class B and have exponentially distributed valuations with a parameter $\lambda = 0.5$. This means that buyers in class A value the service between 0 and 1 with equal probability, while buyers belonging to class B can attach any positive valuation to the service, though higher valuations have lower probabilities. Finally, the service on sale fails 20% of the time (i.e. $\theta = 0.2$).

The approximate best bidding strategies obtained with the algorithm introduced in Section 6.3.3 when the buyers assume that the service does not fail (i.e. $\tilde{\theta}_i = 0$) and when there is no reimbursement (i.e. $q = 0$) are shown in Fig. 6.8b. Please note that the results verify the theoretical properties proved in Subsection 6.3.2.B. Indeed, the bidding functions are increasing with the valuations and for buyers belonging to the same class they coincide.

We proceed as explained in Section 6.4 to compute the approximate buyers' expected payoff and approximate seller's expected revenue, that is we approximate them by Monte Carlo, for different
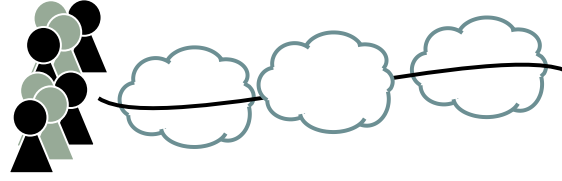
(a) Scenario illustration.

(b) Simulated best bidding strategies. $q = 0$, $\tilde{\theta}_i = 0$, $i = 1\ldots4$.

Figure 6.8: Case study with one class of service and non-homogeneous buyers.

	0	0.2	0.8
0	0.07,-0.04	0.07,0.01	0.09,**0.03**
0.2	0.1,0.0	0.1,0.01	0.12,**0.03**
0.8	**0.19**,0.02	**0.19**,0.02	**0.19**,0.03

(a) Buyers' Expected Payoff, $q = 0$.

	0	0.2	0.8
0	0.45	0.4	0.34
0.2	0.36	0.33	0.27
0.8	0.11	0.1	**0.07**

(b) Seller's Expected Revenue, $q = 0$.

Table 6.2: Non-homogeneous buyers scenario. Expected outcomes for different values of $\tilde{\theta}$ and $q = 0$.

	0	0.2	0.8
0	**0.1**,-0.04	**0.1**,-0.09	**0.1**,0.01
0.2	**0.1**,-0.06	**0.1**,0.01	**0.1**,0.01
0.8	**0.1**,**0.01**	**0.1**,0.0	**0.1**,0.0

(a) Buyers' Expected Payoff, $q = 1$.

	0	0.2	0.8
0	0.36	0.74	**0.33**
0.2	0.38	**0.33**	**0.34**
0.8	**0.35**	0.37	0.34

(b) Seller's Expected Revenue, $q = 1$.

Table 6.3: Non-homogeneous buyers scenario. Expected outcomes for different values of $\tilde{\theta}$ and $q = 1$.

values of q and $\tilde{\theta}_i$. We shall consider $q \in \{0, 1\}$ and $\tilde{\theta}_i \in \{0, 0.2, 0.8\}$ for $i = 1\ldots4$. We recall that we assume that buyers whose valuations' are drawn from the same distribution function and that buy the same service, estimate the same probability of failure of the service. Hence, a symmetry argument can be readily applied to each class to conclude that buyers of class A would assume the same $\tilde{\theta}_i$ at the Nash equilibrium, and likewise for class B. Thus, since we know that at equilibrium rational buyers belonging to the same class have as optimum the same value of probability of failure, we only evaluate those cases. In particular, in this scenario that means that we have to evaluate 3^2 options instead of 3^4. Results for the approximate buyers' expected payoff and seller's expected revenue are shown in Table 6.2 and Table 6.3, for $q = 0$ and $q = 1$ respectively.

Tables should be interpreted as follows. In Table 6.2a and Table 6.3a the evaluated values of $\tilde{\theta}_i$ are listed for classes A and B, A corresponding to the rows and B to the columns. Each entry in the table represents the approximate expected payoff of a buyer from class A and class B, in

that order. Nash equilibrium occurs for the values of $\tilde{\theta}_1 = \tilde{\theta}_2$ and $\tilde{\theta}_3 = \tilde{\theta}_4$ which index a cell where both payoffs are highlighted in bold. That is, for $q = 0$ equilibrium occurs at $\tilde{\theta}_i = 0.8$, $i = 1 \ldots 4$ and for $q = 1$ at the following values of vector $\tilde{\theta}$: $(0.8, 0.8, 0, 0)$, $(0.2, 0.2, 0.2, 0.2)$, $(0, 0, 0.8, 0.8)$ and $(0.2, 0.2, 0.8, 0.8)$.

Once the Nash equilibria for the buyers are found for each value of q, we shall approximate the seller's expected revenue at those equilibria. These results are highlighted in bold in Table 6.2b for $q = 0$ and in Table 6.3b for $q = 1$. Finally, we shall choose as the optimum value of q the one that renders the greatest approximate seller expected revenue. In this case the optimum is $q = 1$ and implies that the seller's expected revenue is ~ 0.34 units, instead of the 0.07 units that would be earned when there is no reimbursement at all (i.e. when $q = 0$).

Results also show that in this case, as in the symmetric buyers case studied in Chapter 5, when buyers assume a probability of failure equal to the real probability of failure, that is when $\tilde{\theta}_i = \theta = 0.2$ for $i = 1 \ldots 4$, the expected outcomes for buyers and seller are the same whether reimbursement is set to 0 or to 100%. Indeed, for the buyers we can see in Table 6.2a that the expected payoff when no reimbursement is in place for the entry $(0.2, 0.2)$ is $(0.1, 0.01)$, the same value obtained when reimbursement is 100%, as shown in Table 6.3a for the entry $(0.2, 0.2)$. Likewise, for the seller's expected revenue, the entries indexed by $(0.2, 0.2)$ in Table 6.2b for $q = 0$ and in Table 6.3b for $q = 1$ are in both cases 0.33 units.

Finally, the *market for lemons* phenomenon commented in Chapter 5 also arises in this scenario. Indeed, in Table 6.2a we can see that at the Nash equilibrium buyers assume a probability of failure equal to 0.8 when reimbursement is set to 0, that is, they assume a probability of failure greater than the real one and this diminishes the seller's expected revenue.
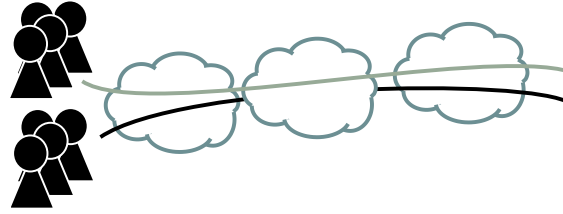
6.5.2 Non-homogeneous Services Scenario

In this scenario we shall consider that buyers are homogeneous and that two different services are offered. The service of type A fails 20% of the time, that is $\theta_A = 0.2$, and the service of type B 50% of the time, that is $\theta_B = 0.5$. There is capacity enough for serving only one service at a time. The scenario is illustrated in Figure 6.9a. Since there are two different services offered, the optimal reimbursement we are looking for is a vector $q = (q_A, q_B)$ corresponding to the reimbursement for each service.

We shall consider four buyers, two who bid for obtaining a service of type A and two who bid for obtaining a service of type B. Buyer's valuations are uniformly distributed on $[0, 1]$. Their approximate best bidding strategies when $q_A = 0$, $q_B = 1$, $\tilde{\theta}_{1,A} = \tilde{\theta}_{2,A} = 0.2$ and $\tilde{\theta}_{3,B} = \tilde{\theta}_{4,B} = 0.5$. are shown in Fig. 6.9b. Please note that as proved in Subsection 6.3.2.B, the strategies are increasing with the valuations and symmetric buyers assuming the same probability of failure have the same bidding functions.

We proceed as in the previous scenario to compute the outcomes of buyers and seller. Results for the buyers' expected payoffs and for the seller's expected revenue are shown in Table 6.4 to Table 6.7, each corresponding to a different reimbursement vector q. Tables should be read as explained in the previous scenario. The optimal reimbursement in this scenario is once again reimbursing 100% for both services. In this case, the seller instead of getting an expected revenue of $0.12, 0.16$ or 0.23 units, would obtain 0.39 units by reimbursing 100%.

Besides, the *market for lemons* phenomenon is also observable in this scenario. Indeed, when there is no reimbursement the equilibrium of the buyers strategies occurs at the highest value of $\tilde{\theta}_i, s$, that is $\tilde{\theta}_i, s = 0.8$ in Table 6.4a.

Previous results were obtained on a regular computer with a i5 processor of 2.67GHz and 3.6 GB of RAM memory running OS Ubuntu 12.04. The time needed to compute all the results related to the first scenario is on the average 3.5 minutes and for the second scenario 15 minutes.

(a) Scenario illustration.
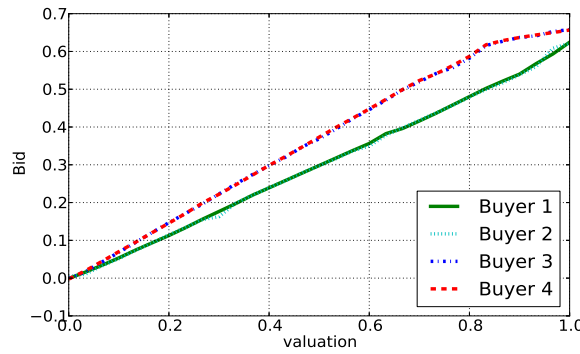
(b) Simulated best bidding strategies. $q_A = 0$, $q_B = 1$, $\tilde{\theta}_{1,A} = \tilde{\theta}_{2,A} = 0.2$ and $\tilde{\theta}_{3,B} = \tilde{\theta}_{4,B} = 0.5$.

Figure 6.9: Case study with two classes of services and homogeneous buyers.

	0	0.2	0.5	0.8
0	0.01,-0.05	0.02,-0.02	0.04,0.02	0.05,**0.07**
0.2	0.04,-0.04	0.04,-0.02	0.06,0.02	0.08,**0.07**
0.5	0.08,-0.02	0.09,-0.0	0.09,0.03	0.1,**0.07**
0.8	**0.14**,-0.01	**0.13**,0.02	**0.13**,0.05	**0.13**,**0.07**

(a) Buyers' Expected Payoff, $q = (0,0)$.

	0	0.2	0.5	0.8
0	0.6	0.53	0.4	0.27
0.2	0.53	0.48	0.36	0.23
0.5	0.4	0.37	0.3	0.17
0.8	0.28	0.23	0.17	**0.12**

(b) Seller's Expected Revenue, $q = (0,0)$.

Table 6.4: Non-homogeneous Services Scenario: $\theta_A = 0.2$, $\theta_B = 0.5$. Expected outcomes for different values of $\tilde{\theta}$ and $q = (0,0)$.

	0	0.2	0.5	0.8
0	0.01,**0.03**	0.01,**0.03**	0.01,0.02	0.01,**0.03**
0.2	0.04,**0.03**	0.04,**0.03**	0.04,**0.03**	0.04,**0.03**
0.5	0.09,**0.04**	0.09, **0.04**	0.09,**0.04**	0.09,**0.04**
0.8	**0.14,0.05**	**0.14,0.05**	**0.14,0.05**	**0.13,0.05**

(a) Buyers' Expected Payoff, $q = (0, 1)$.

	0	0.2	0.5	0.8
0	0.45	0.45	0.45	0.45
0.2	0.39	0.39	0.39	0.39
0.5	0.28	0.28	0.28	0.28
0.8	**0.16**	**0.17**	**0.17**	**0.16**

(b) Seller's Expected Revenue.

Table 6.5: Non-homogeneous Services Scenario: $\theta_A = 0.2$, $\theta_B = 0.5$. Expected outcomes for different values of $\tilde{\theta}$ and $q = (0, 1)$.

	0	0.2	0.5	0.8
0	**0.04**,-0.05	**0.05**,-0.02	**0.07**,0.02	**0.08,0.07**
0.2	**0.04**,-0.05	**0.05**,-0.02	**0.07**,0.02	**0.08,0.07**
0.5	**0.04**,-0.05	**0.05**,-0.02	**0.07**,0.02	**0.08,0.07**
0.8	**0.04**,-0.05	**0.05**,-0.02	**0.07**,0.02	**0.08,0.07**

(a) Buyers' Expected Payoff, $q = (1, 0)$.

	0	0.2	0.5	0.8
0	0.55	0.47	0.34	**0.23**
0.2	0.54	0.47	0.35	**0.23**
0.5	0.54	0.48	0.35	**0.23**
0.8	0.55	0.47	0.35	**0.23**

(b) Seller's Expected Revenue, $q = (1, 0)$.

Table 6.6: Non-homogeneous Services Scenario: $\theta_A = 0.2$, $\theta_B = 0.5$. Expected outcomes for different values of $\tilde{\theta}$ and $q = (1, 0)$.

	0	0.2	0.5	0.8
0	**0.04**, 0.02	**0.04,0.03**	**0.04, 0.03**	**0.04,0.03**
0.2	**0.04,0.03**	**0.04,0.03**	**0.04, 0.03**	**0.04,0.03**
0.5	**0.04,0.03**	**0.04**,0.02	**0.04, 0.03**	**0.04,0.03**
0.8	**0.04,0.03**	**0.04,0.03**	**0.04, 0.03**	**0.04,0.03**

(a) Buyers' Expected Payoff, $q = (1, 1)$.

	0	0.2	0.5	0.8
0	0.39	**0.39**	**0.39**	**0.39**
0.2	**0.39**	**0.39**	**0.39**	**0.39**
0.5	**0.39**	0.39	**0.39**	**0.39**
0.8	**0.39**	**0.39**	**0.39**	**0.39**

(b) Seller's Expected Revenue, $q = (1, 1)$.

Table 6.7: Non-homogeneous Services Scenario: $\theta_A = 0.2$, $\theta_B = 0.5$. Expected outcomes for different values of $\tilde{\theta}$ and $q = (1, 1)$.

6.6 Summary

In this chapter we have focused on evaluating the pricing mechanism proposed in Chapter 5 when relaxing the assumption about buyers' symmetry and a single type of service. When these assumptions are dropped, in the general case, there is no closed-form expression for the best bidding strategy, which constitutes the main pillar of the theory derived in Chapter 5. We have proposed a numerical method that makes it possible to obtain an approximate best bidding strategy when those assumptions are relaxed. The algorithm is based on the Monte Carlo method, which imposes a trade-off between accuracy and computational cost. In order to overcome this trade-off, we have used an automatic translating tool from a high level programming language to native code. The resulting tool performs at least 50 times faster than the original one. This acceleration has allowed us to rely on the developed algorithm to compute the optimum percentage of reimbursement in scenarios with asymmetric buyers and services which fail with different probability.

We have evaluated the pricing scheme through simulations in two different scenarios. In both cases, results have shown that reimbursing 100% provides greater revenues to the seller than not reimbursing at all. A third case is shown in Appendix C. These results verify the theoretical conclusions drawn for the symmetric case in Chapter 5.

Part III

Conclusion

Chapter 7

Conclusion and Perspectives

7.1 Conclusion

In this thesis we have addressed different topics related to the provisioning of Quality of Service (QoS) in interdomain networks. We have divided our study into two different parts. The first part has addressed a problem that arises when computing end-to-end QoS paths. The second part has addressed topics related to multidomain alliances, namely bandwidth allocation, revenue sharing and pricing.

In the first part, we have focused on a problem related to interdomain quality-assured path computation. Several methods to compute end-to-end quality-assured paths have been proposed, mainly based on the Path Computation Element (PCE) framework. However, most of them need QoS metrics related to each domain, or autonomous system (AS). How to compute or how to obtain values related to these metrics is typically not specified by those. In addition, even if a method were available, for instance, end-to-end monitoring information, anomalies in the traffic occur, and those values might no longer be accurate when the time comes to stablish the path. We have conceived a means to compute a bound on the end-to-end delay of traversing an AS as a tool to be used in this context. The method takes into account the uncertainties in the traffic traversing an AS. This uncertainty has been modelled as a polytope in the so-called traffic matrix. This is a robust approach, since the computed value can be announced by the AS in the process of interdomain path selection and it is guaranteed that the bound will hold for a certain period of time. The problem of how to compute this bound was mathematically stated and we have shown that it results to be a non-convex optimization problem, thus standard optimization tools are not suitable to solve it. We have proposed a reformulation of the problem and proved that its solution is found at the extreme points of a polytope obtained by transforming the original polytope, that one that models the traffic uncertainties. In addition, an alternative approximate method was proposed, as a remedy for the high computational cost of the exact method. The approximate method renders rather than a value of the delay, an interval to where the real value of the maximum delay is guaranteed to belong, and this interval can be made arbitrarily small. The latter was theoretically proven and numerically validated, by comparing the results of the approximate method against the real maximum. Both methods, exact and approximate, were tested on real topologies using measurement and synthetic data. The Approximate method was shown, through simulations, to provide acceptable computational times on several scenarios.

In the second part, we have focused on multidomain alliances. Multidomain or Network Service Providers (NSPs) or AS alliances, are an emerging scenario where several ASs get together in order to provide end-to-end QoS, with the objective of overcoming technical and business difficulties of end-to-end QoS delivery. For doing so, coordination principles take place with respect to both technological and economic aspects. For instance, Traffic Engineering techniques are applied to the whole alliance, a common monitoring infrastructure exists and common pricing and revenue sharing principles are in place.

In this context, we have proposed a framework for covering the complete cycle for selling end-to-end quality assured services.

Firstly, we have stated the problem of network bandwidth allocation with QoS constraints and an application based on first-price auctions combined with such problem was proposed as a means to sell quality assured services. In particular, the bandwidth allocation problem was stated as a Network utility Maximization (NUM)-like problem, where end-to-end QoS constraints are considered, and where utility functions are obtained through an auction mechanism. The mechanism, thus, maximizes the revenue of the alliance, while respecting QoS constraints. We have proved that a distributed solution to the allocation problem can be carried out.

Secondly, we have addressed the problem of revenue sharing in the context of ASs alliances, focusing on the case where the income of the alliance is determined by the output of the aforementioned NUM problem. This particular scenario poses new challenges. Indeed, existing methods to perform revenue sharing, which have been reviewed in the thesis, were found to be inappropriate applied to this case. Our contribution in this sense is twofold. We have formally stated the desired properties for the revenue sharing in the context of multidomain alliances and we have proposed a new method to compute the shares. The method is conceived to provide economic stability and efficiency to the alliance and it is flexible enough to be adapted to fulfil additional properties. It is based on solving an optimization problem and considers statistics on the revenues. Its proper behaviour has been evaluated through simulation studies. In particular, we have shown through simulations that, provided the right choice in the objective function of the optimization problem is done, the method presents a proper behaviour. Indeed, it has shown through simulations over several topologies, to provide a fair share and incentives to the ASs belonging to the alliance to increase their resources towards the alliance.

Thirdly, we have provided insight into a pricing mechanism based on first-price bandwidth auctions with reimbursement, which to the best of our knowledge has not been proposed before. The main contribution with respect to this pricing scheme is twofold. We have determined the best bidding strategy under a first-price auction mechanism for network services that are prone to failures and with money-back guarantees; and we have analytically obtained the optimal reimbursement value when bidders are aware of the reimbursement policy.

We have addressed the problem of where to set the percentage of reimbursement assuming different situations and buyer behaviours. In particular, we have studied the situation of rational buyers that are not certain about the probability of failure of the service. In that scenario, we have modelled the pricing problem as a Stackelberg game and shown that there is an equilibrium setting that maximizes seller's revenue, which is given by a reimbursement equal to 100%. In such equilibrium, the *market for lemons* effect and the *moral hazard* one, which were shown to arise for other percentage of reimbursement either than 100%, are overcome and seller's expected and buyers' expected payoff are the same as when having perfect information.

Finally, we have presented a simulative approach in order to evaluate the aforementioned proposed pricing scheme in scenarios where buyers are asymmetric. This accounts for more realistic scenarios, where, for instance, buyers value the service on sale in a different way. The proposed simulator finds an approximation to the best bidding strategies and it is based on an iterative process seeking to maximize each bidder's expected payoff. It relays on the Monte Carlo method to compute expectations and in the Simulated Annealing method to solve the involved optimizations. We have applied the simulator to three different scenarios obtaining in all of them results that are coherent with the results obtained through the analytical analysis, that is, reimbursing 100% was found to provide more revenues to the seller (i.e. the alliance) than no reimbursing at all.

7.2   Perspectives

The subjects addressed by this thesis need an holistic approach. Indeed, technological issues arise but also economic ones must be considered, while the interests of different actors (final users, NSPs, OTTs) may be opposed. However, our approach was rather compartmentalized into different studies. In order to be able to address the problem we have needed to decouple it and to state a

number of hypothesis. Natural perspectives are, thus, to relax those hypothesis and to reassemble all the problem. We detail this in what follows.

With respect to the first part of the thesis, the bound on the delay, we shall explore the case of having uncertainty on the AS topology in addition to traffic uncertainty. For instance, taking into account the case of link or node failures, and being able to provide even in those cases a tight end-to-end delay bound. We shall as well explore the possibility of building a delay curve as a function of ingress traffic. If the delay can be advertised as a function of ingress traffic, this would allow to take more sophisticated routing decisions, leading to lower end-to-end delay values.

As regards the proposed bandwidth allocation mechanism, a straightforward enhancement is considering multipath routes, and showing that even in that case distributed solutions can be found. In addition, we have assumed throughout the study a delay function modelled by the average delay of a M/M/1 queue. This model could be enhanced, for instance through measurements as proposed in [87], through more complex queue models or even using the maximum delay curve mentioned in the previous paragraph.

With respect to the revenue sharing method there is much to do. We have shown through exhaustive simulations that the method behaves well with respect to the properties that we named monotonicity and fairness. The detailed analytical study of these properties was left for future work. It is a very complicated task, since the problem is mathematically complex, no closed-form was found for the expression of the revenue sharing and this renders the proofs hard to derive. In addition, a very interesting challenge is to apply this method in a real case. In order to compute the revenue sharing, all billing information, and more, is needed, which could render the task complicated in a real system. Indeed, billing in telecommunication networks is itself a very complicated task. In our favour, we have to say that the problem does not need to be solved in real time, which from that point of view is much more simpler than the billing problem.

Still about the revenue sharing problem, and regarding the reassemble of the whole problem, we have not considered end-to-end quality constraints in the revenue sharing problem. It would be very interesting to study the impact of the delay on the obtained revenue share. To verify if the properties of the method are still fulfilled, and to what extent the consideration of those additional constraints impacts the implementation of the method.

Concerning the proposed pricing scheme, the first thing to be done is to enhance the model. The probability of failure in our model does not depend on the accepted bandwidth, it is rather a parameter that can account, for instance, for equipment failures or busy servers. We would like to enhance this, by, for instance, having a probability of failure that could account for the probability of exceeding a delay, given that this is a function of the traffic in the network. In addition to provide a more realistic model, this would add feedback from the accepted traffic into pricing.

Also with respect to the proposed pricing scheme, we would like to relax the assumption of independent bidders in order to take into account collusion effects. Collusion can arise if bidders agree to submit a low bid in order to pay a lower price for the service. In this case, since there is an agreement among certain buyers, two of the assumed hypothesis are no longer valid. First, the symmetric buyers one, if some buyers collude but not all, symmetry among all bidders is no longer true. Second, if buyers agree the independence among buyers does no longer hold. Simulative approaches are likely to be needed if both hypothesis are relaxed at the same time.

With respect to the scenario with asymmetric buyers, there are enhancements to be done to the simulative approach we have presented to approximate the best bidding strategies. Even if the computational cost of the implemented algorithm has been dramatically decreased though an automatic translation into native code, it could be further decreased, for instance, through code paralellization. In addition, the method's performance should be enhanced to provide a better performance when the number of bidders is increased. A solution to study in order to tackle both problems convergence time and accuracy of the results, are the variance reduction methods, since the algorithm relays much on the Monte Carlo method.

Finally, regarding the proposed pricing scheme, it would be extremely interesting to contrast the theoretical results with real data about the willingness to pay of buyers that are proposed to be

reimbursed if a failure occurs. Would they indeed act rationally? or are they naive? Does Quality-of-Experience have more influence on willingness to pay than an advertised QoS? Experimental results should be gathered in order to study these aspects.

Beyond the aforementioned enhancements of the proposed models and improvements of the provided solutions, it would be very interesting to study the dynamics of the NSPs alliances. Throughout the thesis, we have assumed that the alliance was somehow formed, and we have not deepen on how that occurs. While it is true that NSPs would get together more because of business reasons and less because of technical ones, some technical aspects could also have influence on how the alliances are formed. In this sense, it is very interesting to study the following questions. What would be guidelines to form an alliance? Could they be form such that the interdomain connections optimize the Internet routing? Can they lead to a better planned interdomain routing diminishing, for instance, BGP table sizes? And more related to the business plane: how does the share of each NSP changes when a NSP joins or leaves the alliance? How would buyers react towards an unstable alliance? Would they distrust it and a phenomenon similar to the *market for lemons* arise? Once more, these questions should be addressed through a multidisciplinary optic, where, for instance, concepts from game theory, microeconomics and networking are blended.

Part IV

Appendices

Appendix A

Bandwidth Allocation Review of Preliminaries Results

In this appendix we recall some results from Dynamic Systems that are used for the proof of the convergence of the primal-dual laws presented in Chapter 3 for the distributed bandwidth allocation problem.

Definition A.1 *Invariant Set. A set M is said to be positively invariant if $x(0) \in M$ implies $x(t) \in M \ \forall t \geq 0$.*

Theorem A.1 *Krasovskii [83]. Given $\dot{x} = f(x)$, x in $X \subset \mathbb{R}^n$ and a Lyapunov function of the form:*

$$V(x) = \dot{x}^T Q \dot{x}.$$

If a positive definite symmetric matrix Q is found such that

$$\left(\frac{\partial f}{\partial x}\right)^T Q + Q \left(\frac{\partial f}{\partial x}\right)$$

is negative semidefinite $\forall x \in X$, then $\dot{V} \leq 0$.

Theorem A.2 *LaSalle Invariance Principle (see e.g. [81]). Let $\Omega \subset D \subset \mathbb{R}^n$ be a compact positively invariant set w.r.t. $\dot{x} = f(x)$. Let $V : D \to \mathbb{R}$ be a continuously differentiable function such that $\dot{V}(x) \leq 0$ in Ω. Let $E \subset \Omega = \{x \in \Omega | \dot{V}(x) = 0\}$. Let M be the largest invariant set in E, then every solution that starts in Ω tends to M when $t \to \infty$.*

Theorem A.3 *LaSalle Generalized Invariance Principle (for Switched Systems) [95]. Given $\dot{x} = F(\sigma, x)$, $\sigma \in \Sigma$ (finite set), x in $X \subset \mathbb{R}^n$ and Ω compact invariant set in $\Sigma \times X$. If there exists $V : \Omega \to \mathbb{R}$ such that:*

1. *For σ fixed, $V(\sigma, x)$ is continuously differentiable w.r.t. x and $\dot{V}(\sigma, x) = \frac{\partial V}{\partial x} \dot{x} \leq 0$*

2. *At switching times, $V(\sigma(t^+), x) \leq V(\sigma(t^-), x)$. Then:*

Any trajectory $(\sigma(t), x(t))$ starting from Ω (as $t \to \inf$) approaches the largest invariant set inside the set of points (σ, x) for which:

i. *either σ fixed and $\dot{V}(\sigma, x) = 0$, $\forall t \geq 0$*

ii. *or σ changes at time t between σ^- and σ^+ but $V(\sigma^+, x) = V(\sigma^-, x)$*

Appendix B

Revenue Sharing, Simulations

In this chapter we present the results of exhaustive simulative studies, which evaluate the behaviour of the revenue sharing method proposed in Chapter 4. In particular, we focus on the projection of the contributions vector v into the stable and efficient set. That is to say, we consider $f(x) = ||x - v||^2$, which is the function that has shown good behaviour in the simulative results present in Chapter 4. Simulations evaluate the fulfilment of the monotonicity and fairness properties.

Network topologies were generated using the automatic Internet topology generator BRITE [1], which automatically generates a graph with directed links and randomly assigns capacities, and receives as input the number of nodes and certain parameters for the algorithms running inside. Over each topology, we have defined an overlay alliance where services are defined between every couple of nodes, if a path in the topology between those nodes exists. For those services, the second shortest path between ingress and egress node has been chosen. The evaluated network topologies are shown in Fig. B.1, while services' paths and nodes' capacities are shown in Table B.1. With respect to utility functions, they ware built up according to the method proposed in Chapter 3, that is to say ordering and summing up bids. The bids' values were randomly generated.

(a) Topology 4.

(b) Topology 5.

(c) Topology 6.
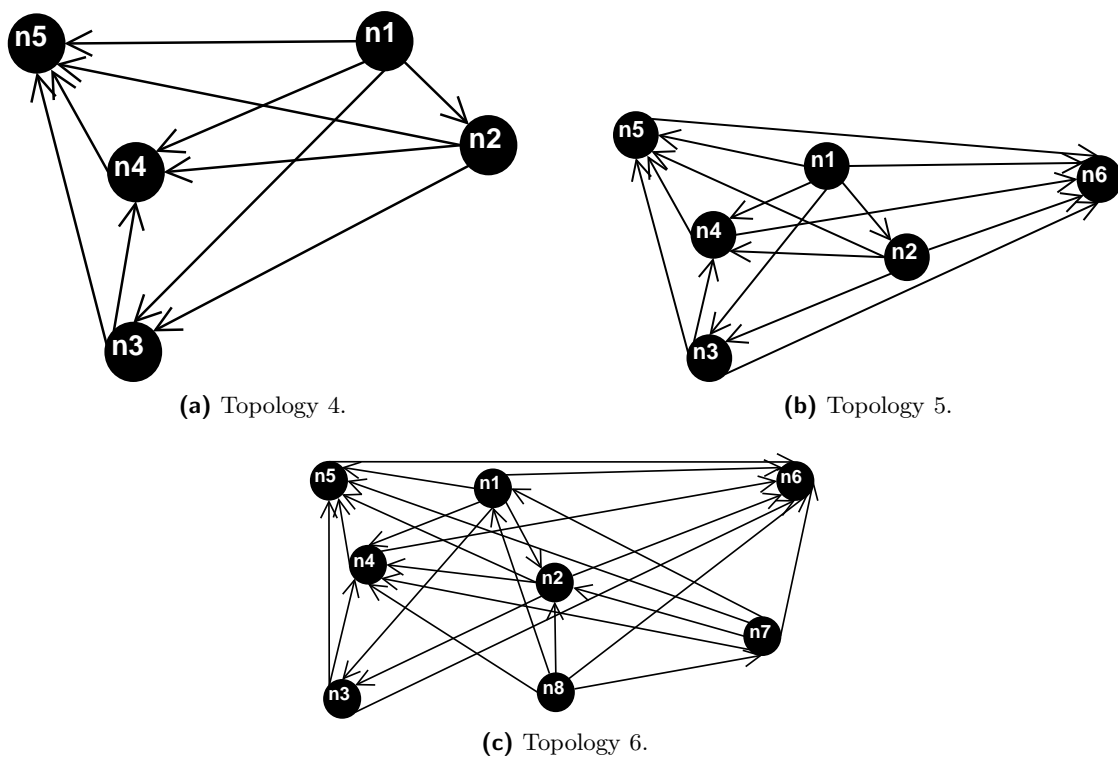
Figure B.1: Networks used on simulations.

Network	Paths	Nominal capacities
Topology 4	[n1,n2,n3] [n1,n2,n4] [n1,n2,n5] [n2,n3,n4] [n2,n3,n5] [n3,n4,n5]	(788,582,483,518,479)
Topology 5	[n1,n2,n3] [n1,n2,n4] [n1,n2,n5] [n1,n2,n6] [n2,n3,n4] [n2,n3,n5] [n2,n3,n6] [n3,n4,n5] [n3,n4,n6 [n4,n5,n6]	(641 , 630, 1072, 877, 788,181)
Topology 6	[n1,n2,n3] [n1,n2,,n4] [n1,n2,,n5] [n1,n2,,n6] [n2,n3,,n4] [n2,n3,,n5] [n2,n3,,n6] [n3,n4,,n5] [n3,n4,n6] [n4,n5,n6] [n1,n7,n8] [n1,n2,,n7] [n1,n2,,n8] [n2,n3,,n7] [n2,n3,n8] [n1,n4,n7] [n1,n4,n8] [n1,n5,n7] [n2,n5,n8] [n1,n6,n7] [n1,n6,n8]	(477,856,807,688,892,380,550,412)

Table B.1: Description of the paths for the different evaluated networks shown in Fig.B.1.

B.1   One-shot Simulations

Over each alliance we have evaluated the revenue share obtained using the proposed method when varying the capacity of one node at a time. Results are shown in Fig. B.5 to Fig. B.8. Each figure shows the revenue shares obtained when using the proposed method with objective function $f(x) = ||x - v||^2$ along with the contributions vector, as functions of the capacity. In each figure, the capacity of only one nodes varies. Shares are stacked up, the bottom one being the share corresponding to the node whose capacity is increasing. The legends respect the same order as the order in which shares are stacked up. The contributions vector is plotted along with the revenue shares and bars respect the same order.

B.1.1   Monotonicity evaluation

The monotonicity property was verified in all the alliances for all the evaluated utility functions. Please note that the results shown in this Appendix correspond to one given utility function. However, different utility functions lead to different revenue shares. As mentioned in Chapter 4, the particularities of our alliances make that the results do not only depend on the topology of the alliance but also on the utility functions. The monotonicity property was however, verified in all the evaluated cases, further than the ones shown in this Appendix. Monotonicity can be readily be checked from the results in Fig. B.5 to Fig. B.8, where for all the cases the share at the bottom of the bars does not decrease when capacity increases.

B.1.2   Fairness evaluation

Fairness can be evaluated by comparing the revenue share with the contributions vector. Each figure shows the revenue share results along with the contribution vector, when the capacity of one node increases while the capacity of all other nodes remain constant. Results show for all simulations that the revenue shares verify the *order preserving* property and the *no free riders* property. Indeed, the order imposed by the contributions vector is respected by the obtained shares, and in any case a node whose contribution is zero receives a non-zero revenue share. With respect to the *equals treatment of equals* property, this was verified in all cases, as it can be roughly verified from the figures.
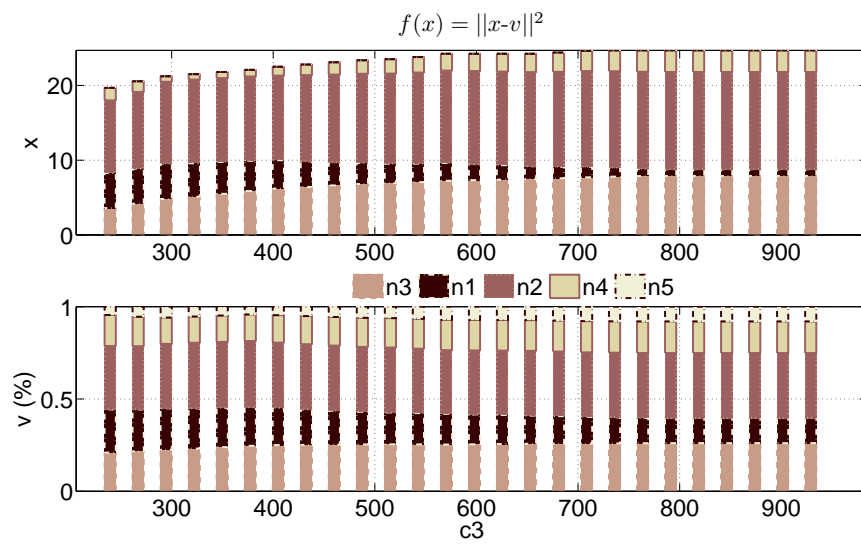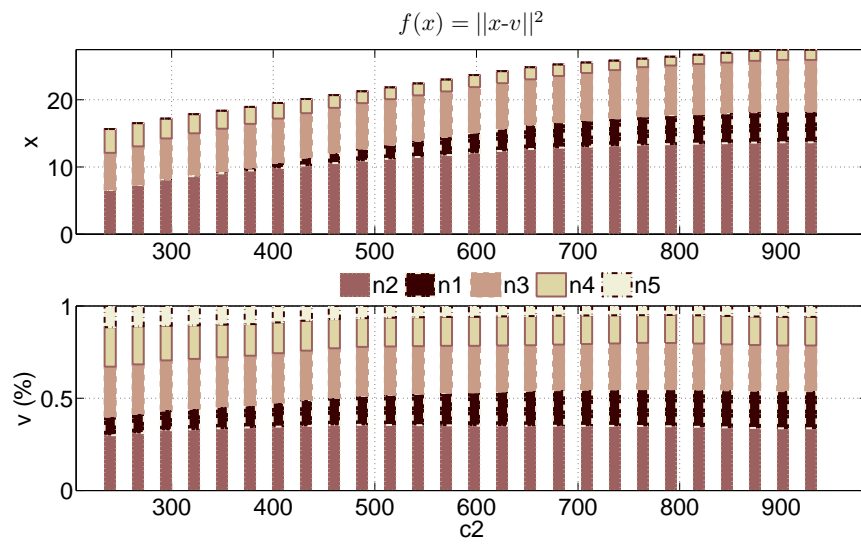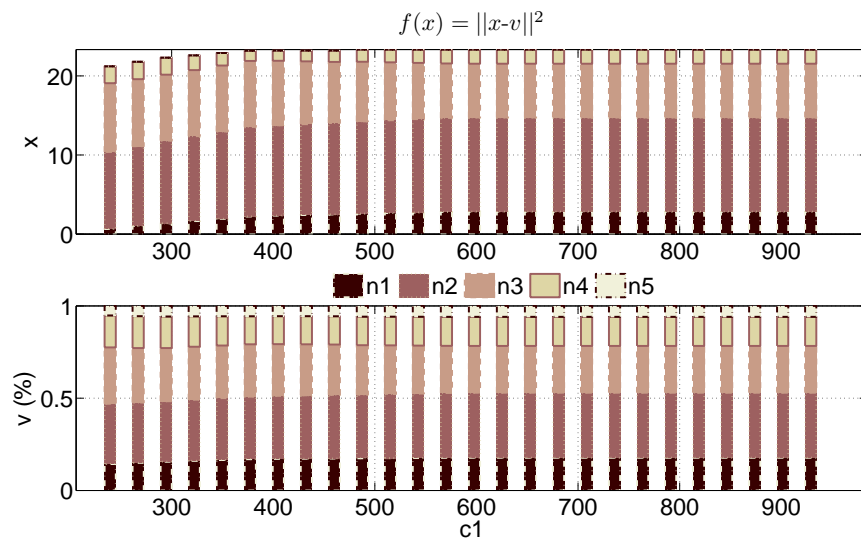
(a) n1.

(b) n2.

(c) n3.

Figure B.2: Revenue sharing when increasing the capacity of one node at a time for Topology 4 nodes 1 to 3.
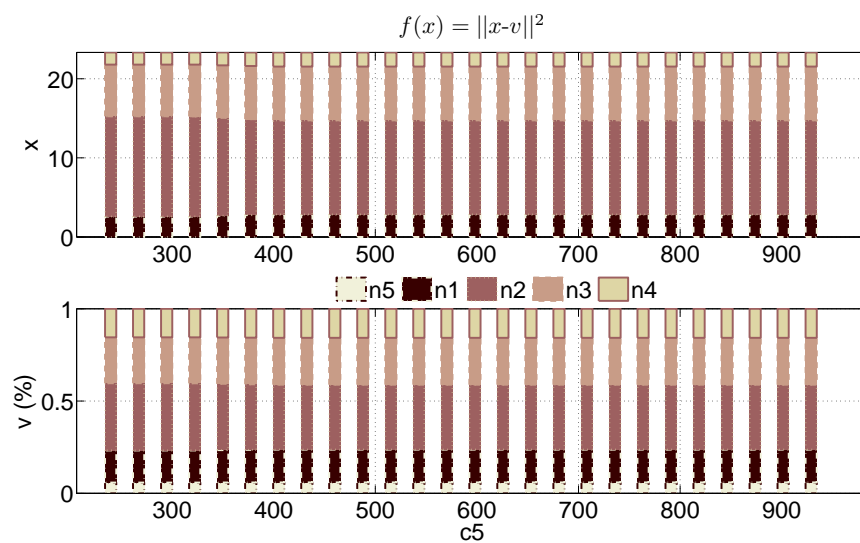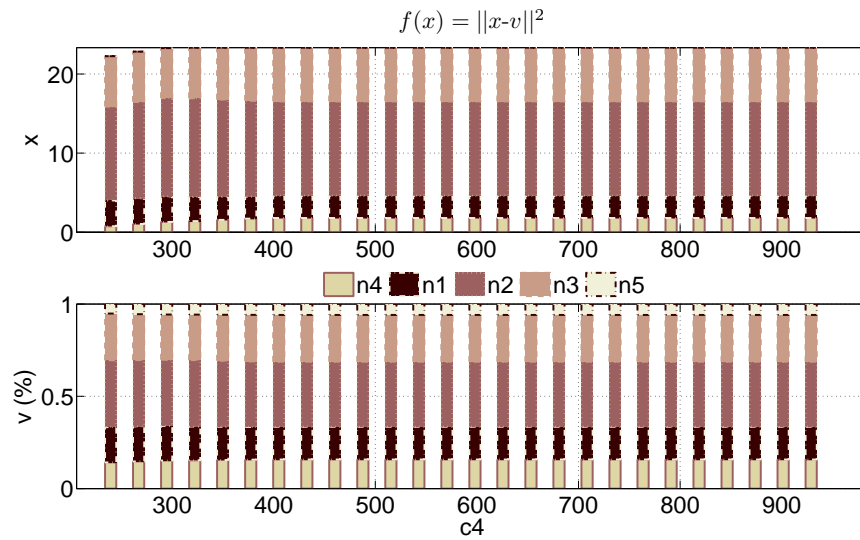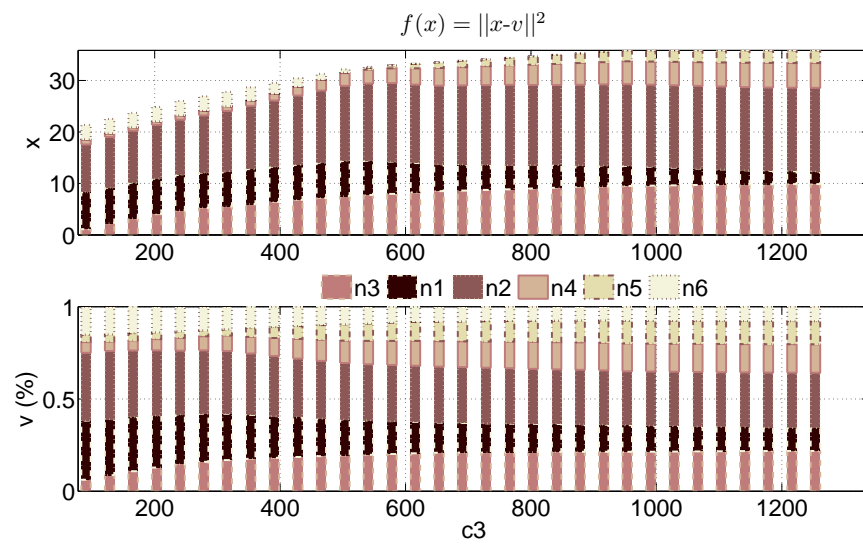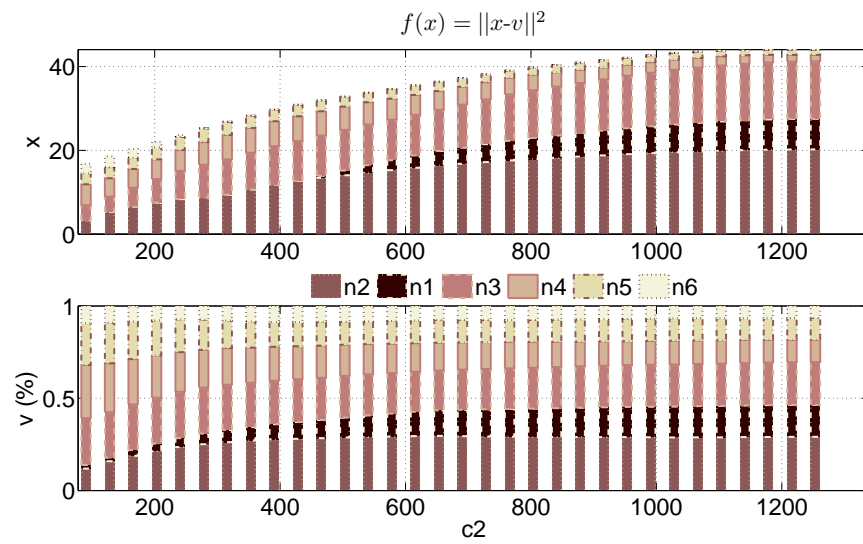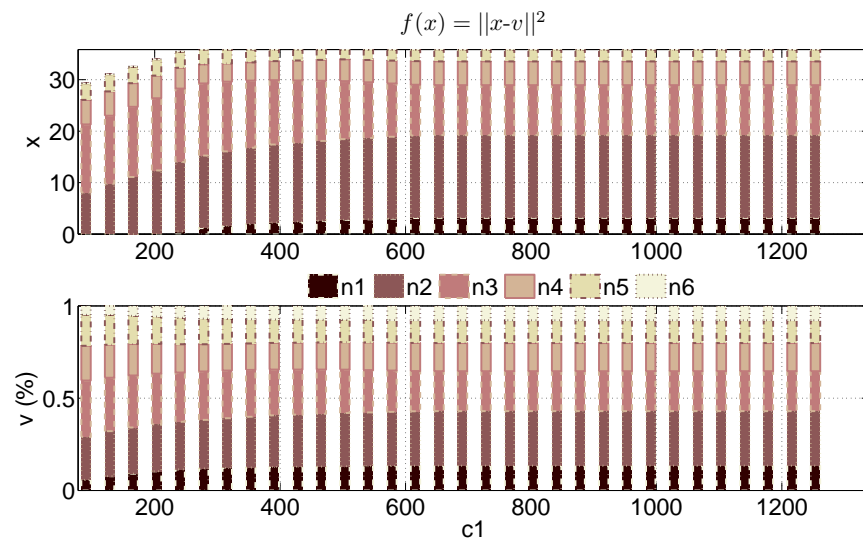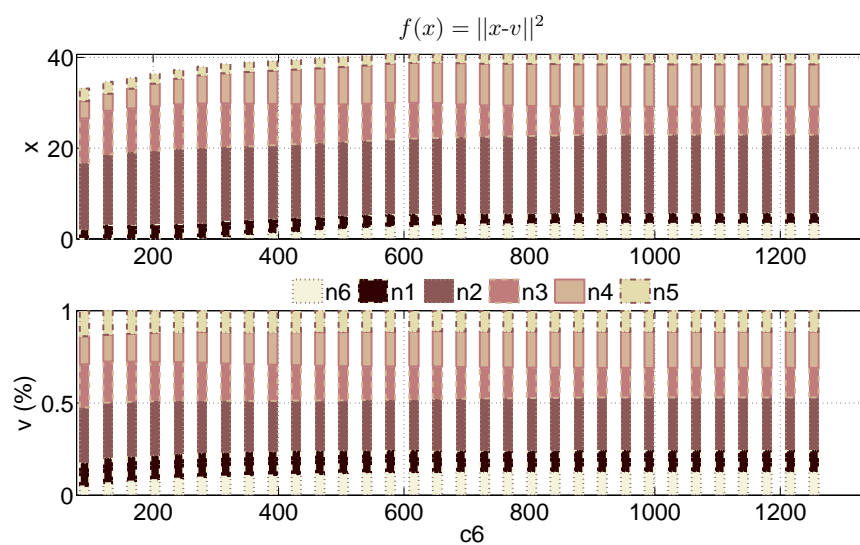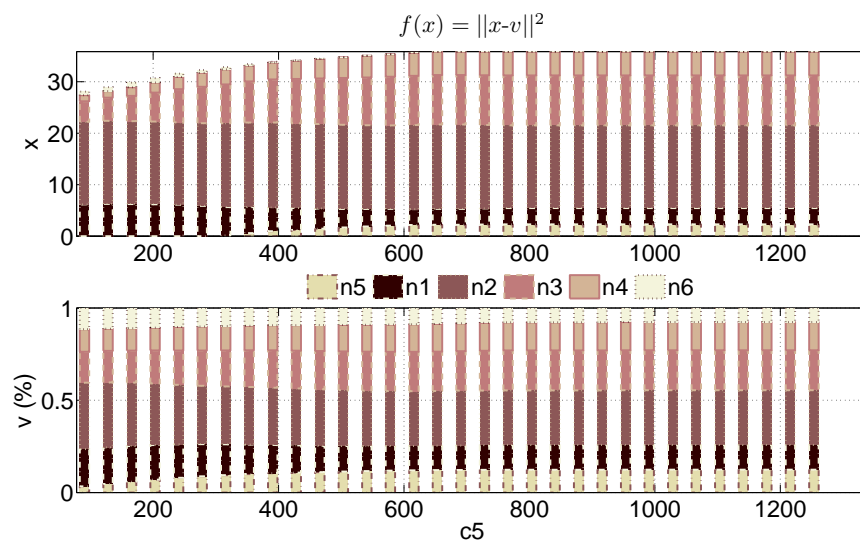
(a) n4.

(b) n5.

Figure B.3: Revenue sharing when increasing the capacity of one node at a time for Topology 4 nodes 4 and 5.

(a) n1.

(b) n2.

(c) n3.

Figure B.4: Revenue sharing when increasing the capacity of one node at a time for Topology 5, nodes 1 to 3.

(a) n4.

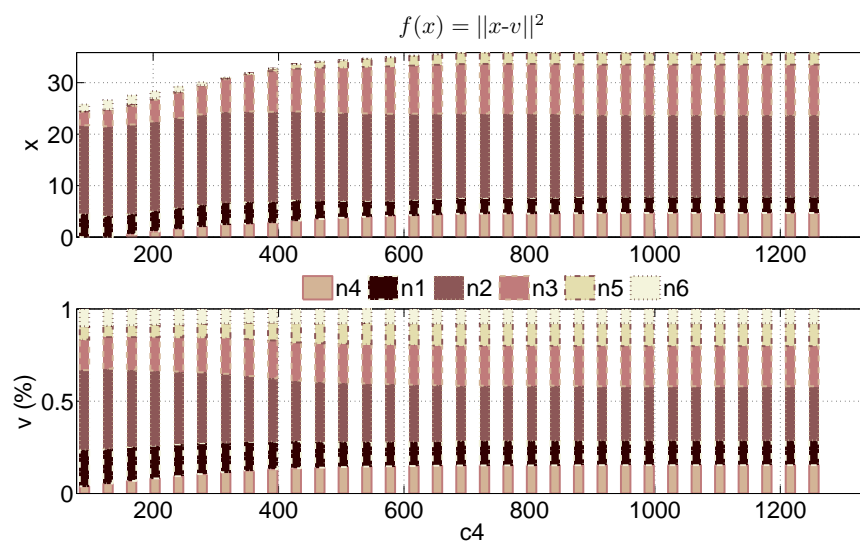
(b) n5.

(c) n6.

Figure B.5: Revenue sharing when increasing the capacity of one node at a time for Topology 5, nodes 4 to 6.
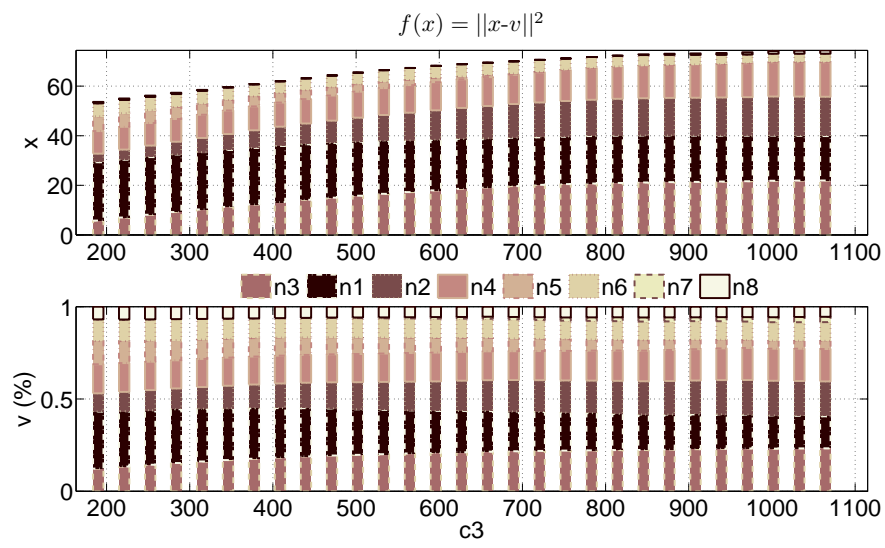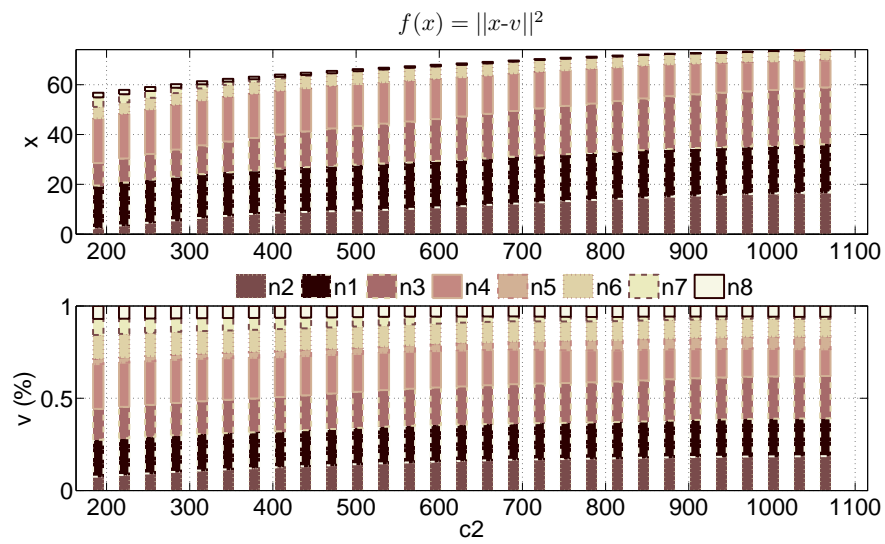
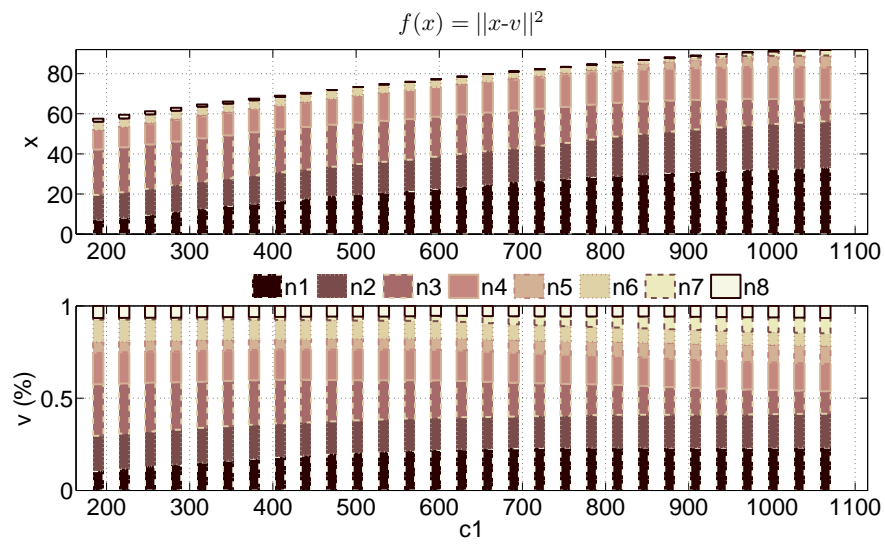**(a)** n1.

(b) n2

(c) n3

Figure B.6: Revenue sharing when increasing the capacity of one node at a time for Topology 6, nodes 1 to 3.
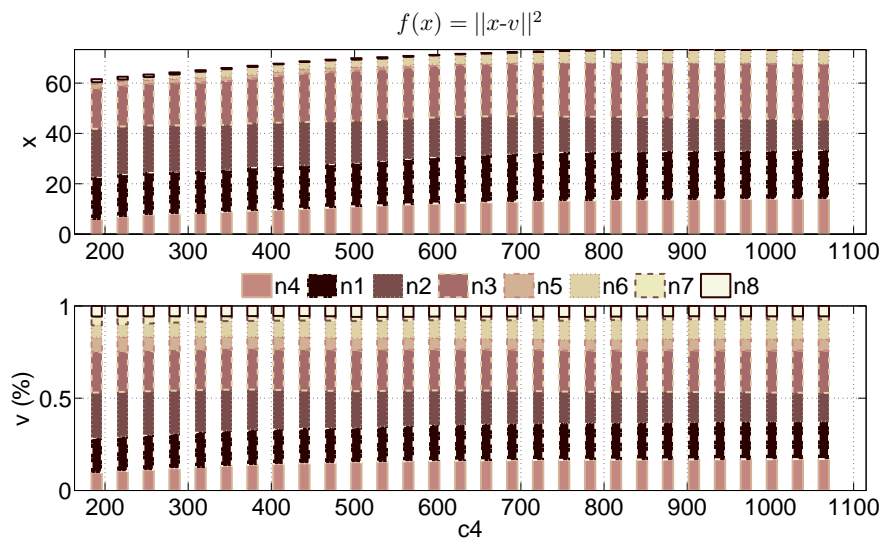
(a) n4

(b) n5

(c) n6

Figure B.7: Revenue sharing when increasing the capacity of one node at a time for Topology 6, nodes 4 to 6.

(a) n7

(b) n8

Figure B.8: Revenue sharing when increasing the capacity of one node at a time for Topology 6, nodes 7 and 8.

B.2   Multiperiod Simulations
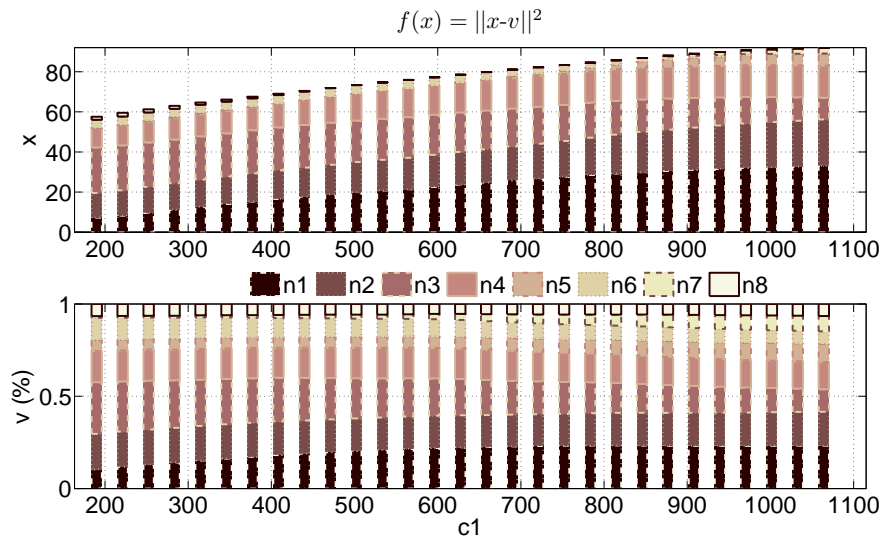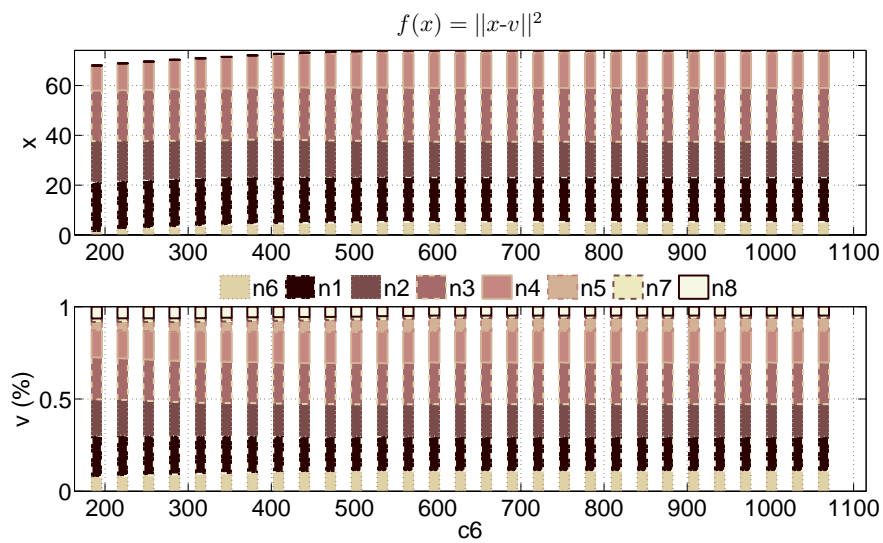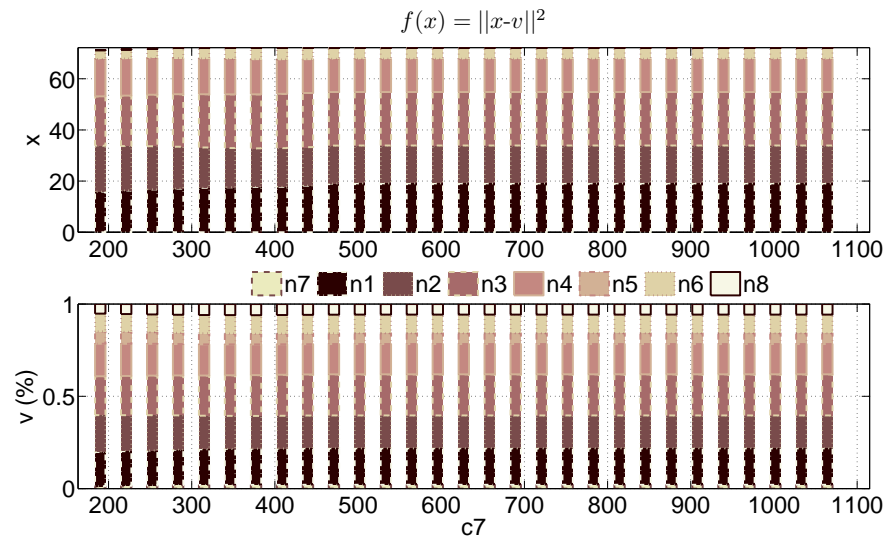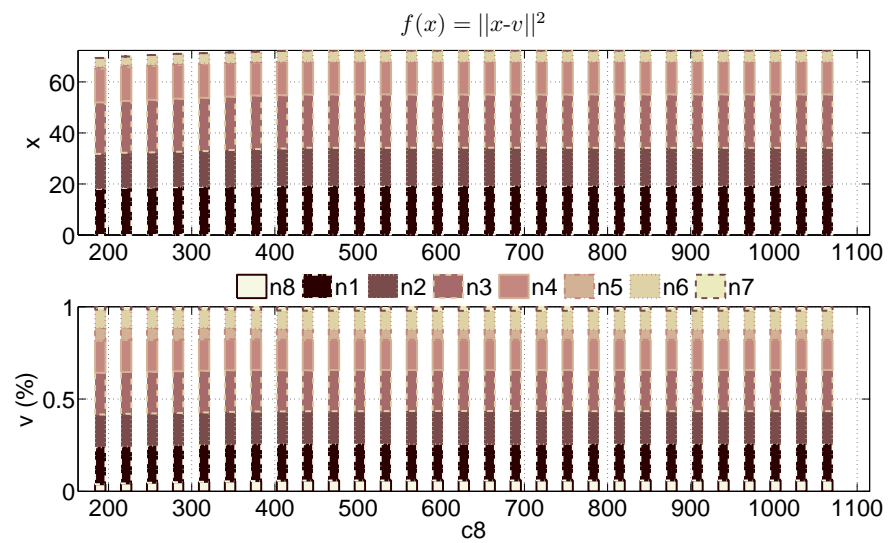
In Chapter 4 we have presented two different approaches to work with the statistics of the revenue in a multiperiod scenario. The objective of these simulations is to study the behaviour of both approaches, and to what extent they provide different revenue shares. We shall use the same topologies as in the previous section, which are shown in Fig. B.1. Revenue sharing was computed using the proposed method and setting the objective function to $f(x) = ||x - v||^2$. Before each revenue share, service selling phase was run 30 times. Fig. B.9, Fig. B.10 and Fig. B.11 show the results obtained for Topologies 4, 5 and 6 respectively. In all cases, the differences between the shares for each node according to Approach 1 and Approach 2 do not vary significantly.
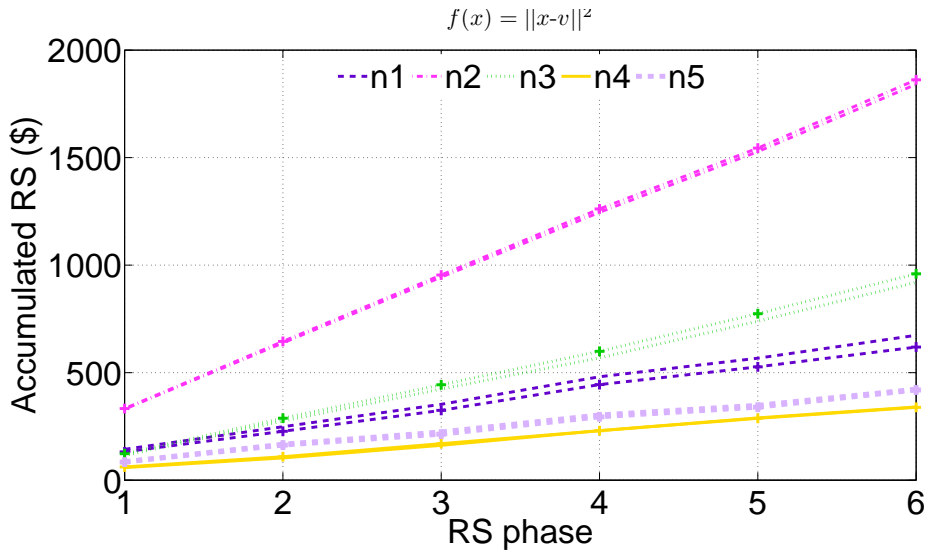

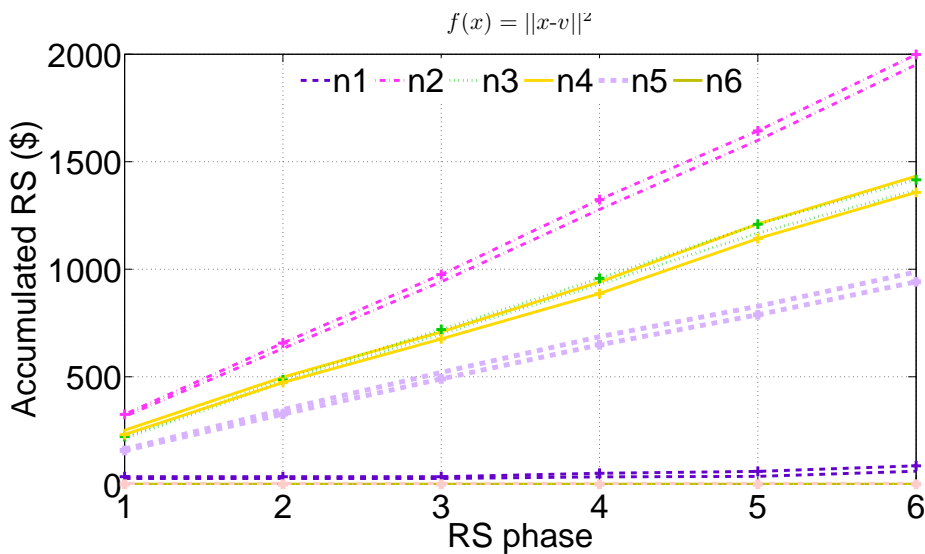**Figure B.9:** RS using Approach 1 (-) and 2 (+), Topology 4.

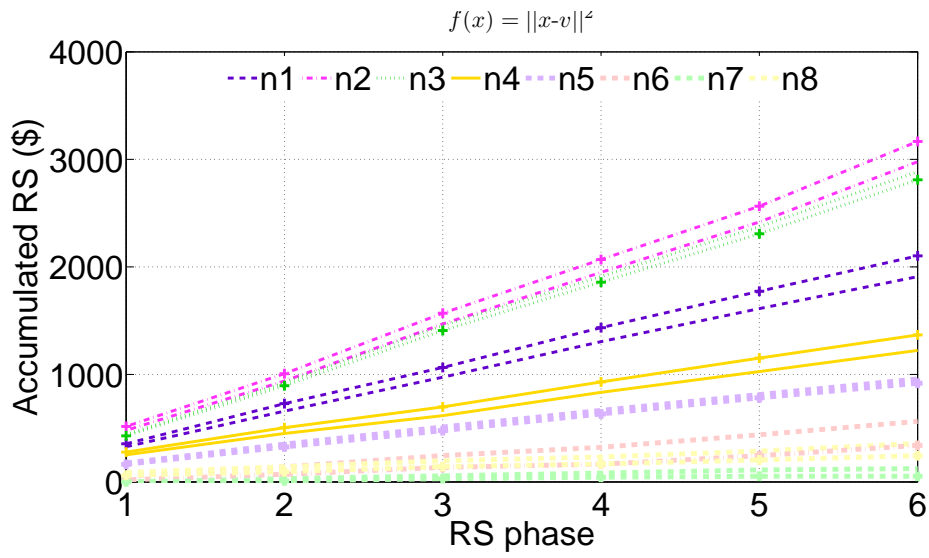**Figure B.10:** RS using Approach 1 (-) and 2 (+), Topology 5.

Figure B.11: RS using Approach 1 (-) and 2 (+), Topology 6.

Appendix C

The Proposed Pricing Scheme: Proofs, Simulations, Validation and Application

In this Appendix we present some additional calculus and simulations related to the proposed pricing scheme applied to the asymmetric case, in Chapter 6.

C.1 Bidding Strategies for Two Bidders and Two Asymmetric Services.

We present the details of the computation of the best bidding strategy for the case stated by Preposition (6.1). We recall that the case with one service, two bidders whose valuations are drawn from two uniform distributions with different supports can be derived in an analogous way to the case we shall show.

Assume there are two buyers whose valuations are drawn from the same uniform distribution and who bid to buy each a different service. Thus, each of them might assume a different probability of failure of the service to buy. Let us call $\tilde{\theta}_{i,s}$ the probability of failure assumed by buyer $i = 1, 2$ for services $1, 2$ and q_s the percentage of reimbursement announced for each service $s = 1, 2$.

Let us define $\phi_i(b) = \beta_i^{-1}(b)$. The probability that bidder 1 wins the auction if he submits a bid b is:

$$Pwin(b_1 = b; b_2) = \mathbb{P}(b > b_2) = 1 - \mathbb{P}(b_2 > b) \tag{C.1.1}$$
$$= 1 - \mathbb{P}(\phi_2(b_2) > \phi_2(b)) = 1 - \mathbb{P}(X > \phi_2(b)) = F(\phi_2(b)),$$

and analogously for buyer 2. Thus we can write the expected payoff for buyer $i = \{1, 2\}$ as:

$$P_{i,s}(b, x) = F(\phi_j(b))(x(1 - \tilde{\theta}_{i,s}) - b(1 - q_s\tilde{\theta}_{i,s})) \, j \neq i \tag{C.1.2}$$

The derivative of the expected payoff with respect to b is:

$$\frac{\partial P_{i,s}(b, x)}{\partial b} = f(\phi_j(b))\phi_j'(b)(x(1 - \tilde{\theta}_{i,s}) - b(1 - q_s\tilde{\theta}_{i,s})) - F(\phi_j(b))(1 - q_s\tilde{\theta}_{i,s}) \, j \neq i \tag{C.1.3}$$

Thus, imposing the first order condition for both buyers we obtain the following system of differential equations.

$$\phi_j'(b) = \frac{F(\phi_j(b))(1 - q_s\tilde{\theta}_{i,s})}{f(\phi_j(b))(\phi_i(b)(1 - \tilde{\theta}_{i,s}) - b(1 - q_s\tilde{\theta}_{i,s}))} \, j \neq i \tag{C.1.4}$$

We shall now assume that valuations are uniformly distributed on $[0, x^{max}]$. Thus, the system expressed by Equation (C.1.4) becomes:

$$\phi'_j(b) = \frac{\phi_j(b)(1 - q_s \tilde{\theta}_{i,s})}{\phi_j(b)(\phi_i(b)(1 - \tilde{\theta}_{i,s}) - b(1 - q_s \tilde{\theta}_{i,s}))} \ j \neq i \tag{C.1.5}$$

Let us define $\alpha_i = \frac{1 - \tilde{\theta}_{i,s}}{1 - q_s \tilde{\theta}_{i,s}}$ for $(i, s) = (1, 1), (2, 2)$. We can rewrite Equation (C.1.5) as

$$\phi'_1(b)\phi_2(b) - \phi'_1(b)\frac{b}{\alpha_2} = \frac{\phi_1(b)}{\alpha_2} \tag{C.1.6}$$

$$\phi'_2(b)\phi_1(b) - \phi'_2(b)\frac{b}{\alpha_1} = \frac{\phi_2(b)}{\alpha_1} \tag{C.1.7}$$

We derive a border condition and decouple the system given by Equation (C.1.6) and Equation (C.1.7). Indeed, summing up those equations we obtain

$$\phi'_1(b)\phi_2(b) + \phi'_2(b)\phi_1(b) = \frac{b\phi'_1(b)}{\alpha_2} + \frac{\phi_1(b)}{\alpha_2} + \frac{b\phi'_2(b)}{\alpha_1} + \frac{\phi_2(b)}{\alpha_1}, \tag{C.1.8}$$

which can be rewritten as:

$$\frac{\partial}{\partial b}\left(\phi_1(b)\phi_2(b)\right) = \frac{\partial}{\partial b}\left[b\left(\frac{\phi_2(b)}{\alpha_1} + \frac{\phi_1(b)}{\alpha_2}\right)\right]. \tag{C.1.9}$$

Integrating Equation (C.1.9) we find that $\phi_1(b)\phi_2(b) = b(\frac{\phi_2(b)}{\alpha_1} + \frac{\phi_1(b)}{\alpha_2})$, which is in particular true at the border of the domain $x = x^{max}$. Please note that the integration constant is zero because $\phi_i(0) = 0$. In addition, we know that both bidding functions $\beta_1(x^{max}) = \beta_2(x^{max})$, since if not one bidder could do better by bidding something smaller (see Lemma (6.3) in Section 6.3.2.B). Let us call this common value b^{max}. Thus we have that:

$$(x^{max})^2 = b^{max}\left(\frac{x^{max}}{\alpha_1} + \frac{x^{max}}{\alpha_2}\right), \tag{C.1.10}$$

from where we can deduce that:

$$b^{max} = \frac{(x^{max})^2}{\frac{x^{max}}{\alpha_2} + \frac{x^{max}}{\alpha_1}}. \tag{C.1.11}$$

From Equation (C.1.9) we can as well conclude that

$$\phi_1(b) = \frac{\frac{b\phi_2(b)}{\alpha_1}}{\phi_2(b) - \frac{b}{\alpha_2}} \tag{C.1.12}$$

Injecting Equation (C.1.12) into Equation (C.1.6) and Equation (C.1.7) we obtain the following decoupled system of first order differential equations:

$$\phi'_i(b) = \phi_i(b)\left(\phi_i(b) - \frac{b}{\alpha_i}\right)\frac{\alpha_i}{b^2}, \ i = 1, 2 \tag{C.1.13}$$

The solution to Equation (C.1.13) can be easily found to be

$$\phi_i(b) = \frac{2b}{(1 + k_i b^2)\alpha_i}, \ i = 1, 2, \tag{C.1.14}$$

where k_i is a constant of integration.

In order to determine k_i we shall use the border condition given by Equation (C.1.11), obtaining:

$$k_i = \frac{\alpha_j^2 - \alpha_i^2}{(x^{max}\alpha_i\alpha_j)^2}, \, i \neq j. \tag{C.1.15}$$

Finally, inverting Equation (C.1.14) we obtain:

$$\beta_i(x) = \frac{1 - \sqrt{1 - \alpha_i^2 k_i x}}{\alpha_i k_i x}, \, i = 1, 2, \tag{C.1.16}$$

where k_i is given by Equation (C.1.15).

We have found β_i at which the first order derivative of the payoff is null. We still need to check that indeed it is an equilibrium, which can be readily done.

C.2   Additional Case Study

In this section we shall evaluate the proposed pricing scheme in another asymmetric scenario, enriching the two scenarios shown in Chapter 6. We now consider a scenario with 3 classes of buyers, 2 of them attach to the one service on sale uniformly distributed valuations, each with different supports. The buyers belonging to the third class attach to the service exponentially distributed valuations. This represents a case, for instance, where there are some buyers that value the service more than others. In particular we consider that buyers belonging to the first class attach to the service a valuation on the interval $[0, 1]$, valuations of buyers belonging to the second class are independently distributed on the interval $[0, 2]$, and those of buyers belonging to the third class are exponentially distributed with parameter $\lambda = 1$. Which could be interpreted as that clients in the latter class usually attach a valuation close to 1 but less frequently they can attach any greater value. We assume that each class has the same number of buyers. The service on sale fails 20% of the time, that is $\theta = 0.2$.

We proceed as explained in Chapter 6. We define a grid for the values of q and for the values of $\tilde{\theta}_i$. For each value of q in that grid, buyers' expected payoffs over the discretized space of $\tilde{\theta}_i$ are computed. With this values, for each value of q we determine the Nash equilibrium, that is the values of $\tilde{\theta}_i$ that maximize each bidders' payoff. Please note that since the equilibrium bidding strategy among buyers belonging to the same class is the same, we only need three different values of $\tilde{\theta}_i$.

Results are shown in Table C.1. This table should be read as follows. For each column $\bar{P}_j^{MC}$ rows with the same color on represent settings where the actions of buyers $i \neq j$ are fixed (the reader can check this against the column in the left). In order to find the Nash Equilibrium at each column $\bar{P}_j^{MC}$ for each color, the setting that maximizes j's payoff is highlighted in bold. Those settings that are highlighted in bold at all columns constitute a Nash equilibrium. Finally, we compare the expected seller's revenue at these settings (red cells) in order to chose the value of q that renders the greatest expected revenue to the seller. In this scenario, once again the optimum is setting $q = 1$.

Finally, it is worth noting that the results obtained in this scenario are coherent with the analytical results presented in Chapter 5. That is, that when buyers are uncertain about the performance of the service to buy, in order to avoid a decrease on the expected seller's revenue, the seller should propose a percentage of reimbursement equal to 100%.

$(\tilde\theta_1,\tilde\theta_2,\tilde\theta_3)$	$\bar P_1^{MC}$	$\bar P_2^{MC}$	$\bar P_3^{MC}$	$\bar R^{MC}$	$\bar P_1^{MC}$	$\bar P_2^{MC}$	$\bar P_3^{MC}$	$\bar R^{MC}$
(0, 0, 0.2)	**0.01**	0.05	0.3	0.39	**0.01**	0.07	0.32	0.35
(0, 0, 0.9)	**0.01**	0.05	**0.54**	0.11	**0.01**	**0.07**	**0.33**	**0.34**
(0, 0, 0)	0.0	0.05	0.26	0.42	**0.01**	**0.07**	0.31	0.35
(0, 0.2, 0.2)	**0.01**	0.07	0.3	0.36	**0.01**	**0.08**	0.32	0.35
(0, 0.2, 0.9)	**0.01**	0.06	**0.54**	0.11	**0.01**	0.07	**0.33**	**0.35**
(0, 0.2, 0)	0.0	0.07	0.27	0.41	**0.01**	0.07	0.32	0.33
(0, 0.9, 0.2)	**0.01**	**0.22**	0.38	0.22	**0.01**	0.07	0.32	0.34
(0, 0.9, 0.9)	**0.01**	**0.21**	**0.56**	**0.07**	**0.01**	**0.07**	0.32	0.34
(0, 0.9, 0)	**0.01**	**0.23**	0.37	0.23	**0.01**	**0.07**	**0.33**	**0.34**
(0.2, 0, 0.2)	**0.01**	0.06	0.3	0.36	**0.01**	**0.08**	0.31	0.35
(0.2, 0, 0.9)	**0.01**	0.05	**0.55**	0.1	**0.01**	**0.07**	**0.33**	**0.33**
(0.2, 0, 0)	**0.01**	0.06	0.27	0.39	**0.01**	**0.07**	**0.33**	**0.35**
(0.2, 0.2, 0.2)	**0.01**	0.07	0.32	0.33	**0.01**	0.07	**0.33**	0.33
(0.2, 0.2, 0.9)	**0.01**	0.06	**0.56**	0.1	**0.01**	**0.07**	0.31	0.35
(0.2, 0.2, 0)	**0.01**	0.07	0.3	0.36	**0.01**	**0.07**	**0.33**	**0.34**
(0.2, 0.9, 0.2)	**0.01**	**0.22**	**0.41**	**0.18**	**0.01**	0.07	0.32	0.35
(0.2, 0.9, 0.9)	**0.01**	**0.21**	0.18	0.11	**0.01**	**0.07**	0.32	0.34
(0.2, 0.9, 0)	**0.01**	**0.22**	**0.41**	**0.19**	**0.01**	**0.07**	0.32	0.35
(0.9, 0, 0.2)	**0.01**	0.1	0.34	0.35	**0.01**	**0.08**	**0.33**	**0.34**
(0.9, 0, 0.9)	**0.01**	0.04	**0.58**	0.05	**0.01**	0.06	**0.33**	0.35
(0.9, 0, 0)	**0.01**	0.1	0.29	0.39	**0.01**	0.06	**0.33**	0.35
(0.9, 0.2, 0.2)	**0.01**	0.12	0.35	0.31	**0.01**	0.07	**0.32**	0.34
(0.9, 0.2, 0.9)	**0.01**	0.05	**0.57**	0.05	**0.01**	**0.07**	0.32	**0.34**
(0.9, 0.2, 0)	**0.01**	0.12	0.31	0.35	**0.01**	**0.07**	0.32	0.35
(0.9, 0.9, 0.2)	**0.01**	**0.21**	0.46	0.07	**0.01**	0.07	**0.33**	0.35
(0.9, 0.9, 0.9)	**0.01**	**0.2**	**0.59**	**0.04**	**0.01**	0.07	0.32	0.34
(0.9, 0.9, 0)	**0.01**	**0.21**	0.44	0.07	**0.01**	**0.07**	0.32	0.35
(a) Settings	**(b)** $q = 0$			**(c)** q=0	**(d)** $q = 1$			**(e)** q=1

Table C.1: Information structure for the game in Case study 2. Actions $(\tilde\theta_1, \tilde\theta_2, \tilde\theta_3)$ and q, approximate bidders' expected payoff ($\bar P_i^{MC}$, $i = 1\ldots 3$) and approximate seller's expected revenue $\bar R^{MC}$.

Bibliography

[1] BRITE: Boston Representative Internet Topology Generator [Online] Available: http://www.cs.bu.edu.

[2] ETICS Deliverable D3.5 "Final Business Models Analysis". Technical report.

[3] ETICS Deliverable D4.4 "Final ETICS Architecture and Functional Entities High Level Design". Technical report.

[4] ETICS: Economics and Technologies for Inter-carrier Services. European FP7 Research Project [Online] Available: http://www.ict-etics.eu.

[5] Fiber-to-the-Home (FTTH) Council Americas [Online] Available: http://www.ftthcouncil.org.

[6] IBM ILOG CPLEX Optimizer [Online] Available: http://www-01.ibm.com/software/integration/optimization/cplex-optimizer/.

[7] Multi-Parametric Toolbox (MPT) A tool (not only) for multi-parametric optimization [Online] Available: http://control.ee.ethz.ch/~mpt/.

[8] Path Computation Element (PCE) IETF Working Group [Online] Available: http://www.ietf.org/html.charters/pce-charter.html.

[9] PYTHRAN An authomatic PYTHON to C++ converter [Online] Available: https://github.com/serge-sans-paille/pythran .

[10] RFC 2475: An Architecture for Differentiated Services [Online] Available: http://www.ietf.org/rfc/rfc2475.txt.

[11] RFC 2702: Requirements for Traffic Engineering Over MPLS [Online] Available: http://www.ietf.org/rfc/rfc2702.txt .

[12] RFC 3209: RSVP-TE: Extensions to RSVP for LSP Tunnels [Online] Available: http://www.rfc-editor.org/rfc/rfc3209.txt.

[13] RFC 3954: Cisco Systems NetFlow Services Export Version 9 [Online] Available: http://www.rfc-editor.org/rfc/rfc3954.txt.

[14] RFC 4726: A Framework for Inter-Domain Multiprotocol Label Switch [Online] Available: http://tools.ietf.org/rfc/rfc4726.txt.

[15] RFC 5151: Inter-Domain MPLS and GMPLS Traffic Engineering - Resource Reservation Protocol-Traffic Engineering (RSVP-TE) Extensions [Online] Available: http://www.rfc-editor.org/rfc/rfc5151.txt.

[16] RFC 5441: A Backward-Recursive PCE-Based Computation (BRPC) Procedure to Compute Shortest Constrained Inter-Domain Traffic Engineering Label Switched Paths [Online] Available: http://tools.ietf.org/rfc/rfc5441.txt.

[17] RFC 6805: The Application of the Path Computation Element Architecture to the Determination of a Sequence of Domains in MPLS and GMPLS [Online] Available: http://tools.ietf.org/rfc/rfc6805.txt.

[18] RFC 6812: Cisco Service-Level Assurance Protocol [Online] Available: http://www.rfc-editor.org/rfc/rfc6812.txt.

[19] The ATM Forum, Private Network-Network Interface. Specification version 1.0, March 1996.

[20] The GÉANT network [Online] Available: http://www.geant.net.

[21] Cisco Systems: Entering the Zettabyte Era. Technical report, May 2011.

[22] Cisco Visual Networking Index, The Zettabyte Era. Technical report, May 2012.

[23] G. A. Akerlof. The Market for Lemons: Quality Uncertainty and the Market Mechanism. *The Quarterly Journal of Economics*, 84(3):488–500, 1970.

[24] A. Al Daoud, T. Alpcan, S. Agarwal, and M. Alanyali. A Stackelberg game for pricing uplink power in wide-band cognitive radio networks. In *47th IEEE Conference on Decision and Control*, pages 1422–1427. IEEE, December 2008.

[25] E. Altman, P. Bernhard, S. Caron, G. Kesidis, J. Rojas-Mora, and S. Wong. A model of network neutrality with usage-based prices. *Telecommunication Systems*, pages 1–9, June 2011.

[26] E. Altman, R. Marquez, R. El-Azouzi, D. Ros, and B. Tuffin. Stackelberg approach for pricing differentiated services. In *Proceedings of the 2nd international conference on Performance evaluation methodologies and tools*, ValueTools '07.

[27] I. Amigo, P. Belzarena, F. Larroca, and S. Vaton. Network Bandwidth allocation with end-to-end QoS constraints and Revenue Sharing in Multidomain Federations. In *ICQT 2011*, pages 50–62, Berlin, Heidelberg, 2011. Springer-Verlag.

[28] I. Amigo, P. Belzarena, and S. Vaton. A Pricing Scheme for QoS in Overlay Networks Based on First-Price Auctions and Reimbursement, Submitted. *Submitted*.

[29] I. Amigo, P. Belzarena, and S. Vaton. On the problem of revenue sharing in multi-domain federations. In *Networking 2012, Prague, Czech Republic*, may 2012.

[30] I. Amigo, S. Vaton, T. Chonavel, and F. Larroca. Maximum delay computation for interdomain path selection. *International journal of network management*, 2011.

[31] I. Amigo, S. Vaton, T. Chonavel, and F. Larroca. Maximum delay computation under traffic matrix uncertainty and its application to interdomain path selection. In *INOC 2011: Intermational Network Optimization Conference*, volume 6701, pages 182 – 195, 2011.

[32] O. Armantier, J.-P. Florens, and J.-F. Richard. Approximation of Nash equilibria in Bayesian games. *J. Appl. Econ.*, 23(7):965–981, 2008.

[33] K. J. Arrow, L. Hurwicz, and H. Uzawa. *Studies in Linear and Non-Linear Programming*. Stanford University Press, 1958.

[34] R. J. Aumann and L. S. Shapley. *Values of non-atomic games*. Princeton University Press, 1974.

[35] P. Bajari. *The First Price Auctions with Asymmetric Bidders: Theory and Applications*. PhD thesis, University of Minnesota, 1997.

[36] P. Bajari. Comparing competition and collusion: a numerical approach. *Economic Theory*, 18(1):187–205, July 2001.

[37] D. Barth, L. Echabbi, and C. Hamlaoui. Stock provisioning strategies for QoS at the interdomain level. *Telecommunication Systems*, pages 1–10, June 2011.

[38] R. C. Bartlett and S. D. Collins. *Aristotle's Nicomachean ethics*. University of Chicago Press, 2012.

[39] T. Basar and R. Srikant. A Stackelberg Network Game with a Large Number of Followers. *J. Optimization Theory and Applications*, 115(3):479–490, December 2002.

[40] A. Belghith, S. Said, B. Cousin, and S. Lahoud. Proactive and reactive collaboration schemes for multi-domain networks monitoring. In *Computing, Communications and Applications Conference*, pages 150–157, January 2012.

[41] P. Belzarena, A. Ferragut, and F. Paganini. Bandwidth Allocation via Distributed Auctions with Time Reservations. In *IEEE INFOCOM, Procedings. Rio de Janeiro, Brazil*, 2009.

[42] W. Ben-Ameur and H. Kerivin. Routing of Uncertain Traffic Demands. *Optimization and Engineering*, 6(3):283–313, September 2005.

[43] G. Bertrand and G. Texier. Ad-hoc Recursive PCE Based Inter-domain Path Computation (ARPC) Methods. In *Fifth International Working Conference on Performance Modelling and Evaluation of Heterogeneous Networks (HET-NETs)*, February 2008.

[44] D. Bertsekas and J. Tsitsiklis. *Parallel and Distributed Computation: Numerical Methods.* Prentice-Hall, 1989.

[45] S. Bessler and P. Reichl. A Network Provisioning Scheme Based on Decentralized Bandwidth Auctions. In S. Raghavan and G. Anandalingam, editors, *Telecommunications Planning: Innovations in Pricing, Network Design and Management*, volume 33 of *Operations Research/Computer Science Interfaces Series*, pages 89–100. Springer US, 2006.

[46] H. K. Bhargava and S. Sundaresan. Contingent Bids in Auctions: Availability, Commitment and Pricing of Computing as Utility. In *System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on*, page 10. IEEE, January 2004.

[47] A. Bogomolnaia, R. Holzman, and H. Moulin. Sharing the Cost of a Capacity Network. *Mathematics of Operations Research*, 35(1):173–192, February 2010.

[48] M. Bowling and M. Veloso. An Analysis of Stochastic Game Theory for Multiagent Reinforcement Learning. In *Computer Science Department, Carnegie Mellon University*, 2000.

[49] A. Brondsted. *An Introduction to Convex Polytopes.* Springer, December 1982.

[50] P. Casas, S. Vaton, L. Fillatre, and I. Nikiforov. Optimal volume anomaly detection and isolation in large-scale IP networks using coarse-grained measurements. *Computer Networks*, 54(11):1750–1766, August 2010.

[51] P. Casas, S. Vaton, L. Fillatre, and T. Chonavel. Efficient methods for traffic matrix modeling and on-line estimation in large-scale IP networks. In *21st International Teletraffic Congress (ITC 21)*, pages 1–8, September 2009.

[52] J. Corral, M. Ouarraou, G. Texier, and L. Toutain. Saturne: une architecture pour mesurer les performances des classes de service IP. *Technique et science informatiques (TSI) - ISSN 0752-4072*, 24(4):423–448, 2005.

[53] C. Courcoubetis and R. Weber. *Pricing and Communications Networks.* John Wiley & Sons, Ltd, 2003.

[54] L. De Vito, S. Rapuano, and L. Tomaciello. One-Way Delay Measurement: State of the Art. *Instrumentation and Measurement, IEEE Transactions on*, 57(12):2742–2750, 2008.

[55] A. Dhamdhere and C. Dovrolis. Ten years in the evolution of the Internet ecosystem. In *ACM SIGCOMM 2008*, pages 183–196. ACM, 2008.

[56] M. Dramitinos, G. D. Stamoulis, and C. Courcoubetis. An Auction Mechanism for Allocating the Bandwidth of Networks to their Users. *Computer Networks*, 51:4979–4996, December 2007.

[57] N. G. Duffield, P. Goyal, A. Greenberg, P. Mishra, K. K. Ramakrishnan, and J. E. van der Merive. A flexible model for resource management in virtual private networks. In *SIGCOMM '99: Proceedings of the conference on Applications, technologies, architectures, and protocols for computer communication*, pages 95–108, New York, NY, USA, 1999. ACM.

[58] A. Fabrikant, C. Papadimitriou, and K. Talwar. The complexity of pure Nash equilibria. In *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, pages 604–612. ACM, 2004.

[59] D. Feijer and F. Paganini. Krasovskii's method in the stability of network control. In *American Control Conference, 2009. ACC '09.*, pages 3292–3297, june 2009.

[60] D. Feijer and F. Paganini. Stability of primal-dual gradient dynamics and applications to network optimization. *Automatica*, 46:1974–1981, December 2010.

[61] N. Feng, S. Mau, and N. B. Mandayam. Pricing and Power Control for Joint Network-centric and User-centric Radio Resource Management. *Communications, IEEE Transactions on*, 52(9):1547–1557, september 2004.

[62] G. Fibich and N. Gavish. Numerical simulations of asymmetric first-price auctions. *Games and Economic Behavior*, 73(2):479–495, November 2011.

[63] L. Fillatre, D. Marakov, and S. Vaton. Forecasting Seasonal Traffic Flows. In *Workshop on QoS and Traffic Control.*, December 2005.

[64] E. Friedman and H. Moulin. Three Methods to Share Joint Costs or Surplus. *Journal of Economic Theory*, 87(2):275–312, August 1999.

[65] W. R. Gayle and J. F. Richard. Numerical Solutions of Asymmetric, First-Price, Independent Private Values Auctions. *Computational Economics*, 32(3):245–278, October 2008.

[66] R. A. Guérin and A. Orda. QoS routing in networks with inaccurate information: theory and algorithms. *IEEE/ACM Trans. Netw.*, 7(3):350–364, 1999.

[67] L. He and J. Walrand. Pricing and revenue sharing strategies for Internet service providers. *Selected Areas in Communications, IEEE Journal on*, 24(5):942–951, May 2006.

[68] A. Hernandez and E. Magana. One-way Delay Measurement and Characterization. In *Proceedings of the Third International Conference on Networking and Services*, ICNS '07, pages 114–, Washington, DC, USA, 2007. IEEE Computer Society.

[69] R. Holenstein. Using Sampling To Compute Bayes-Nash Equilibrium In Auction Games [Online] Available: http://www.cs.ubc.ca/ romanh/courses/cpsc540. Technical report, 2005.

[70] C. Homburg, N. Koschate, and W. D. Hoyer. Do Satisfied Customers Really Pay More? A Study of the Relationship Between Customer Satisfaction and Willingness to Pay. *Journal of Marketing*, 69(2):84–96, Avril 2005.

[71] S. Hon-Snir, D. Monderer, and A. Sela. A learning approach to auctions. *Journal of Economic Theory*, 82(1):65–88, 1998.

[72] J. L. Hougaard. *An Introduction to Allocation Rules*. Springer, 2009.

[73] J. L. Hougaard and L. P. Østerdal. Monotonicity of social welfare optima. *Games and Economic Behavior*, 70(2):392–402, 2010.

[74] W. Htira, O. Dugeon, and M. Diaz. A mesh aggregation scheme for call admission control. In *Telecommunication Networks and Applications Conference, 2007. ATNAC 2007. Australasian*, pages 265–270, dec. 2007.

[75] M. Johansson and A. Gunnar. Data-driven traffic engineering: techniques, experiences and challenges. In *3rd International Conference on Broadband Communications, Networks and Systems, 2006. BROADNETS 2006*, pages 1–10, October 2006.

[76] C. N. Jones. *Polyhedral Tools for Control*. PhD thesis, Pembroke College, 2005.

[77] H. T. Karaoglu and M. Yuksel. Value flows: Inter-domain routing over contract links. pages 327–332, December 2010.

[78] H. Keiding. Maximizing selections from the core of a cooperative game. *Journal of Global Optimization*, 50:107–118, 2011.

[79] F. P. Kelly, A. K. Maulloo, and D. K. H. Tan. Rate Control for Communication Networks: Shadow Prices, Proportional Fairness and Stability. *The Journal of Operational Research Society*, 49(3):237–252, 1998.

[80] L. Khachiyan, E. Boros, K. Borys, K. Elbassioni, and V. Gurvich. Generating All Vertices of a Polyhedron Is Hard. *Discrete and Computational Geometry*, 39(1):174–190, March 2008.

[81] H. Khalil. *Nonlinear Systems*. Prentice Hall, New Jersey, 1996.

[82] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by Simulated Annealing. *Science*, 220(4598):671–680, May 1983.

[83] N. N. Krasovskii. *Stability of motion: applications of Lyapunov's second method to differential systems and equations with delay.* Stanford university press, 1963.

[84] V. Krishna. *Auction Theory*. Academic Press/Elsevier, 2009.

[85] K. Kultti and H. Vartiainen. Bargaining with many players: A limit result. *Economics Letters*, 101(3):249–252, 2008.

[86] A. A. Kurzhanskiy and P. Varaiya. Ellipsoidal Toolbox. Technical Report UCB/EECS-2006-46, EECS Department, University of California, Berkeley, May 2006 2011.

[87] F. Larroca and J. L. Rougier. Robust regression for minimum-delay load-balancing. In *Teletraffic Congress, 2009. ITC. 21st International*, pages 1–8, September 2009.

[88] A. A. Lazar and N. Semret. Design and Analysis of the Progressive Second Price Auction for Network Bandwidth Sharing. *Telecommunications Systems, Special Issue on Network Economics*, 2000.

[89] B. Lebrun. Existence of an equilibrium in first price auctions. *Economic Theory*, 7(3):421–443, 1996.

[90] B. Lebrun. First Price Auctions in the Asymmetric N Bidder Case. *International Economic Review*, 40(1):125–142, 1999.

[91] J. Levendovszky and C. Orosz. Developing Novel Statistical Bandwidths for Communication Networks with Incomplete Information. In *Experimental and Efficient Algorithms*, pages 614–617. 2005.

[92] H. Li and J. G. Riley. Auction choice. *International Journal of Industrial Organization*, 25(6):1269–1298, December 2007.

[93] S. Lima, P. Carvalho, and V. Freitas. Toward Scalable Management of Multiple Service Levels in IP Networks. pages 223–232, February 2007.

[94] X. Lin and N. B. Shroff. Utility Maximization for Communication Networks With Multipath Routing. *IEEE Transactions on Automatic Control*, 51(5):766–781, May 2006.

[95] J. Lygeros, K. Johansson, S. Simic, J. Zhang, and S. Sastry. Dynamical properties of hybrid automata. *Automatic Control, IEEE Transactions on*, 48(1):2–17, 2003.

[96] R. T. B. Ma, D. M. Chiu, J. C. S. Lui, V. Misra, and D. Rubenstein. Internet Economics: The Use of Shapley Value for ISP Settlement. *IEEE/ACM Transactions on Networking*, 18(3):775–787, June 2010.

[97] J. MacKie-Mason. A Smart Market for Resource Reservation in Multiple Quality of Service Information Networks. Technical report, 1997. University of Michigan, USA.

[98] P. Maillé and B. Tuffin. Pricing the Internet with Multibid Auctions. *IEEE/ACM Transactions on Networking*, 14:992–1004, October 2006.

[99] P. Maillé and B. Tuffin. Why VCG Auctions can Hardly be Applied to the Pricing of Interdomain and Ad hoc Networks. In *3rd EuroNGI conference, Trondheim, Norway*, pages 36 – 39, 2007.

[100] P. Maillé and B. Tuffin. A Progressive Second Price Mechanism with a Sanction Bid from Excluded Players. In B. Stiller, G. Carle, M. Karsten, and P. Reichl, editors, *Group Communications and Charges. Technology and Business Models*, volume 2816 of *Lecture Notes in Computer Science*, pages 332–341. Springer Berlin Heidelberg, 2003.

[101] O. L. Mangasarian. Nonlinear Programming Problems with Stochastic Objective Functions. *Management Science*, 10(2), 1964.

[102] D. P. Mann and J. P. Wissink. Money-Back Contracts with Double Moral Hazard. *The RAND Journal of Economics*, 19(2), 1988.

[103] R. C. Marshall, M. J. Meurer, J.-F. Richard, and W. Stromquist. Numerical Analysis of Asymmetric First Price Auctions. *Games and Economic Behavior*, 7(2):193–220, 1994.

[104] A. Mas-Colell, M. Whinston, and J. Green. *Microeconomic Theory*. Oxford student edition. Oxford University Press, 1995.

[105] X. Masip-Bruin, S. Sanchez-Lopez, J. Sole-Pareta, and J. Domingo-Pascual. QoS routing algorithms under inaccurate routing for bandwidth constrained applications. In *ICC '03*, pages 1743–1748. IEEE, 2003.

[106] E. Maskin and J. Riley. Asymmetric auctions. *The Review of Economic Studies*, 67(3):413–438, 2000.

[107] E. Maskin and J. Riley. Equilibrium in sealed high bid auctions. *The Review of Economic Studies*, 67(3):439–454, 2000.

[108] E. Maskin and J. Riley. Uniqueness of equilibrium in sealed high-bid auctions. *Games and Economic Behavior*, 45(2):395–409, 2003.

[109] P. R. Milgrom and R. J. Weber. A Theory of Auctions and Competitive Bidding. *Econometrica*, 50(5):1089–1122, 1982.

[110] M. Minoux. *Programmation mathématique: Théorie et algorithmes*. Tec & Doc Lavoisier, 2nd edition, December 2007.

[111] V. Misra, S. Ioannidis, A. Chaintreau, and L. Massoulié. Incentivizing peer-assisted services: a fluid shapley value approach. In *Proceedings of the ACM SIGMETRICS International conference on Measurement and modeling of computer systems*, SIGMETRICS '10, pages 215–226, New York, NY, USA, 2010. ACM.

[112] H. Moulin and F. Laigret. Equal-need sharing of a network under connectivity constraints. *Games and Economic Behavior*, 72(1):314–320, 2011.

[113] M. Mycek, S. Secci, M. Pioro, J.-L. Rougier, A. Tomaszewski, and A. Pattavina. Cooperative multi-provider routing optimization and income distribution. In *DRCN 2009*, pages 281–288, October 2009.

[114] N. Le Sauze et al. ETICS : QoS-enabled Interconnection for Future Internet Services. In *Future Network and Mobile Summit*, 2010.

[115] J. F. J. Nash. The Bargaining Problem. *Econometrica*, 18(2):pp. 155–162, 1950.

[116] A. Nedić and A. Ozdaglar. Approximate Primal Solutions and Rate Analysis for Dual Subgradient Methods. *SIAM J. on Optimization*, 19(4):1757–1780, February 2009.

[117] P. S. Neelakanta and R. C. Sardenberg. Consumer benefit versus price elasticity of demand: a nonlinear complex system model of pricing internet services on QoS-centric architecture. *NETNOMICS: Economic Research and Electronic Networking*, 12:31–60, 2011.

[118] S. Niccolini, M. Molina, F. Raspall, and S. Tartarelli. Design and implementation of a one way delay passive measurement system. pages 469–482.

[119] M. J. Osborne and A. Rubinstein. *A Course in Game Theory*. The MIT Press, first edition, July 1994.

[120] S. Park, E. H. Durfee, and W. P. Birmingham. An adaptive agent bidding strategy based on stochastic modeling. In *Proceedings of the third annual conference on Autonomous Agents*, AGENTS '99, pages 147–153, New York, NY, USA, 1999. ACM.

[121] C. Pelsser and O. Bonaventure. Path selection techniques to establish constrained interdomain MPLS LSPs. In *NETWORKING 2006*, pages 209–220. Springer, 2006.

[122] H. Pouyllau and G. Carofiglio. Inter-carrier SLA negotiation using Q-learning. *Telecommunication Systems*, pages 1–12, June 2011.

[123] H. Pouyllau and R. Douville. End-to-end QoS negotiation in network federations. In *IEEE/IFIP Network Operations and Management Symposium Workshops (NOMS Wksps)*, pages 173–176, 2010.

[124] H. Pouyllau, R. Douville, N. B. Djarallah, and N. Le Sauze. Economic and technical propositions for inter-domain services. volume 14, pages 185–202. Wiley Subscription Services, Inc., A Wiley Company, 2009.

[125] B. Quoitin, C. Pelsser, L. Swinnen, O. Bonaventure, and S. Uhlig. Interdomain traffic engineering with BGP. *IEEE Communications Magazine*, 41(5):122–128, May 2003.

[126] P. Reichl. From Charging for Quality of Service to Charging for Quality of Experience. *Annals of Telecommunications*, 65(3):189–199, April 2010.

[127] P. Reichl, S. Bessler, and B. Stiller. Second-Chance Auctions for Multimedia Session Pricing. In G. Ventre and R. Canonico, editors, *Interactive Multimedia on Next Generation Networks*, volume 2899 of *LNCS*, pages 306–318. Springer, 2003.

[128] P. Reichl, B. Tuffin, and R. Schatz. Logarithmic laws in service quality perception: where microeconomics meets psychophysics and quality of experience. *Telecommunication Systems*, pages 1–14, 2011.

[129] P. Reichl and S. Wrzaczek. Equilibrium market prices for multi-period auctions of internet resources. In *12th GI/ITG Conference on Measuring, Modelling and Evaluation of Computer and Communication Systems*, page 25, 2004.

[130] P. J. Reny. On the Existence of Pure and Mixed Strategy Nash Equilibria in Discontinuous Games. *Econometrica*, 67(5):1029–1056, 1999.

[131] RFC 4656 - A One-way Active Measurement Protocol (OWAMP) [Online] Available: http://tools.ietf.org/rfc/rfc4656.txt .

[132] J. G. Riley and W. F. Samuelson. Optimal Auctions. *American Economic Review*, 71(3):381–92, June 1981.

[133] A. Rubinstein. Perfect Equilibrium in a Bargaining Model. *Econometrica*, 50(1):pp. 97–109, 1982.

[134] R. Y. Rubinstein and D. P. Kroese. *Simulation and the Monte Carlo Method (Wiley Series in Probability and Statistics)*. 2 edition, 2007.

[135] M. Saad, A. Leon-garcia, and W. Yu. Rate Allocation under Network End-to-End Quality-of-Service Requirements. In *GLOBECOM 2006*.

[136] A. Sackl, P. Zwickl, and P. Reichl. From Quality of Experience to Willingness to Pay for Interconnection Service Quality. In *NETWORKING 2012 Workshops*, volume 7291 of *LNCS*, pages 89–96. Springer Berlin Heidelberg, 2012.

[137] E. Salhi, S. Lahoud, and B. Cousin. Global Vs. Per-domain monitoring of multi-domain networks. In *Proceedings of the 2011 IEEE 36th Conference on Local Computer Networks*, LCN '11, pages 391–398, Washington, DC, USA, 2011. IEEE Computer Society.

[138] L. Shapley. A value for n-person games. In H. Kuhn and A. Tucker, editors, *Contributions to the Theory of Games*, volume 28, pages 307–317, 1953.

[139] H. Shen and T. Basar. Differentiated Internet pricing using a hierarchical network game model. In *Proceedings of the American Control Conference*, volume 3, pages 2322–2327 vol.3. IEEE, June 2004.

[140] Y. Shoham and K. Leyton-Brown. *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*. Cambridge University Press, Cambridge, UK, 2009.

[141] R. Stanojevic, I. Castro, and S. Gorinsky. CIPT: using tuangou to reduce IP transit costs. In *CoNEXT 2011*, New York, NY, USA, 2011. ACM.

[142] R. Teixeira, N. Duffield, J. Rexford, and M. Roughan. Traffic matrix reloaded: Impact of routing changes. In *Passive and Active Network Measurement*, pages 251–264. Springer, 2005.

[143] H.-H. Tran, B. Tuffin, et al. A distributed auction-based algorithm to allocate bandwidth over paths. In *NET-COOP 2010-4th Workshop on Network Control and Optimization*, 2010.

[144] T. Trinh, L. Gyarmati, and G. Sallai. Understanding the impact of loyal user behaviour on Internet access pricing: a game-theoretic framework. *Telecommunication Systems*, 48:43–61, 2011.

[145] B. Tuffin, H. Le Cadre, and M. Bouhtou. Optimal Pricing Strategy with Compensation when QoS is not Satisfied. *Telecommunication Annals*, 65(3-4):163–170, 2010.

[146] Y. Vardi. Network Tomography: Estimating Source-Destination Traffic Intensities from Link Data. *Journal of the American Statistical Association*, 91(433):365–377, March 1996.

[147] V. Černý. Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm. *Journal of Optimization Theory and Applications*, 45(1):41–51, January 1985.

[148] D. Veit and C. Czernohous. Automated bidding strategy adaption using learning agents in many-to-many e-markets. In *Proceedings of the Workshop on Agent Mediated Electronic Commerce V, AMEC-V, Melbourne, Australia*, 2003.

[149] W. Vickrey. Counterspeculation, Auctions, and Competitive Sealed Tenders. *The Journal of Finance*, 16(1):8–37, 1961.

[150] H. von Stackelberg. *Market Structure and Equilibrium*. Springer, 1st edition, December 2010.

[151] H. Walid, D. Olivier, and D. Michel. An Aggregated Delay—Bandwidth Star Scheme for Admission Control. In *Personal, Indoor and Mobile Radio Communications, 2007. PIMRC 2007. IEEE 18th International Symposium on*, pages 1–8, sept. 2007.

[152] B. Wang, Y. Wu, and K. J. R. Liu. Game theory for cognitive radio networks: An overview. *Computer Networks*, 54(14):2537–2561, October 2010.

[153] D. Wang and Z. Li. On Traffic Pricing Model Based on Incentive Stackelberg Strategies for Communication Networks. In *Communications and Information Processing*, volume 288 of *Communications in Computer and Information Science*, pages 498–505. Springer Berlin Heidelberg, 2012.

[154] M. Yannuzzi, A. Fonte, X. Masip-Bruin, E. Monteiro, S. Sanchez-Lopez, and J. Domingo-Pascual. A self-adaptive QoS routing framework for multihomed stub autonomous systems. In *Proceedings of Eunice, IFIP. Madrid, Spain*, pages 241–247, July 2004.

[155] H. Young. Monotonic solutions of cooperative games. *International Journal of Game Theory*, 14:65–72, 1985.

[156] F. Zhang and P. Verma. Pricing multi-class network services using the Shapley Value. *NETNOMICS: Economic Research and Electronic Networking*, 12:61–75, 2011.

[157] R. Zhang and J. P. Vasseur. MPLS Inter-Autonomous System (AS) Traffic Engineering (TE) Requirements, November 2005.

[158] Y. Zhang. The Abilene Dataset [Online] Available: http://www.cs.utexas.edu.

[159] Y. Zhang, M. Roughan, C. Lund, and D. L. Donoho. Estimating point-to-point and point-to-multipoint traffic matrices: an information-theoretic approach. *IEEE/ACM Trans. Netw.*, 13:947–960, October 2005.