# Contents

# List of Figures

# List of Tables

# Publications

To date, the work presented in this thesis has produced the following publications:

- Beltran, F. and Roggendorf, M. (2007). Multiple equilibria in symmetric strategies for simultaneous auctions in next-generation bandwidth markets. *To appear in the Proceedings of the 1st Workshop for Games in Communication Networks (GameComm 2007), October 22, Nantes, France*

- Roggendorf, M. and Beltran, F. (2006c). A heuristic approach to revenue maximisation in a competitive bandwidth-on-demand wireless market. *Proceedings of the 1st International Workshop for Bandwidth-on-Demand (BOD), San Francisco, USA*

- Roggendorf, M. and Beltran, F. (2006b). Flow-based resource allocation in a multiple access wireless market using an auction. *Proceedings of the Second International Workshop on Incentive-Based Computing (IBC), July 4-7, Lisboa, Portugal.*

- Beltran, F. and Roggendorf, M. (2006a). A simulation-based approach to bidding strategies for network resources in competitive wireless networks. *Proceedings of the Joint International Conference on Measurement and Modeling of Computer Systems, June 26-30, 2006, Saint-Malo, France.*

- Roggendorf, M., Beltran, F., and Gutierrez, J. (2006). Architecture and implementation of an agent-based simulation tool for market-based pricing in next-generation wireless networks. *Proceedings of the TridentCOM 2006, 2nd International Conference on Testbeds and Research Infrastructures for the Development of Networks and Communities, March 1-3, Barcelona, Spain,* pages 282 288.

- Beltran, F. and Roggendorf, M. (2005). A simulation model for the dynamic allocation of network resources in a competitive wireless scenario. *Proceedings of the Second International Workshop on Mobility Aware Technologies and Applications, MATA*

*2005*. Montreal, Canada, October 17-19, 2005. Lecture Notes in Computer Science 3744, ISBN 3-540-29410-4. 54-64.

- Roggendorf, M. and Beltran, F. (2005). Innovative pricing and charging in next-generation wireless networks, *Short Paper in the Proceedings of the IRMA conference, San Diego, USA*

# Chapter 1
# Introduction

## ▶ 1.1 Introduction

The introduction of the Internet Protocol (IP) for packeting, global addressing and transportation of digital information in data communication networks has opened up a tremendously broad range of possibilities. The potential of IP-based technology has also been recognised by the wireless industry; traditional cellular telephony providers, as well as new entrants, are already operating IP-based networks to support high speed data transport services, as well as interactive multimedia applications which provide audio and video content as additional features. The next generation of wireless networks, which currently emerge from existing cellular network standards and wireless data communication networks, promises to be an all-IP ubiquitous network, capable of providing multiple service types with guaranteed Quality-of-Service (Berezdivin et al., 2002).

Wireless technology continues to advance at a fast pace pushing the efficiency of the scarce wireless frequency spectrum and reducing the packet delay by introducing sophisticated data flow mechanisms. Another area of current research is the development of suitable mechanisms to let users seamlessly roam between different wireless network technologies while providing guaranteed Quality-of-Service. Together with the advances in microelectronics it is expected to create mobile terminals which are capable of using different radio technologies concurrently and to automatically select the wireless network most appropriate for the current service request.

In contrast to the technical development, the principal economic models to market wireless services have not changed tremendously over the last few years. Despite the introduction of innovative data products such as IP-based TV, music downloads and push-based email applications, the business models used to promote such services are still based on long-term customer contracts, which bind the user to a specific network over an extended period. While the available wireless technology already supports dynamic network selection in IP-based networks, current business models do not cater for the implementation across provider boundaries.[1] The most widespread charging schemes currently used in wireless communication are flat pricing and usage-based pricing (Soldatos et al., 2005). While flat rate pricing charges a fixed amount independent from actual usage, usage-based charging is the predominant way of pricing in wireless networks due to the stringent limitations of wireless resources.

A development, which has the potential of disrupting the current charging practices prevailingly used in mobile networks, is the introduction of services that work independently from the underlying transport network. Unlike circuit-switched voice services, which are closely interlinked with the provisioning of the wireless network itself (and which still produce the largest share of revenues for mobile service providers), such new service types can also be used in a heterogeneous network environment without a fixed relationship to only one provider. Examples include Voice-over-IP and Video-On-Demand, which allow customers to use information and data services independently from the underlying wireless transport network. With this development ongoing, wireless transport providers may need to rethink their current revenue models and find new ways of pricing for network transport services than simply by flat rates or usage-based models.

In a world that is so much influenced by the impact of technological advances in wireless communications, the pricing of these services is expected to play an important role in future developments. Pricing of mobile voice and data services is, in the public, primarily perceived as a tool for maximising provider profits in a highly competitive environment. Secondary, many other functions of pricing can be discovered. One important function, which plays a central role in this thesis, is pricing for network control and signalling. By increasing prices for a communication service, a provider can reduce the demand, reduce congestion, and can ensure that the service is provided to the users who benefit most and are most willing to pay (Courcoubetis and Weber, 2003). Pricing in this respect can also be seen as a mechanism for communication between market participants. By changing a price for a communication service, providers can provide incentives to change user demand and to adapt to the new situation. Conversely, from the way users respond to prices, providers can learn about their preferences and their intended network use and can adapt price levels accordingly.

One aspect in wireless networks, which has not yet received a lot of attention, is the possibility of direct competition between network providers on shorter time-scales. In

[1] An exception is the so-called roaming in cellular networks, which allows users to make use of other wireless networks outside their home countries. With this approach the billing relationship is handled by the provider the user is subscribed to.

wired networks, the physical channel (usually the copper cable or fiber) is owned by one network operator, which can exclusively supply a customer with communication services. Through market liberalisation it has become possible for market entrants to access this channel to provide users a choice and to foster competition. However, provider switching is a long-term decision, which cannot be easily reversed and may be associated with high switching costs.

In contrast, radio access networks provide a direct interface to the customer with any suitable end device. Switching between networks of different providers can be realised in the order of seconds or minutes based on the signals received from these networks. In this respect, prices can play an influential role in the customer decision about which network to join.

Supported by the fast convergence toward purely packet-based network structures, mobile service providers will increasingly face competition stemming from emerging providers using IP native wireless access technologies. As in today's WiFi hotspots, access to such networks will be handled more flexibly, allowing for short-term use of resources instead of being based on long-term contracts. Additionally, emerging technologies such as WiMAX (802.16x) are currently becoming available on a larger scale, creating additional competition for mobile network providers and suppliers of fixed broadband services.

### ► 1.1.1 Main developments in next-generation wireless networks

Mobile wireless technologies beyond the currently implemented third generation (3G) are being investigated from multiple perspectives (Berezdivin et al., 2002). The wireless communication industry, including content players and application providers, are interested in understanding the potential behind next-generation networks and are trying to anticipate the potentially disruptive character of technology on current business models. Researchers from the technical departments at universities and other research institutions investigate new concepts for increasing the efficiency of the wireless air interface as well as develop innovative ways of managing wireless resource allocation to enable high-speed, high-quality wireless data transmission. In other fields such as computer science and information management new application and usage scenarios are developed and tested. Yet another research stream investigates the consequences for society and the changes in how humans and commercial entities will communicate over highly available wireless multimedia networks.

The term of *next-generation wireless networks* (NGWN) is highly ambiguous and is continually adjusting as new technologies for wireless networks become available on different layers of the network stack. For some researchers NGWN denotes a completely new IP-based network with a high-performing physical layer and an open and programmable architecture.[2] However, most recent research describes NGWN as a vision

[2]Regularly, the upcoming fourth generation (4G) standards, of which some of them are already deployed commercially, are seen as the enabling technology to realise the vision of full integration of voice and data services over a common IP-based transport layer.

of a ”truly seamless, multi-technology, multi-provider network, providing mobile and fixed users with a variety of multi-media services” (Berezdivin et al., 2002, p. 1). Part of this vision is the creation of a ”multinetwork”, consisting of multiple underlying network technologies, which are fully transparent to the user. This creates the prospect of creating a mobile ubiquitous service environment in which customised and personalised services can be provided by heterogeneous network access (Yang et al., 2006). Another aspect of ubiquitous services is the revocation of access barriers between network technologies and service providers.

Next-generation wireless networks ”comprise concepts and technologies for innovations in architectures, spectrum allocation, and utilisation, in radio communications, networks, and services and applications” (Berezdivin et al., 2002, p. 108). Many such concepts are intertwined on different levels of abstraction, and knowledge in multiple fields is required to produce substantial outcomes. Some of the key concepts of NGWN are a high-performing physical layer beneath the IP layer, an adaptive resource management, and the flexibility of services and applications to adapt to changing network conditions. In the following we review the main trends in all three areas.

A high-performing physical layer is often seen as the first prerequisite of NGWN. New technologies for increasing spectral efficiency are continuously evolving. Orthogonal Frequency-division multiplexing (OFDM) is seen as the most promising candidate to provide a highly adaptive transmission environment. Together with advanced processing techniques this will allow for data rates up to 50-100Mb/s (Berezdivin et al., 2002). Alternative ideas for increasing spectral efficiency include the dynamic allocation of radio spectrum, which allows the spectrum allocations to adapt over time according to the demand structure in a local area.

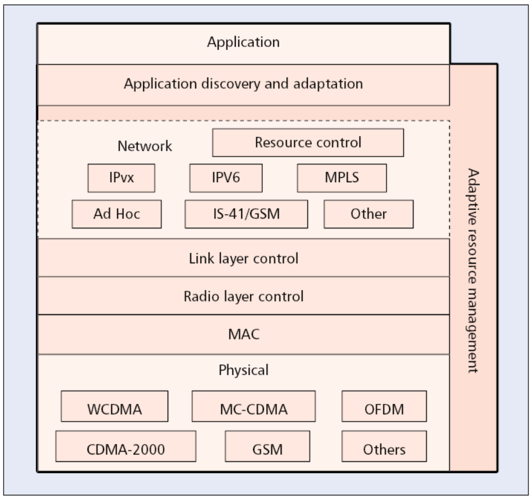



**Figure 1.1:** Fully adaptable multilayer architecture. Source: Berezdivin et al. (2002).

On the next layer, the high-performing physical layer needs to be complemented by an adaptive resource management (ARM) function. This function controls and adjusts parameters to supply upper layers with the desired Quality-of-Service, data throughput, or to manage resource usage based on overall budget restrictions. Such a function requires spanning multiple layers to adjust parameters on different levels simultaneously. ARM functions will also provide the dynamic management of resources from multiple networks. Resource allocation can then be dynamically switched between networks depending on factors such as current connection quality, congestion levels or pricing. ARM may also utilise resources from multiple networks simultaneously for increasing bandwidth or stability of the connection. Figure 1.1 provides an example architecture including a layer-spanning ARM.

Finally, NGWN requires an application layer, which is able to continuously adapt to new network conditions. An application discovery and adaption layer (Figure 1.1) provides adaption to the available bandwidth and quantity of the underlying network resources. A second function is the dynamic discovery of new application, which may become available with location changes of the terminal device. This layer may also include functionality to negotiate between user preferences and and communications resource availability, which also includes the possibility of using variable pricing options (Berezdivin et al., 2002).

### ▸ 1.1.2 Economic theory for network pricing

Economics has become an important and indispensable background for network pricing for various reasons. *Network economics*, a sub field of general economic theory, has evolved to analyse specific problems related to networks, and, in particular, the effects stemming from network externalities. Positive network externalities relate to the effect that consumers' willingness-to-pay increases by an increasing number of all consumers subscribing to the same service. Negative externalities, in contrast, arise from the shared consumption of limited network resources. Besides the effect of externalities, most networks share other properties such as very high installation costs (for example, to install a communication network or an electricity grid) but very low production costs (Shapiro and Varian., 1999). Such special properties have a strong influence on the economic analysis and the pricing and market structure.

Beside the special field of network economics, economic theory in general provides a lot of insights into market institutions and market rules and their effects on the equilibrium properties of such markets. A free market often produces economically efficient results with little centralised control. Microeconomic theory provides a large body of literature on market equilibria and decentralised control, offering a solid foundation for decentralised systems in network engineering. Market institutions in societies usually evolve from the interaction between individual entities and only limited possibilities for influencing such market rules exist. However, many markets are also designed by setting certain pricing and allocation rules. A market can then be designed in a way that individual entities behave in a certain way (Zhang et al., 2003).[3] Such rules are often formalised in a mechanism. A mechanism, in this sense, is a "set of procedures, penalties and rewards designed to guide selfish entities toward a desired outcome" (Dash et al., 2003, p. 40). Mechanisms can make use of different market institutions such as auctions, but can also rely on one-to-one negotiations. Prices play an important role in these mechanisms functioning as a tool for communication between entities and signalling back the current conditions in the network.

Another good reason for integrating economic thinking with network engineering is to broaden the optimisation perspective from the purely technical network view, in which performance is usually measured in maximum throughput and delay, to an economic perspective, in which additional measures, such as customer value and satisfaction, are considered (Courcoubetis and Weber, 2003). By adapting an integrated view of technology and customer demand, network mechanisms need to take into account user preferences. Economics can teach us how individual market entities behave under different circumstances. The concept of utility allows us to describe individuals by their valuation for goods and resources and let them behave accordingly. Thus, by complementing technical models with economic tools, the optimisation objective changes toward providing higher customer satisfaction while simultaneously achieving high resource utilisation.

### ▸ 1.1.3 The Role of Game Theory for Network Pricing

Game theory is a sub-field of economics that studies conflict and cooperation between interdependent agents, which are typical for an environment in which resources are allocated between multiple, independently-acting entities. Game theory provides the tools for structuring and analysing the problem of strategic choice (Turocy and von Stengel, 2001). A game, in this context, is defined as a formal model of an interactive situation, which involves several players. In a *cooperative game*, groups of players may enforce cooperative behaviour depending on the relative amount of power or knowledge held by the agents.

In contrast, *noncooperative game* theory analyses the strategic choices of players, which make choices only in their own interest. A central assumption in noncooperative game theory is the rational behaviour of the players, which lets them always choose an action that maximises the most preferred outcome. The goal of the analysis is to predict how a game will be played by such rational players.

The concept of *Nash equilibrium* is central in noncooperative game theory to identify the set of strategies, one for each player, such that no player has incentive to unilaterally change his action. In this context, a player's strategy describes his complete plan of action available to him. If a game has more than one Nash Equilibrium, the interaction

[3] A good example for this is the market for radio spectrum. In many countries regulators have decided to use auction formats to allocate scarce radio spectrum between interested parties. The authority's primary goal is to ensure the efficient allocation of spectrum. By choosing specific auction formats bidders in such auctions can be motivated to bid their true valuation for such resources.

between players can guide them to the "most reasonable" solution (Turocy and von Stengel, 2001). If no direct equilibrium exists with the strategies available to the players, they may randomise between several strategies with a certain probability. Randomising strategies in this way is called a *mixed strategy*.

In *simultaneous or static games*, players all select their action at the same time, and cannot observe the actions of other players in the game. This is, for example, the case for bidding in a sealed bid auction, in which players simultaneously decide on their bids to submit to the auction (Gibbons, 1992). In contrast, in *sequential games*, players select their strategy by some kind of predefined order. At least some players are then able to observe the actions of the other players to make their choice of strategy. Games, which are repeatedly played by the same players are called *dynamic or repeated games*. In this case players have at least some information about the strategies chosen by others and thus may base their play on past moves (Shor, 2006).

In the above cases it is assumed that players are aware of the payoff structures of their opponents. However, this may not always be the case and players may be unaware of certain characteristics of the other players. This is called a *game of incomplete information* or a *Bayesian game*, in which a player's type is private information. Players, only aware of their own type, need to form beliefs of the opponents types by using probability functions. Such games cannot be solved with the concept of Nash equilibrium. Instead, the concept of Bayesian-Nash equilibrium is used to understand the strategy choices of players in the game. A Bayesian-Nash equilibrium is defined as a strategy profile together with the beliefs held by each player about the types of the other players that maximises the expected payoff for each player, given their beliefs about the other players' types and given the strategies played by the opponents (Gibbons, 1992).

The popularity of game theory is closely connected with the rising importance of auctions in theory and practice (Turocy and von Stengel, 2001). By interpreting auctions as games between different players, it is possible to understand the properties of different auction formats in respect to different objectives such as efficiency and optimality. Furthermore, economists have used game theory to develop a detailed picture of how bidders would behave in different types of auctions.

## ► 1.2 Research Motivation

This work has been motivated by the vision about next-generation wireless networks and the extended potential of open network structures, in which network resources can be supplied on-demand rather than on a long-term basis. This opens up new possibilities for innovative business models to sell network capacities based on the actual user demand instead of binding customers in static contractual relationships. In the face of such new possibilities, customers and providers need to find ways of managing the additional complexity introduced by the increased flexibility network access provides.

When reviewing the existing literature on network pricing in wireless networks we can ascertain a strong focus on the technical aspects of next-generation networks. Most of the recent papers develop sophisticated concepts for using pricing on the lower network levels. This includes the use of pricing for power control and pricing for the allocation of limited buffer space between users. Economic concepts are used to prioritise network access based on the users' willingness-to-pay to complement the technical criteria such as channel conditions and user mobility. Pricing in such models is often used independent of individual user valuation but is based on predefined demand profiles (virtual pricing). By selecting a certain price plan a user is assigned to a specific demand category, which prioritises traffic accordingly.

While such concepts are crucial for effectively allocating resources in next-generation networks, we argue that it is also important to develop models which complement low-level resource allocation with customer-centric model, which reflect potentially real pricing policy of a wireless operator (Badia et al., 2004). Developing such models increases the insight in how users react to certain price models and how providers can use pricing to maximise profits. While it is expected that, from a practical view, resource negotiations will be handled by software-based agents, users will be able to communicate their preferences to such agents and are able to adapt behaviour according to their individual needs.

Another research focus in academia and the applied industries is the development of suitable concepts to manage network access in heterogeneous network structures based on user preferences and congestion levels. Many studies either assume that such networks are either owned by a single provider or, alternatively, that networks are cooperatively optimising network operations.

An alternative to this perspective, which we will follow in our research, is the assumption that providers optimise resource allocation independently without central control. Consequently, decentralised mechanisms are needed, which are suitable for a market consisting of multiple, independent entities, offering resources to customers which demand network resources. While such mechanisms do not need to be fundamentally different from the one-provider case, complexity increases due to the extended possibilities of users being faced with multiple options to acquire network resources.

Another important aspect, which motivates our research, is the notion that pricing can play a regulating role on multiple time-scales in order to build a robust wireless system. Figure 1.2 shows the connection between the traffic time-scale and the control time-scale. In each control layer pricing can be used to complement the technical traffic management function and can fulfill different tasks to police user demand. In this thesis we explore the use of pricing mechanisms on two different time-scales, namely the flow level and the admission control level, to optimise allocation efficiency and provider revenue in a setting in which providers directly compete for customers.

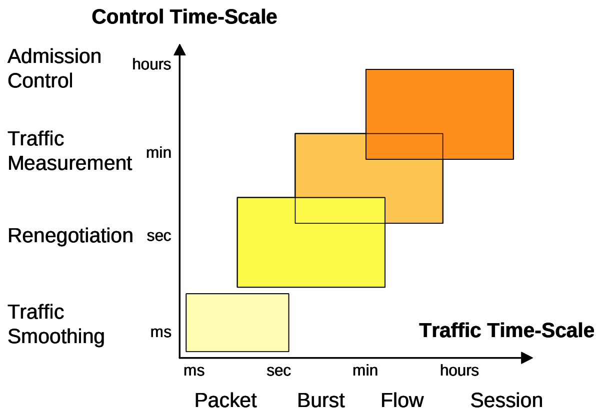



**Figure 1.2:** The correlation between the traffic time-scale and the control time-scale. Adapted from (Grossglauser, 1999).

## ► 1.3 Research Challenges

The need for revised, and potentially dynamic pricing strategies in wireless networks comes from various sources. First, if wireless network resources are seen as a public good which needs to be efficiently shared among users, dynamic pricing can help to provide signals to users to reduce waste and to adapt demand to the actual situation. Secondly, if resources are provided by profit-seeking firms, dynamic pricing is a tool for increasing revenue and gaining competitive advantage. Finally, from a more technical perspective, pricing can be used as a controlling tool, which complements the radio resource management functions such as admission control, flow control, and scheduling.

Despite the many advantages of dynamic pricing many different challenges need to be overcome to replace or complement current static tariffing structures. Some commentators have argued that it is easier to overprovision network capacity instead of building complicated control mechanisms. Others believe that more complex charging models will lack user acceptance if customers are directly faced with comparing and understanding such tariffing schemes. Martin Odlyzko, one of the most noted critics of dynamic pricing in communication networks notes about pricing in wireless networks (Odlyzko, 2001):

> *Another such area [for dynamic pricing] is likely to be in wireless communication. Although the bandwidth there is growing, it is orders of magnitude lower than on fiber, and will remain orders of magnitude lower. Hence wireless bandwidth will continue to be relatively scarce (at least relative to that on fiber backbones) and technical and economic methods to ration it may continue to be required.*

On the implementation side, one fundamental issue in designing dynamic pricing policies is the trade-off between engineering efficiency and economic efficiency (Semret, 1999). This trade-off includes many different factors. Particularly, the designer of a new pricing mechanism needs to consider the granularity of differently priced service offerings, the level of resource aggregation in time and space, and the information requirements for the mechanism to work properly. Additionally, the underlying network technology may limit the possibilities of implementing sophisticated pricing mechanisms.

Courcoubetis and Weber (2003) discuss the question about the feasibility of dynamic pricing in future network structures by comparing the advantages of flat pricing with the potential upside of adaptive pricing. While the main advantage of flat rate pricing is seen in the simplicity and predictability of charges for the end customer, disadvantages are the waste of resources, and unfairness between users. Two technological facts are presented to support the use of dynamic pricing, which allows users to adapt their demand according to the state of the network. First, in future networks it is likely that decisions are delegated to intelligent software agents, which reside in the customer end devices. While the customer keeps control over the overall spend, the software agent can absorb the decision complexity on shorter time-scales. Secondly, Courcoubetis and Weber (2003) argue that it is possible to implement charging structures in which sophisticated charges are potentially attributable to many stakeholders in the value chain of a service bundle and that the customer can be shielded from this complexity. One way to do this is by using a distributed service architecture in which brokering agents take over the task of finding and buying adequate network resources. While end customers only face simple tariffing schemes, brokers have to deal with the risk of fluctuating prices and service quality.

## ► 1.4 Categorisation of Research in this Thesis

The work in this thesis concentrates on the development and performance testing of suitable control models for resource allocation in a multi-provider environment. In all models we use pricing as the central element for communicating the network state between the participating entities, namely providers and network users.

The matrix shown in Figure 1.3 differentiates the chosen approach from the existing work. Most control models using dynamic pricing focus on the lower row under the assumption that only one provider supplies network resources to network users. We concentrate on models in which network resources are supplied by multiple, independently-acting providers. We distinguish two cases: (a) the case where multiple customers concurrently compete for network resources from multiple providers and (b), where multiple providers are faced with one customer at a time requesting network resources. While the first case is typical for a situation in which customers with elastic demand can adapt their consumption according to the level of congestion in the network, the latter case applies to an access control scheme in which users arrive in an asynchronous fashion and providers compete for the customer over price.

| | One user | Multiple users |
|---|---|---|
| **Multiple providers** | Admission-based pricing under competition | Flow-based pricing under competition |
| **One provider** | Usually not considered | Typical setting in most pricing models |

**Figure 1.3:** Differentiation of pricing models according to the number of providers and users considered.

## ▸ 1.5 Thesis Synopsis

This thesis has been structured into six chapters. Chapter 2 presents the research objectives and the research methodology. We differentiate our work from the existing research body and provide a substantiation of the selected methodology in the context of the defined research objectives.

Chapter 3 introduces the reader to the various objectives and perspectives of network pricing in wireless networks. In the second part of the chapter we introduce a classification structure based on the time-scale of the pricing decision, which is used in the third part of the chapter to categorise existing studies. The main focus of the literature review is on research using dynamic pricing concepts for the optimisation of network operations in wireless networks.

Chapter 4 presents a decentralised, flow-based congestion pricing scheme based on an auction mechanism in a wireless multi-provider setting. In this setting network resources are seen as a public good with the main objective of efficiently allocating resources among users based on their willingness-to-pay. A central element of the work is the development of a bidding strategy for users for distributing their demand among the available wireless networks. We examine the main properties of the bidding strategy and contrast them with alternative bidding strategies. The second part of the chapter presents the results from the simulation experiments to understand the performance of the proposed approach in settings that cannot be fully described by an analytical discussion.

Chapter 5 presents an admission-based pricing approach for provider revenue maximisation when providers face price competition at the time of user demand. The two main optimisation variables for each provider are the admission price and the cell size within which customers are served. In the first part of this chapter we introduce the general optimisation model and explain how competition can be integrated in the decision process. The second part of the chapter presents the simulation experiments, which show how optimal prices and cell-sizes change when varying the model parametrisation such as user density or the size of the shared coverage area.

Chapter 6 provides the reader with the details of the agent-based simulation platform used to conduct the experiments in Chapter 4 and 5. The first part of the chapter briefly reviews the platform selection process. The second part reviews the theory and main concepts of agent-based software development and, in particular, the use of agent-based architectures for computer simulation. Since the development of the simulation environment was a central achievement of this research, the third part of the chapter presents in detail the developed architecture and ontology structures which both simulation platforms have been based on. Finally, we briefly present the extensions needed to implement the functionality needed for the two pricing mechanisms in the different contexts.

Chapter 7 provides a summary of the conclusions that can be drawn from the work presented here and points out directions for future research.

# Chapter 2
## Research Objectives and Methodology

### ► 2.1 Introduction

The purpose of this chapter is to explain the scope of the research and to justify the methodological approach taken as the most appropriate for the defined objectives. The first section discusses the research objectives and the arising research questions to be answered by this study. Section 2.3 provides an overview of the research limitations. In Section 2.4, the contributions of this study are presented. Finally, Section 2.5 describes the chosen research methodologies and the connection between them. Since simulation, one of the main methodologies, is not common in MIS research, we provide a short explanation of the term and describe the theoretical foundations of simulation as a valid research methodology.

### ► 2.2 Research Objectives

The objective of this study is to *improve the understanding of decision-making in a wireless noncooperative multi-provider environment*. In such an environment, rational customers are free in their decision to join or leave a network depending on the signals they receive from all available networks and their preference structure for the resources offered. Providers are interested in attracting customers, to increase their revenue and to control access

to keep their network load within the feasible operating region. In particular, we are interested in *how to use pricing in such an environment as a means of signalling the current states of the network to and between users so that certain objectives can be reached by the designer of the system*. Another main objective set for this research is to *understand behaviour of rational entities in a wireless network environment under defined rules of interaction, when they are faced with multiple choices of fulfilling their demand*.
Since the above formulations of the research objectives have been formulated as open and do not define the distinct goals to be accomplished by this research, we split them into a series of questions:

Q1 What is a suitable categorisation framework for pricing in wireless communication networks, with which the existing literature can be classified?

Q2 What are relevant studies and research articles on pricing of wireless resources and how do they fit into the developed categorisation framework?

Q3 What is the optimal behaviour of a rational user with the possibility to connect to multiple wireless networks, when faced with competition from other customers?

Q4 What is the behaviour of a revenue-maximising wireless provider when faced with price competition from other wireless providers partly or fully covering its service area?

Q5 How can a simulation platform be developed, which allows us to experiment with different pricing mechanisms in a wireless multi-provider network?

The above research questions provide a general frame for our research. In each of the chapters we refer back to selected research objectives.

## ▶ 2.3 Delimitations of Scope and Key Assumptions

The above research questions put a strong focus on the economic side of competition rather than on implementing sophisticated resource allocation mechanisms on the technological side. While it has been essential for this research to understand the underlying technological principles, especially in the sense of what will be feasible and what will be infeasible in next-generation wireless networks we put our focus on understanding how technology can be complemented by economic models to control user demand and to reach certain design objectives. In the following we describe the key assumptions taken on the modelling side of user behaviour as well as on the side of modelling of wireless network capacity.

When modelling resource allocation in communication networks by means of economic methodologies, many different objectives can be followed. Our research questions strongly focus on how network entities, customers and providers, behave in situations in which network access is provided by multiple suppliers and users can switch between resources at no cost. Very different ways to model such situations are possible depending on the sophistication of the agents' decision framework and their ability to interact. The key assumptions for this research can be summarised as follows:

A1.1 Users behave rationally and selfishly to maximise their satisfaction in their own interest. Utility is drawn exclusively from consumption.

A1.2 Users only act upon price and take a predefined quantity and quality of the resource as given and non-negotiable.

A1.3 All entities are myopic or shortsighted in a way that they maximise their payoffs at the given point in time. They are assumed to be unable to foresee the consequences of their current actions for future outcomes.

A1.4 Users do not cooperate with each other to jointly optimise resource allocation but act independently. Cooperation may only occur indirectly as motivated by the incentives of the implemented market institution.

With the above assumptions we follow standard assumptions taken in network economics. One important consideration is that resource negotiation is rarely directly influenced by human behaviour but taken over by software agents, which fulfill the automated task of negotiating for resources (Courcoubetis and Weber, 2003). While humans will be required to input their preferences in the form of simplified negotiation rules, they have no influence on the actual negotiation process.

To be able to concentrate on the economic aspects of competition in the described setting we largely abstract from the complexity of the technological implementation. To represent wireless resources in this setting, we therefore take simplifying assumptions, which can be summarised as follows:

A2.1 We only model the downlink (base station to mobile terminal). While, in principle, the same mechanisms presented in this work can be used for modelling resource allocation in the uplink, we have decided to concentrate on the downlink as the main bottleneck in future IP-based wireless networks.

A2.2 Communication for negotiating for resources is always free of errors.

A2.3 In the first part of this research we assume that the available wireless capacity in the form of bandwidth is fixed and can be arbitrarily distributed among users. We also assume that Quality-of-Service can be provided to all users irrespective of their position and level of mobility.

A2.4 In the second part of the research we assume a simple distance-based propagation model, in which the received bandwidth depends only on the relative distance between the base station and the mobile terminal. We assume that Quality-of-Service is measured only by the bit error ratio (BER) experience of the user.[1] We also do not consider the effects of cell handovers and user mobility.

The technological key assumptions given above are clearly abstracting from the real complexity and practical issues in wireless networks but allow us to search for general concepts for pricing network resources rather than working on technology-specific solutions. It is important to understand that the work presented in this study does not directly lead to practical pricing and charging schemes which could be implemented in next-generation wireless networks right away. In the next section we therefore explain the distinct contributions of this study.

## ► 2.4 Contributions of this Research

In order to achieve the goals defined in the research objectives, this thesis makes the following contributions:

C1 **Classification framework:** We have defined a classification framework using the time-scale of pricing to differentiate between the different pricing approaches in the literature and to categorise our own work. While the use of time-scales is not new, we propose to complement the existing classification frameworks developed for fixed networks, to be suitable in the wireless domain.

C2 **Literature review:** By surveying the literature on network pricing in wireless networks we provide a comprehensive overview of existing concepts and proposals.

C3 **Implementation of the Progressive-Second-Price (PSP) auction in a multiprovider environment :** We use the Progressive-Second-Price Auction format as the basis for resource allocation in an environment in which customers can access multiple providers.

C4 **Bidding strategies:** We have developed a bidding strategy for agents faced with multiple access and the possibility to bundle resources from multiple access networks. We have shown analytically that this strategy is the truthful best-reply of an agent in such a situation. We have also examined the resulting properties of the entire multi-auction system if *BalancedBid* is applied by all agents. Additionally, we have proposed several alternative bidding strategies when agents are limited in a certain dimension of their behaviour.

C5 **PSP simulation results:** We show by simulation how the proposed bidding strategies behave when changing certain input parameters such as agents' valuations

[1] Since we model resource allocation on the physical level, the BER has a direct influence on packet delay and jitter on the link layer. With a higher error rate packets need to be retransmitted more often, which increases the overall delay experienced by the user. However, we do not further consider such effects in our research.

and distribution of resources among sellers. We also provide extensive results for a multi-cell environment and compare the performance with alternative resource allocation mechanisms.

C6 **Pricing strategies for providers in a wireless multiprovider access scenario:** We describe optimal pricing strategies of providers facing direct competition from other providers when network cells partly or fully overlap and if prices can be set at admission time. We provide the results of the game played among providers and propose a heuristic to approximate optimal behaviour to maximise provider revenue. We provide extensive simulation results from implementing the heuristics in the simulation environment.

C7 **Development of a simulation platform:** One tangible practical contribution of this study is the development of a flexible and highly modular simulation platform for simulating resource negotiation from an economic point of view. This platform will be the base for subsequent research on various resource allocation mechanisms.

## ▸ 2.5 Methodology

A solid methodology is one of the most important aspects of a research journey. A research methodology is commonly defined as *"a combination of the process, methods, and tools that are used in conducting research in a research domain"* Nunamaker et al. (1990, p. 41). This thesis employs a multi methodological approach, which covers two of the four research categories distinguished by Nunamaker et al. (1990) in the proposed research framework (Figure 2.1). This research framework generally describes the different methodologies, which can be used to conduct research in the field of MIS.

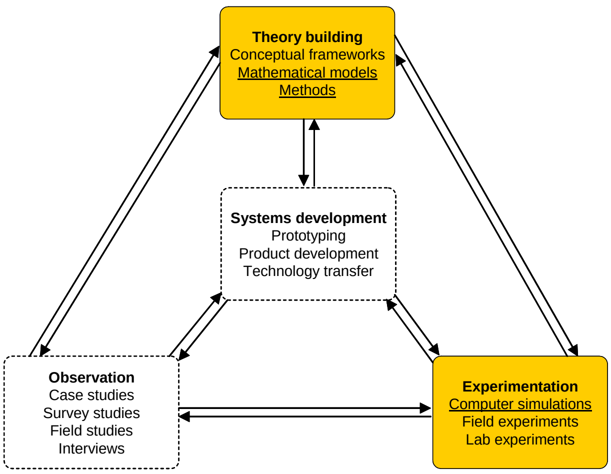



**Figure 2.1:** The research framework proposed by Nunamaker et al. (1990).

### ▸ 2.5.1 Description of the applied research methodologies

One core part of this thesis consists of mathematical models to formally describe the situation in a wireless multi provider network. We make use of standard tools commonly used in economic theory to describe individual entities, their preferences, their mode of interaction, and the level of knowledge each entity possesses. Game-theory and constrained maximisation methods play important roles in understanding the outcome of the interactions between such entities.

Another methodological element extensively used in this thesis is agent-based simulation. Since we are focusing on the economic aspects of network control and want to understand how pricing can be used to manage demand, we need a method which allows us to gain information on the effects of the defined market rules, individual behaviour and the resulting outcomes. Agent-based simulation has been widely accepted as valid research methodology in economics because it lets the researcher model each entity with individual behaviour, preferences, and beliefs. Results from experiments can be used to analyse the micro level as well as the macro level by aggregating the results for the entire market.

While we could have reached simular results by using traditional simulation approaches, the use of the agent paradigm has substantially increased the flexibility of the simulation model. To introduce competition, for example, only required the creation of a new agent representing the base station of a wireless cell. By applying a modular approach each agent can adapt behaviour simply by means of its initial parameterization.

Beside producing the data for the numerous simulation experiments presented in this study we have made extensive use of simulation to build up intuition about the properties of the developed behavioural strategies before describing them formally. By learning from preliminary experimental results we could collect evidence about the existence of certain properties such as maximisation criteria or stability. The intution gained by analysing the simulation results could then often be proven by the analytical model. The use of simulation as an exploratory tool is also mentioned in the literature (MacKie-Mason and Wellman, 2006). Figure 2.2 shows the principle feedback cycle, which was used in this research in both directions.

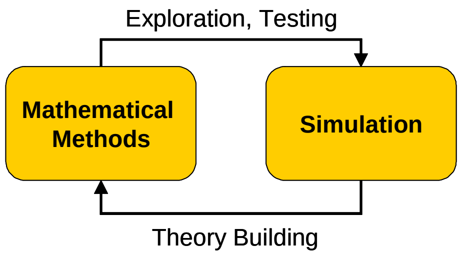



**Figure 2.2:** The feedback cycle between mathematical models and simulation.

To analyse the simulation experiments in a structured and acceptable way, and to produce results that are statistically significant, we make use of standard statistical methods. While some of the simulation experiments have been based on a deterministic setting, in many other experiments input parameters have been randomised. In such cases we use multiple repetitions and statistical methods to calulate the confidence intervals of the derived output variables.

The remainder of this chapter further defines and substantiates simulation as a research methodology and describes the theoretical foundation for simulation as a formal research methodology. At this point we do not further elaborate on agent-based simulation techniques but refer the reader to Chapter 6 of this thesis for a detailed description of the agent paradigms and the use as a tool for simulation.

### ► 2.5.2 Definition of the term simulation

As with many terms, multiple definitions and meanings of simulation have emerged depending on on the field of study and the particular interests of researchers using simulation as research methodology. While simulation, in general, is not necessarily connected with the use of computers, due to the influence of increasing computing power and the sophistication of available modelling tools, it has become closely associated with computer models. Fishwick (1994) defines simulation as follows: *"Computer simulation is the discipline of designing a model of an actual or theoretical physical system, executing the model on a digital computer, and analyzing the execution output."*

Gilbert and Troitzsch (2005) describe simulation as a particular type of modelling. A model is a simplified description of a real or hypothetical system or structure. By using a model the researcher can abstract from the complexity of the real system and can concentrate on some few factors, which are of direct interest.[2]

Beside the term simulation it is also important to understand the typical elements of a simulation model to establish a boundary to other methodologies such as abstract mental experiments or prototyping. Sonnessa (2004) distinguishes four elements of a simulation model:

- a set of parameter variables
- a set of entities or objects, which can either be structurally present or volatile over the simulation period
- a set of static or dynamic relations among entities
- a well-defined time representation or formalism.

[2]For example, when simulating a new communication protocol for a large communication system, researchers can take simplifications such as assuming an error-free transmission channel or disregarding certain processing times of message retrieval. If such variables have no direct influence on the communication protocol itself they can be kept constant or be omitted from the model.

### ► 2.5.3 Theoretical foundation of simulation as research methodology

Simulation differs considerably from the more established and better known methodologies in research (Nigel, 1996). Depending on the field of research, it has been more or less accepted as a valid methodology for conducting research.

In mathematics and natural sciences simulation has often been considered as an extreme technique for solving hard problems. For example, simulation is a common methodology used in queuing theory, numerical analysis, or operations research (Sonnessa, 2004).

In recent years, simulation has become a more and more established research methodology in social sciences, which often adopts an agent-based approach for modelling micro-structures and to analyse emerging phenomena on the macro-structure. It is often used in an interdisciplinary sense, combining different fields of knowledge, such as artificial intelligence, learning algorithms, genetic programming or economics.

In economics itself, simulation has found wide acceptance as valid research methodology. The application of game-theory models usually becomes infeasible as soon as the problem complexity of the model grows (MacKie-Mason and Wellman, 2006). In contrast, with computational experiments, researchers can systematically investigate new agent strategies or market mechanisms for testing their performance.

Axelrod (2003) introduces a very strong notion of simulation as a formal research methodology by arguing that simulation can be thought of as *a third way of performing research*. In contrast to the two standard research methods of *inductive reasoning* and *deductive reasoning*, one can define simulation to lie in between or to combine certain elements of both. While induction is usually described as moving from the specific to the general, deductive reasoning starts with general observations to make conclusions for a specific case. Induction is often used to extract general rules from a limited set of experiments or case studies such as interviews or laboratory trials. In contrast, deductive reasoning uses general rules, which have been shown to be applicable in all cases. Researchers may conduct multiple deductive experiments to demonstrate that the law holds true in many different circumstances.

Simulation is able to combine both types of reasoning by using the given premises and rules, and producing rich data, which can then be analysed by inductive methods such as statistical tools or empirical analysis. While induction is used to identify patterns in the data, deductive methods are used to "find consequences of assumptions" (Axelrod, 2003). In consequence, Axelrod (2003) argues that *"simulation modelling can be used as an aid to intuition."* While simulation rules can be simple in nature, they can lead to consequences that may not be obvious to the researcher. Such a phenomenon is called *emergent property* or *emergent behaviour*.

Fromm (2005) defines emergence as follows: *"A property of a system is emergent, if it is not a property of any fundamental element, and emergence is the appearance of emergent properties and structures on a higher level of organization or complexity"*. Emergent behaviour can appear when simple entities form a more complex behaviour as a collective. This behaviour is not the property of a single element but emerges from the interactions of agents in the system. One reason for the difficulty of predicting emergent behaviour is the number of interactions between components of a system. However, a large number of interactions is not sufficient to guarantee some form of emergent property and many different communication acts may weaken or cancel out emergent patterns. Sometimes, the large volume of interactions may even hinder the identification of such properties by creating ”noise” obstructing from the actual ”signal” (Fromm, 2005). Still, emergent behaviour is usually robust and resilient and can be recognised even if some elements fail or are missing.

# Chapter 3

# Knowledge Domains and Literature Review

## ▶ 3.1 Introduction

The previous chapters have set the frame for this research by defining the research motivation, the objectives and the methodology. This chapter introduces the reader to the various aspects of network pricing with a focus on wireless networks and to review the existing research in this area.

Network pricing has become a topic of multi-disciplinary interest, which has been studied from multiple perspectives (Wang, 2006). Besides introducing a common terminology, we cluster the objectives of pricing in four main categories and explain the different aspects within each objective category (Section 3.2). Besides the theoretical aspects of pricing we also provide an overview about pricing practices in mobile and wireless networks and briefly describe the different stages affecting the worldwide diffusion of the technology.

An important contribution in this chapter is the pricing time-scale framework, which has been developed with the intention of providing a classification structure for different pricing approaches in the literature (Section 3.3). We distinguish five different time-scales, based on the unit of measure and the lengths of the pricing intervals used by the pricing model. Even though the developed framework is partly based on existing frameworks developed for pricing in fixed networks, it introduces new aspects unique to pricing in wireless networks.

This chapter (Section 3.4) provides an extensive overview of the existing body of literature about network pricing with a focus on wireless networks. We use the classification framework developed in the previous section to cluster the different studies. Besides providing an overview we review in detail the work most closely related to the research in this thesis. Most existing work concentrates on the one-provider case and competition between providers can only be found in very few studies.

In Section 3.5 we substantiate the selection of the scenarios in the following two chapters based on the time-scale framework. Finally, in Section 3.6, we conclude this chapter by summarising our findings.

## ▶ 3.2 Aspects of Network Pricing

In this section we explain some of the central aspects of pricing in communication networks. Part of the material serves as background for the work presented in the following chapters; other aspects help to delimit pricing concepts from other related disciplines such as technical resource allocation or the provisioning of differentiated services.

### ▶ 3.2.1 Terminology

Throughout the thesis we mostly adopt the terminology defined by Stiller et al. (2001a) and Courcoubetis and Weber (2003).

A *price* represents a monetary value associated with a unit of a specific service type. *Pricing* covers the specification and setting of prices for goods, specifically networking resources and services in an open market situation (Stiller et al., 2001b). Pricing strategies can involve many different factors for determining the price of a product or service. Prices can be based on the underlying costs for a resource, on an return-on-investment measure, on the value they deliver for a customer, on the competitive situation, or on a combination of all factors.

A *charge* is the amount of money that is billed for a service (Courcoubetis and Weber, 2003). *Charging* is used as an overall term, depicting all tasks required to calculate the finalised content of a billing record (Stiller et al., 2001b). Charging is not to be confused with the billing process, which also includes the customer invoice process and customer data management tasks.

We have to stress that the distinction between *price* and *charge* is sometimes blurred. In this thesis we prefer the expression pricing for expressing the actual determination of the unit price for network resources. In this sense the term *charge* can be used interchangeably and stands for the unit price multiplied by the amount of resources consumed by the user.

*Tariff* is an expression that refers to the general structure of prices and charges for a service usage. It defines how the charges for a service are computed. For example, a tariff of the form $a + pt$ defines a subscription fee $a$ independent from usage, together with a charge $p$ depending on the duration of the service usage $t$. The process of deciding upon the algorithm used to determine a tariff is called *tariffing*. For calculating charges the tariff may contain discount strategies, rebate schemes, or marketing information (Stiller et al., 2001b).

*Billing* defines the collection of charging records, summarising charging content, and delivering an invoice, including an optional list of detailed charges to a user (Stiller et al., 2001b). The billing process collects all relevant usage information and translates it into charges using the pre-defined tariff structures (Cushnie, 2003). The billing process might collect elements from different decentralised databases to create an overall billing statement for all services relevant to the customer.

Besides the terms above it is also useful to provide definitions for the terms *utility*, *willingness-to-pay*, and *consumer surplus*. Utility refers to the the value a customer derives from using network services (Wang, 2006). Utility can be defined as a time- or usage-based measure or can be defined as a one-time event if a connection has been successful. The willingness-to-pay expresses the monetary value a customer is willing to spend for using network services. Assuming rationality, the willingness-to-pay equals the user's utility. The difference between the utility and the overall charges for network usage is defined as consumer surplus.

Looking at the aggregated market we can define the sum of utilities of all users minus the costs of providing the network services as *social welfare*, which equals the sum of the consumer surplus plus the *provider profits*. (Courcoubetis and Weber, 2003). The maximisation of social welfare results in economic efficiency.

### ► 3.2.2 Delimitation of network transport services

In today's competitive environment wireless service providers frequently sell more than pure data transport capacity to their customers. Part of the strategy to sell services as bundles is to complicate direct price comparisons between operators. *Value-added services* bundle functionality from different levels into one product For example, a Voice-over-IP telephony service (VoIP) can be seen as value-added services because it combines a directory service, a signalling service, a data transport service, and a billing service into one product (Courcoubetis and Weber, 2003). Regularly, providers offer bundled voice and data services, which already include a certain amount of free usage, to increase the intransparency of charges for individual services.

In its core, most of the network services offered by wireless service providers include the actual data transport service or telecommunication service. The US Federal Communications Act of 1996 defines *communication transport services* as "*the transmission, between or among points specified by the user, of information of the user's choosing, without change in the form or content of the information as sent and received*" (Aufderheide, 1999, p. 144). In contrast, *information services* are defined as "*the offering of a capability for generating, acquiring, storing, transforming, processing, retrieving, utilizing or making available information via telecommunications*". (Aufderheide, 1999, p. 144).

With the widespread use of packet-switched networks, the above definitions have become problematic, since data is processed either inside the network or at the network edges (Courcoubetis and Weber, 2003). Since this type of data manipulation is closer to the lower-level transport services, *packet-based transport services* refer to the transmission of data including the processing of data needed to enable packetised network transport.

In this study we concentrate on the pricing of transport services that may complement higher-level services. We use the term *network resources* as a generic expression for the capacity of a communication channel. The capacity of such a channel can be measured by its throughput in terms of bandwidth or the available time- or code-slots. We use the the term *communication services* interchangeably with *network transport services*. With the term *differentiated services* we refer to data transport services for which certain quality guarantees, for example in terms of minimum throughput or package delay, have been made.

In many studies the term *telecommunication services* stands for voice-centered applications while *communication services* refer to packet-based data applications (Courcoubetis and Weber, 2003). We will use the latter term as a generic way to refer to transport data between entities.

### ▶ 3.2.3 Objectives of network pricing

Pricing in wireless networks is an interdisciplinary topic, which involves the interest of many stakeholders (Wang, 2006). Depending on the objectives of the designer and the granularity of the chosen pricing approach we can distinguish several, possibly conflicting, objectives. In the following section we distinguish four main objectives.

#### Pricing for economic efficiency

When communication transport services are seen as a public good, a central objective of network pricing becomes the regulation of demand so as to maximise *economic efficiency*. In general, economic efficiency is "*about producing as much value as possible with given resources, preferences, and technology*" (Varian et al., 2005).

A natural question arises as to who could be the stakeholders aiming for maximising economic efficiency? There are two main perspectives; first, a market regulator is usually interested in economic efficiency to allocate resources in the most beneficial way for society. There are usually several ways to influence a monopolist's output and service offerings, but one of the most important ones is by defining prices so that users adapt their demand according to their valuation for the services.

Second, a network provider itself may be interested in the maximisation of economic efficiency. While the maximisation of revenue is usually the first goal, long-term customer satisfaction may be as important to secure long term profit. Since customer valuation expresses the customers satisfaction or "happiness", a provider can use this information to improve customer satisfaction and general acceptance of service offerings.

Economic efficiency is of interest for two reasons. The first reason is a positive one, which explores the search for value as the driving market force. For example, for a provider, it is essential to understand the customer's valuation in order to design attractive products and services and to gain as much market share as possible. The second reason for the interest in economic efficiency is normative. The measurement of economic efficiency allows us to compare different policies (such as tariffing schemes) and understand the differences in terms of value created. In this way different resource allocations can be brought into a normative order (Varian et al., 2005).

Economic theory tells us, under some stringent assumptions[1], that there exists a price under which producers and consumers choose services (and quantities) in such a way that social welfare is maximised. In this case, prices equal the supplier's marginal cost and each consumer's marginal utility at equilibrium (Courcoubetis and Weber, 2003). The main advantage of this perspective is that no central control is needed to enforce such outcome but market participants can reach equilibrium in a decentralised fashion. Different ways exist to identify the equilibrium price. If all cost and utility functions are known, the price can simply be calculated by solving the optimisation problem of the system. However, in most cases, such information is only known by the market participants themselves. In this case, the price can be found through an iterative process (also called tâtonnement), in which prices are adapted in each period after the demand has been observed (Varian et al., 2005).

We can also distiniguish different time-scales of economic efficiency and the associated economic theory. Looking at the entire market for wireless network services in one economy implies a different notion of economic efficiency than a single wireless access-point with a countable number of active users. While in the first case interactions between the service provider and potential customers are indirect and customers are price-takers, the latter situation can be modelled as a game in which customers are aware of their influence on prices.

### Pricing for cost recovery

The most important concern of any wireless network operator in a competitive setting is cost recovery. If costs are not covered by the revenues obtained from selling network transport services, the supply of the network service may not be sustainable long-term. However, setting prices solely according to cost aspects may reduce user demand and

[1]Such stringent assumptions are, for example, the concavity and convexity of utility and cost functions, and market transparency by all participants.

social welfare (Courcoubetis and Weber, 2003). Consequently, a provider needs to develop the basic understanding about the pricing boundaries and constraints in the market given on the one side by customer demand, and, on the other side, by the competitive situation.

One possibility to ensure that costs are recovered is to set prices according to marginal costs. While marginal cost pricing under competition leads to economic efficiency, several other problems arise. First, marginal costs in networks may be difficult to obtain, especially on a packet-based level (Courcoubetis and Weber, 2003). Second, marginal costs do not contain fixed costs, which need to be also recovered from running the network. One possibility described in the literature is to reinterpret marginal costs as long-term marginal costs, which contain components to cover fixed costs and a continuous expansion of network capacity (Courcoubetis and Weber, 2003).

In a market which is partly regulated due to partial competition, an extensively discussed approach is to allow providers to charge prices higher than the short-term marginal costs. Prices formed in such way are referred to as *Ramsey prices* (Laffont and Tirole, 2000). The advantage of this approach is that besides cost recovery, social welfare can be increased. If the cross-elasticities of the different products offered by the provider are zero, Ramsey prices are inversely proportional to the demand elasticities. To implement Ramsey prices, a provider (or the regulator) needs to have full knowledge of the product demand functions and the costs. If this information is not available to the regulator, an iterative process can be used to obtain estimations.

The solution most often used in practice are non-linear tariff models of which the two-part tariff is the most popular one. With a two-part tariff scheme a provider charges a fixed subscription amount $a$ together with usage-based charges $p$. The joint revenues cover the fixed and marginal costs. If a customer is faced with a two-part tariff he optimises $u(x) - a - px$, where $x$ is the quantity consumed. As a result, he adapts his consumption so that his marginal utility equals the usage-based price $p$ (Mitchell and Vogelsang, 1991). Depending on the definition of the subscription fee and the usage-based charge, two-part tariffs can reduce social welfare if consumers with smaller demand are deterred from subscribing to the service. Therefore, such parameters need to be carefully chosen. A possible way to increase the provider's profits is to discriminate between customer segments and to set fixed charges in proportion to the net benefits that customers in the respective segment receive (Courcoubetis and Weber, 2003). However, this requires some market power by the provider and information about customers' valuation for services.

### Pricing for revenue maximisation

While economic theory suggests that under perfect competition, prices are driven toward marginal costs, in practice, markets are never perfectly competitive nor fully transparent. Additionally, effects such as switching and lock-in costs deter customers from always choosing the best offer available on the market. For example, we have also already discussed how providers hide true costs by bundling multiple services. In all such cases providers have some degree of freedom to set prices above marginal costs to increase revenues from selling network resources. However, the customers' willingness-to-pay provides the upper limit for maximising revenue from a provider perspective.

Besides the two extreme cases, a monopoly situation in which the provider sets marginal revenue equal to the marginal cost, and the situation of perfect competition, in which providers are price-takers, the oligopoly is often seen as the most realistic scenario for modelling the situation in communication markets. Since investments in wireless networks are usually high and entry is limited by licensed spectrum bands, only a limited number of providers will be able to setup the required infrastructure for a cellular network.[2] Participants in an oligopoly game need to consider the consequences of their actions on the market and on their net benefit when deciding for a particular strategy (Courcoubetis and Weber, 2003). Standard models have been developed to explain different aspects of competition in an oligopoly such as the *Cournot game*, in which players set their production quantities and prices are formed dynamically; in contrast, the *Bertrand game* lets players set prices and, in turn, customers decide on the offer with the lowest price. The outcome of both games depends on a series of conditions such as capacity constraints and additional qualitative decision variables.

In an oligopoly, games can also be formulated from different perspectives. One possibility is to define the customers as the players of the game. A provider can maximise its revenues by carefully choosing the design of the game (the pricing and allocation rules). For example, by using an auction, a provider may learn about customer preferences in an iterative process and may also be able to skim as much of the customers' benefit to increase its revenue. However, obtaining such information usually comes at a cost, which leaves some surplus to the customers (Krishna, 2004).

Another important issue with revenue maximisation is the availability of information to the different market participants. Customers, who have no transparency about available providers and tariffing options, will be limited in finding the optimal solution. On the other hand, providers need to learn about customer preferences to make optimal pricing decisions in terms of revenue maximisation. One way to obtain more information about the users' valuation for resources is through pricing (Courcoubetis and Weber, 2003). For example, a provider may offer a set of different tariffing schemes to its customers. The user will choose the tariff which minimises his expected overall charges. The tariff choice therefore reveals important information about user preferences and usage patterns.

### Pricing for traffic and resource management

In the past decade, network pricing has not only attracted economists and practitioners, but has also found wide acceptance with network engineers, who started to explore

[2]The situation looks different for the setup of unlicensed wireless data networks, which operate in a very limited range and where investments for setting up a new access point are very low. If network access becomes transparent on an IP-level and smaller, locally limited providers enter the market, the competition model for wireless networks may change considerably.

pricing as a complementing function of network control and resource management. Many of such pricing schemes are designed to improve purely technical solutions to include the user view. In wireless networks the technical functions of real-time network control of the radio interface are summarised in the *Radio Resource Management* (RRM) function (Wu, 2005).[3] It ensures that the required connection quality is available during the connection. At the same time it increases the capacity of the wireless channel by efficiently using the available resources. Depending on the technology used for the wireless air interface different components of RRM are used.

Pricing can be used with the majority of RRM components to control the traffic load induced by different users and to keep the network within stable conditions. Figure 3.1 shows a generic model of RRM functions in a wireless network with differentiated service support. Note that for a specific implementation, only some of the components may be used. In addition, the sophistication of the RRM modules can vary widely. In the following, we provide a brief description of each function and explain the role of pricing to complement the technical function.

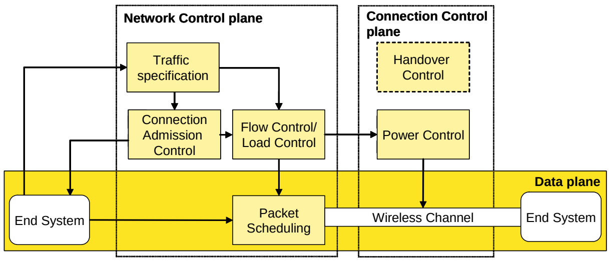



**Figure 3.1:** Generic radio resource management components for providing differentiated services in wireless networks. Adapted from Wu (2005).

Two different levels of RRM can be distinguished. The *Network Control Plane* contains all RRM functions that are centrally performed either before or during the connection setup. The *Connection Control Plane* manages the resources of a specific connection.

**The traffic specification function** defines the source characteristics and QoS requirements. It can either be measurement-based or given a priori by the traffic source. The a priori specification may be given by different parameters such as minimum guaranteed throughput, maximum average delay or maximum error rate. Users may be influenced in selecting such parameters by the prices defined for certain quality guarantees and will adapt their request accordingly.

**Connection admission control (CAC)** decides if a new incoming service request should be admitted or rejected given the current state of the network and the traffic specification of the new source. CAC serves two functions; first, it protects existing flows from

[3]Other functions, which are out of the scope of this work, but which have also been influenced by network pricing, are *network planning* and the *offline optimisation of network resources* (Wang, 2006)

experiencing degraded service quality by admitting new flows to the network. Second, it ensures that sufficient resources are available for the given QoS requirements of the new flow. Admission-based pricing can control the number and structure of requests made to the network. Prices can be used as a way for signalling to users the current congestion level in the network. In times of high demand, higher prices can prevent low-value customers from entering the network. In multi-cell networks, the CAC function has to additionally plan for handovers of established connections.

**Load control (LC)** Once the new connection is admitted, the *load control* function polices each flow to ensure that the resources are shared in the predefined fashion. In multi-cell networks, load control may also be used to balance load in the different cells to optimise overall resource usage. Load control continuously measures the uplink and downlink traffic and the transmission rates of the different data sources. In an overload condition, LC reduces the load and brings the network back into operating state. Within this framework prices can serve as a signal to the users to adjust their demand. Prices can either be set centrally by the RRM or can be formed decentrally by the mobile terminals declaring a price they are willing to pay for a certain share of resources in the given moment. Auction mechanisms are often used to realise this type of decentralised optimisation.

**Packet scheduling (PS)** allows the network to give priority to certain flows. PS mainly controls the packet delay experienced by the data flows by prioritising packets to be admitted into the wireless channel. Simple scheduling algorithms work with one queue and ensure the fair sharing of the wireless channel. Complex packet schedulers employ multiple queues for different service classes and include information about the current channel quality in the scheduling decision. Prices can influence the priority of packets given in the scheduler. For example, users running a connection with stringent delay requirements may be willing to spend more for the prioritisation of their data packets to achieve a satisfactory performance. In contrast, best-effort users may be indifferent about delay and may only care about throughput no matter how bursty the connection may become. Prices can be an indicator of such preferences and the correlated willingness-to-pay for prioritisation.

**Power control (PC)** is an essential function of all network technologies in which the wireless channel is shared between users. In CDMA-based systems, PC is essential to make efficient use of the available spectrum. PC schemes have two main tasks; first, PC minimises the interference between users (intracell) and cells (intercell) in the reverse link. Therefore, it needs to find a feasible power allocation, which minimises overall transmission power of all terminals. Second, it needs to ensure adequate signal quality and a sufficient signal level at the receiving end. Power control schemes often use pricing to limit the transmission power used by each terminal. Prices motivate users to find the smallest power allocation which maximises their net utility. Other approaches use pricing to allocate power between users in the downlink. This can be, for example, done by an auction in which each user announces a price per power share.

**Handover control (HC)** manages the switch of users from one cell to another. In simple schemes, a mobile terminal switches between cells if signal quality in the existing cell falls below a certain threshold. In CDMA-based networks, mobile terminals perform a soft handover, which allows the terminal to stay connected until a stable connection to the next cell can be established. In a heterogeneous wireless network HC can help deciding which access network to use for the particular service. For example, if a mobile terminal has access to two base stations, it can decide the best station to connect to based on the current price level. Price structures could even be more complex and define prices for different service classes and traffic characteristics. In this case the mobile terminal needs to evaluate the different options and decide for the option according to some predefined decision logic.

### ► 3.2.4 Pricing practices in mobile and wireless communication networks

Pricing in wireless networks has lived through different phases and has been heavily influenced by market regulation and the level of competition. We provide a brief overview of the history of pricing in mobile cell networks and wireless data networks and discuss the main pricing models currently used by the mobile and wireless industry. Since pricing in wireless data networks has followed a considerably different evolutionary path, we divide this section into two parts; the first part discusses pricing practices in mobile voice networks, while the second part elaborates on pricing in wireless data networks. In both parts we will concentrate on access-pricing practices and will exclude the discussion of interconnection pricing and price regulation in the mobile communication industry.

#### Pricing practices in mobile telecommunication networks

The mobile telecommunication industry can look back on an interesting history of different pricing practices, depending on the intensity of competition, product innovation, and customer behaviour (Gruber, 2005). In the first stage of mobile telecommunications, which began during the early 1980s, the typical mobile customers were business users with high usage levels and a high willingness-to-pay. This type of customer behaviour was reflected in the tariffing plans of that time, consisting of high connection charges and high monthly access fees. Prices were designed more with the intention of extracting monopoly rents than oriented to costs because of the limited competition in the mobile voice market.

With the introduction of digital cellular standards and intensifying competition, operators started to explore the potential of non-business customers, which were attracted by alternative tariffing schemes (Gruber, 2005). Such schemes were characterised by lower access fees at higher usage charges to match the usage patterns of residential customers.

However, while operating costs per terminal in mobile networks declined relatively quickly and approached the same level as the cost per line in fixed networks, consumer prices remained on a high level.

From the mid 1990s the market for mobile voice services became a commodity with further declining prices for access and usage. With the widespread introduction of second-generation digital standards, of which GSM was by far the most successful worldwide, sufficient capacity was available to supply a large customer base and to further reduce charges to increase demand. Tariffing models were creatively extended to accommodate the special usage patterns of prospective customers(Ahonen et al., 2004). One of the most successful strategies was to bind customers in long-term contracts and to substantially subsidise handsets to attract new customers. Another important pricing innovation was pre-paid tariffing plans, in which customers buy a fixed amount of voice minutes for later usage. Through pre-paid contracts providers were able to stabilise their revenue flow and to increase the value of their networks (positive externality) by adding more active users to their networks. In many economies the number of pre-paid customers has by far outreached the number of post-paid contracts because of better control over expenditures and the possibility of being connected to a mobile network even without having credit.

With more and more companies entering the market for mobile voice services the full range of marketing tools was used to attract and retain customers. Customer loyalty was driven by price-related and non-price-related elements such as improved geographical coverage or improved voice quality. Another marketing element, which started to get increasingly popular during this time, was the so-called service bundles. Service bundles consist of several single services, which are priced in combination so that single prices can not be separated with the added effect that customers started using services, which they would never have tried without bundling (Ahonen et al., 2004).

Introducing data services to mobile voice networks has created new challenges for the pricing design and tariffing structures of mobile services. While flat-rate pricing models for data had been widely established in fixed networks, the capacity constraints of wireless voice networks prohibited the introduction of flat fees for second-generation wireless networks (Koutsopoulou et al., 2004). To prevent network congestion and to maintain the existing service level for voice services, network providers decided to introduce time-based pricing schemes with rates only slightly lower than those of voice services. The introduction of 2.5G extensions, such as GPRS and EDGE, enabled fully transparent IP-services over wireless cellular networks with a constant connectivity to the Internet. This has facilitated the introduction of usage-based pricing mechanisms based on the network traffic instead of time-based schemes.

### Pricing in wireless data networks

Pricing in wireless data networks (*Wireless Local Area Networks* (WLAN)) has followed a different evolutionary path compared to pricing in mobile cellular networks. Instead of providing coverage over large geographical areas, wireless data networks, of which the IEEE802.11 standard is the most successful, were developed as cable replacement and for connecting portable computing devices. WLANs were mainly used in closed user environments such as companies or universities. Pricing with the objective of cost recovery was not in the focus of such deployments.

With increasing technological maturity and wider acceptance among users, business models emerged to implement wireless data networks in public locations such as hotels and airports to supply a wider customer base with wireless connectivity to the Internet. Opening up such networks to the public and commercially offering network services meant that suitable tariffing models needed to be developed, together with a billing and accounting infrastructure for handling the monetary transfers. In contrast to the limited competition in mobile voice networks, providers of wireless data services were often local companies, supplying multiple sites in a geographically limited area.

Since usage patterns are different to mobile voice communication and coverage of WLANs is very limited, operators regularly implemented access-based, time-based pricing models instead of requiring long-term contractual agreements with fixed tariffing schemes. Wireless data users could decide on-the-go which network to connect to and which tariff structure to use. High-usage customers often had the possibility to buy pre-paid time-units, which can be used at multiple visits to the same network. Also, subscription-based models have emerged, giving users unlimited access to the wireless network. In contrast to mobile voice communication, where a user expects that the service is available for the duration of the voice call (and handovers between cells being managed transparently without user interaction), resources in wireless data networks are sold on a best-effort basis. Since network performance in WLAN is strongly depending on the number of users concurrently being connected, applications with stringent Quality-of-Service requirements may not perform well in times of network congestion. Also, service quality cannot be guaranteed once an application has started with the transmission.

With an increasing diffusion of wireless access points large mobile voice providers started to show interest in the technology. The interest was motivated by the opportunity to offer complementary services to their customers as well as by the concern of losing important market share of data services. Most of today's thousands of "hot-spots" are now operated by the large mobile operators. In addition, regional WLAN providers supply metropolitan areas such as hotels, airports and train stations with wireless data services.

With the new industry structure in wireless data networks pricing practices have significantly changed. Large mobile operators often bundle WLAN services with their mobile voice and data offerings and include free connection time in the subscription plans. In some cases, WLAN access is not even available as a separate service offering. Alternative providers still offer more flexible pricing schemes with which users can connect on demand. Due to larger geographical coverage it has become common to sell pre-paid services, for example in the form of a scratch card, which can be used with any wireless network belonging to the same provider.

Another development, which has only recently found commercial application, is the deployment of Wireless *Metropolitan Area Networks* (WMANs) with the aim of substituting high-speed wired connections such as DSL and Cable. With relatively low implementation cost, comparable data rates to fixed access options, and quick user installation, WMANs circumvent the local loop, which in many countries, has not yet been fully deregulated. Additionally, such networks can offer portability and limited mobility, which opens up new ways to use such services.

Tariffing plans in such networks closely follow the approach used in fixed broadband networks by offering flat-rate tariffs with limited or unlimited capacity. As with WLANs, capacity is sold on a best-effort basis, which does not allow the stable usage of quality-sensitive applications in times of congestion. Since the air interface of such networks is highly volatile compared to a wired connection, degradation in service quality can be observed more frequently and is often influenced by factors such as weather conditions and large moving objects.

### Latest developments with an influence in pricing practices

With further technological advancement and the introduction of third-generation mobile networks, the potential of mobile data communication has been vastly extended. The *third generation* (3G) and some recently released 3.5G standards multiply the available capacities by using more efficient multiplexing and transmission schemes. While such enhancements will have limited consequences for the pricing of voice services, charges for data services are expected to significantly decline. With an increasing diffusion of 3G technology, innovative pricing models for data services can be observed. Some providers have introduced flat-rate pricing for best-effort data services or offer fixed rates for a certain amount of usage per month. Applications such as "push e-mail"[4] can successfully exploit spare capacity of the networks by using the lowest-priority class.

With the extended capabilities of end-devices such as smart phones and *personal digital assistants* (PDAs), which can handle bandwidth-intensive multimedia applications, the bandwidth demand is expected to grow with even greater speed than the available capacity in wireless networks (Gruber, 2005). Also, such applications will require higher quality guarantees to successfully run over mobile networks, which may not have been required in fixed networks due to higher transmission reliability and larger capacities. The question remains what pricing models will suit such applications and if flat-rate models will also be sustainable with better-than-best-effort traffic classes.

One obvious area of concern for mobile operators is the development of mobile IP-based communication such as Voice-over-IP-over-Wireless. Since 80-85% of revenues are still generated by voice services it is essential for operators to protect their main source

[4]Push e-mail is a technology that provides a copy of your email to your cell-phone or PDA. It is a subscription service that the e-mail provider has to implement.

of revenue and to block such services over their own networks for as long as possible (Osborn, 2005).

With the introduction of high-speed IP-based extensions such as HSDPA and EV-DV, networks become capable of providing an acceptable quality for time-sensitive applications. The pricing strategies for such offers remain open but are expected to be similar to todays tariffing schemes. Especially the question if differentiated services will be offered and at what price remains an unsolved question to be answered by the large mobile operators.

### ▶ 3.2.5 Feasibility of network pricing

When developing a new pricing scheme for differentiated network transport services, various criteria need to be considered to assure that the the scheme is feasible from multiple perspectives. Reichl et al. (2001b) identify three main categories, *network efficiency*, *user acceptance* and *technological feasibility*. The requirements from those categories are usually conflicting and the designer of the pricing mechanism needs to understand the priorities of the different requirements depending on the particular intention of the new pricing scheme. Reichl et al. (2001b) therefore call the trade-off between the requirements the "NUT Trilemma" (Figure 3.2).

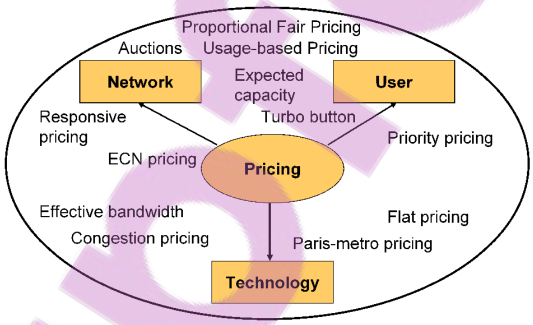



**Figure 3.2:** The NUT Trilemma. Source: Reichl et al. (2001b)

Network efficiency relates to the economical side of pricing. It asks the question if the attained allocation distributes resources so as to maximise social welfare. Another question is how fast the pricing mechanisms can be adapted to react to new changes in the network. For example, time-of-day pricing, as used in many wireless tariffing schemes, reaches an asymptotically efficient outcome, while being simple from an engineering and user point of view (Hayel et al., 2005). Other pricing mechanisms, which adjust prices on shorter time-scales are able to reach higher efficiency but may score lower in other categories due to technological complexity.

The second requirement category used by Reichl et al. (2001b) is user acceptance. Several empirical studies (see, for example, INDEX (Chu and Altmann, 1999), CATI (Stiller et al., 1999), and M3I (M3I, 2001)) have shown that from a user perspective, transparency and predictability of charges are the most important criteria for the acceptance of a new pricing scheme. These requirements do not automatically speak against dynamic pricing schemes per se. Courcoubetis and Weber (2003) argue that the complexity of such pricing algorithms can be shaded by intelligent end-devices, which shield the charging complexity and take over the task of adjusting demand on behalf of the user.

The last category relates to technological feasibility of the pricing scheme. Feasibility in this context relates to the existence of tools for technical accounting. Only if charges can be accurately captured by the accounting system, is it possible to implement a pricing scheme in a commercial setting. Since technical conditions vary from technology to technology, pricing schemes need to consider whether the underlying network structures provide sufficient support. Another question connected with technological feasibility is the costs connected with data collection. Usually it is possible to collect usage data or congestion data on the network management level. However, the cost for storing or processing such data may prohibit widespread use.

Reichl et al. (2001b) argue that not all requirements have the same priority when designing a new pricing scheme. Especially, technical feasibility is a precondition for establishing a new pricing mechanism. Therefore, technical feasibility is seen as a hard criterion, whereas designers of the pricing mechanism may be able to balance between the two other categories.

## ► 3.3 A Classification Framework for Pricing of Network Transport Services

In this section we introduce a classification framework for pricing of wireless network transport services, which uses the pricing unit and time-scale of the pricing approach as categorisation criteria. Before we introduce our framework we briefly review the existing classification frameworks which have been developed in fixed networks.

### ► 3.3.1 Classification categories for network pricing

Pricing and charging in communication networks can be classified by various means, depending on the purpose. The following, possibly incomplete list of categories has been collected from the literature.

- *Flat pricing versus usage pricing versus congestion pricing:* With flat pricing, network users are only charged a flat fee, which is independent from the actual resource consumption. In contrast, usage pricing refers to a charging scheme in which customers are charged according to resource consumption. Resource consumption can either be measured in time or volume or a combination of both. Congestion pricing takes into account the actual state of the network. In times of low resource congestion charges become zero while during times of high congestion, users are charged according to their contribution to the congestion (Reichl et al., 2001a). All three pricing models can be combined to form complex tariffing schemes.

- *Static pricing versus dynamic pricing:* With static pricing, charges are predetermined before service requests are made to the network (Courcoubetis and Weber, 2003). With dynamic pricing, some or all tariffing elements are determined at the time of the actual resource consumption.

- *Connection-oriented versus connection-less pricing:* Connection-oriented pricing is concerned with a flow of information instead of the single information packets. With connection-less pricing each packet is handled separately and no link is made between packets belonging to one data source or sink.

- *Access-pricing versus end-to-end pricing:* Access-pricing is concerned with price setting in a specific link. With wireless networks the designer is usually concerned only with the radio link as the bottleneck. Many studies assume that the capacity in the core network is sufficiently large and that providers have long-term interconnection agreements (for an extended discussion see, e.g., Maillé and Tuffin (2006)). With end-to-end pricing, a the charge is the sum of all sub-charges from each single link used by the network flow.

- *Centralised price setting versus distributed pricing:* With centralised price setting the price is defined by a central authority. An alternative approach is the use of a market institution such as auctions to let users reveal part of their private information in a bidding process.

### ▸ 3.3.2 Alternative classification approaches

In the following we review some of the relevant classification frameworks proposed in the literature.

Wang (2006) presents a pricing classification framework, which describes how pricing models can complement engineering functions in wired networks on different time-scales (Figure 3.3). Three pricing functions are distinguished according to the time-scale they operate in. On the highest, most strategic level, pricing complements *multi-period network capacity planning function*, which has been traditionally treated as a cost minimisation problem in which demand forecasts are taken as input parameters. By means of optimisation the equipment combinations with the lowest net present value can be determined which exactly fulfill the forecasted future demand. Additionally, a provider needs to consider the order of investments and if a delay of investment may lower infrastructure costs due to falling prices. Pricing, in such models, plays a central role for the overall strategy of a provider for profit maximisation.

On the medium level offline pricing schemes complement the *offline resource management function*. Based on the existing network infrastructure, offline resource management selects the topology of explicit paths for different services and source-destination pairs as well as allocates capacity to the single links. Pricing on this level has become an integral

| Time-Scale | Relevance | Typical Actions | Units |
|---|---|---|---|
| Atomic | Communication | Packets, roundtrip times | ms |
| Short-term | Applications | FTP, IP phone call | sec/min |
| Medium-term | Billing | Phone bills, rents | weeks |
| Long-term | Contracts | ISP-customer contracts | year |

**Table 3.1:** Relevant time-scales for tariff schemes. Adapted from Reichl and Stiller (2001)

part for the joint optimisation of revenue and traffic engineering and for load balancing so that a provider balances the risk of over- and under-usage of link capacity in different parts of the network.

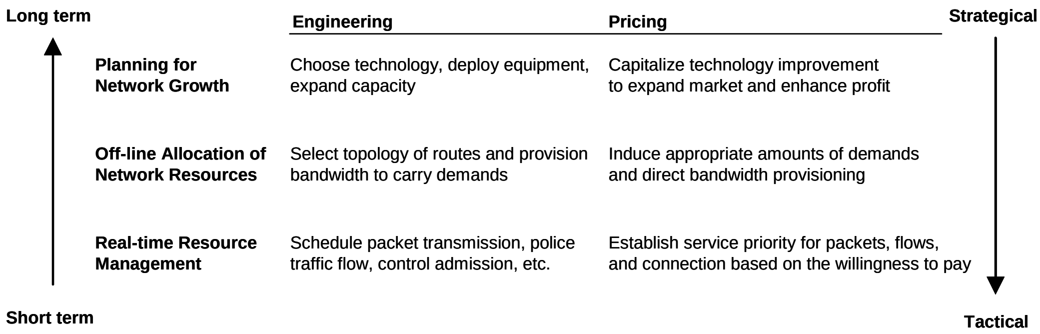



**Figure 3.3:** Combining pricing and engineering at different levels. Source: Wang (2006).

On the lowest level, pricing complements the *real-time* or *on-line resource management*, which controls traffic flows in the network. On-line resource management is further divided into four sub levels. Pricing at the *physical level* includes approaches which assign prices to the physical units of the wireless channel such as transmission power or time slots. Many approaches also make use of the fact that network services can be provided by different combinations of resources such as buffer and bandwidth. Pricing on the *IP-packet transmission level* denotes pricing approaches which associate a price with each individual packet. Pricing and *flow control* formulates the pricing problem as a function of end-to-end data rate. Finally, pricing on the *admission level* is described as a method of controlling Quality-of-Service on a *session level* by only admitting new requests if sufficient resources are available.

Reichl and Stiller (2001) present a classification, which is called the "time-scale methodology". It defines four different pricing time-scales, which are given in Table 3.1. The *atomic view* relates to the packet level on which each packet is marked with a price to obtain a priority for scheduling. On the short-term level Reichl and Stiller (2001) see the *pricing of applications and sessions*. Medium-term pricing relates to the *billing cycle*, which is usually done on a monthly basis. *Subscription contracts* are sorted into the long-term pricing time scale.

The model is used in conjunction with providing an integrated framework for the feasibility of pricing schemes, which has been presented in Section 3.2.5. It shows how the different requirement categories match with the time-scales defined by the time-scale methodology.

Fulp and Reeves (2002) analyse the consequences of the pricing time-scale on provider profits, allocation efficiency and feasibility of the pricing scheme. While the paper does not provide an explicit classification framework for sorting existing pricing approaches it provides a valuable analysis of the effect of using different pricing time-scales. On the one extreme, when prices stay fixed for long periods, such as weeks or months, a provider has no influence on controlling congestion and user demand. Also, price predictability from a user perspective is optimal. On the other extreme, prices can be adjusted on a very small time-scale, which allows the provider to quickly reach equilibrium between supply and demand. However, frequent price changes may be difficult for users to comprehend. The authors propose a time-of-day pricing approach to balance between these extremes.

### ► 3.3.3 The wireless pricing time-scale classification framework

In this section we present the time-scale classification framework for pricing models in wireless networks. We distinguish five different time-scales: *subscription*, *admission*, *flow*, *packet*, and *physical channel*. Furthermore, we define an additional category for pricing models spanning multiple categories, by, for example, combining flow control and admission control through a common pricing scheme. Figure 3.4 provides an overview and brief description of each level.

| | Description | Motivation |
|---|---|---|
| **Subscription** | Design of price plans to let customers switch from competing providers in long-term | Provider revenue |
| **Admission** | Prices dynamically formed at the time of request depending on situation in network | Provider revenue |
| **Flow/session** | Pricing of bandwidth shares to manage congestion and maximise allocation efficiency | Economic efficiency |
| **Packet** | Price attached with each packet to set priority in times of congestion (e.g., smart market) | Technical and economic efficiency |
| **Physical unit** | Use of pricing to allocate radio resources such as power or transmission slots | Technical efficiency (Engineering) |

**Figure 3.4:** The pricing time-scale framework.

In the following we provide a short description of the main characteristics of each category, give some examples how to apply pricing, and discuss the advantages and potential disadvantages.

### The physical channel time-scale

The lowest level of the time-scale pricing framework is given by physical layer, in which Radio Resource Management (RRM) is responsible for an efficient use of the available radio spectrum and the limited transmission power of the end devices. The objective of pricing, which usually complements the engineering functions such as power control, is to optimise resource usage from a economical perspective by including user utility as an additional measure of efficiency.

The increasing diffusion of Code Division Multiple Access (CDMA) based wireless systems has produced a high demand for sophisticated power allocation mechanisms. Since a common frequency spectrum is shared by all active users, efficient power control mechanisms have become an important area of research to improve the transmission efficiency. The use of economic models has been seen as a promising way to strengthen a customer-oriented perspective by using utility-based measures for power allocation. Since power can, in parts, be treated as any other economic good, the same methodological tools can be applied in a similar fashion. Since users do not directly derive value from transmission power, but from data transmission, many models build a relationship between the signal-to-interference ratio (SIR) on the physical level to measures on higher layers such as the achievable bit-error-ratio (BER).

Very different economic approaches have been transfered to the domain of power control. Since users do not usually cooperate and are expected to behave selfishly to maximise their utility, the situation can be modelled as a non-cooperative game between network users. Alternatively, cooperative games, for example between the different cells of mobile network, can help to optimise the downlink power of the base station belonging to the same provider.

In most approaches, pricing is decoupled from the actual charging process but is used solely to transfer the principle concepts to RRM. In principle, such approaches will not prohibit users from cheating on their actual utility since the incentives are not directly coupled with the payment function. However, since the control functions on the power control level are usually not accessible for the network user, there is only a small danger of manipulation.

### The packet time-scale

With packet-based pricing, charges are applied to each packet transmitted over the network. Each packet may be marked with a ”price tag”, defining the willingness-to-pay of the data source. Each single packet is treated individually without establishing an explicit connection between transmitted packages.

Packet-based pricing is closely related to packet scheduling performed by routers in the network. If multiple users simultaneously transmit packages, the packet scheduler managing the forwarding of packets needs to decide, which packets to buffer and which to forward. The traditional scheduling mechanism of the Internet is *First-In-First-Out*

(FIFO).[5]. With FIFO, the packet longest in the queue is given highest priority for the next available slot. If the buffer becomes full, the newly arriving packets are discarded. Many alternative scheduling schemes have been developed to increase technical efficiency and to allow for differentiated services. For example, with priority scheduling, separate queues for different service categories are introduced, with higher priority queues being served first (Plasser et al., 2002).

In wireless networks, scheduling is often coupled with lower network management functions to increase the efficiency of the wireless channel. If the scheduler is aware of the current channel quality, it can rearrange the priorities of the packets in the queue to improve the overall throughput. When the wireless channel is in a burst error state, most retransmission attempts fail, thereby causing poor utilization of the wireless channel. If the scheduler is aware of this, it can rearrange packets and first serve users with a low error state while queuing packets from users with bad channel for subsequent transmission (Shakkottai and Srikant, 1999).

Pricing schemes on the packet layer can complement the technical scheduling function in wireless networks. Prices serve as signal to sort packets according the utility a user receives from transmitting such packets. As long as routers are non-congested, they will forward all packets with some defined scheduling algorithm. When the path becomes congested, the pricing information contained in each packet serves as a prioritisation signal for the router.

While interesting as a concept, not many packet-based pricing models have been developed. The main criticism is the technical feasibility in a large, distributed network structure such as the Internet. Another issue relates to the implementation of suitable accounting schemes, which keep track of the number of successful packets submitted by each network node.

### The flow time-scale

Flow-based pricing refers to pricing models which define a charge for a connection-oriented traffic flow. A flow is defined as a set of packets related to an instance of some session (Roberts, 2004). Network flows can have different characteristics. Depending on the application type and the coding schemes used, flows may either generate a *constant bitrate* (CBR) or may produce bursty traffic with *variable bitrate* (VBR). While for CBR traffic the flow can be characterised by the guaranteed available bandwidth, VBR traffic can define minimum and maximum bandwidth requirements.

With flow-based pricing, the utility of users can be modelled as a function of the end-to-end data rate and pricing can be used to control the data flow (Wang, 2006). When users are flexible in their demand they may be able to reduce the flow size if demand increases. Users with fixed demand may try to keep the committed flow size or may terminate their service.

[5]This is often described as First-Come First-Serve (FCFS) scheduling

With static pricing, charges may be defined by some kind of pricing function, which is known to the users. By introducing dynamic pricing, charges can be adapted depending on the current demand in the network link. Then, flow-based pricing mechanism becomes a tool for managing and avoiding congestion in a network link and charges reflect the current congestion level users are imposing by choosing a certain data rate.

With the existence of a central price setter, users adapt their demand according to the charges imposed by the network. In contrast, providers may implement a distributed mechanism to let users reveal their valuation for the network resources. Price models using a distributed approach often make use of auction mechanisms to let users bid for resources describing the flow. Depending on the auction type the iterative bidding process may lead to a situation in which total user utility is maximised.

A potential disadvantage of flow-based pricing from a user perspective is the continuous price fluctuation and the difficulty of predicting the charges. As demand in wireless networks may change regularly, a user needs to update his input rate. Another potential disadvantage of flow-based pricing is the requirements on the charging signals that need to be recorded for the billing process. However, the billing requirements are less demanding as with packet-based pricing.

### The admission time-scale

Pricing models on the admission time-scale allow users to request prices at the actual time of demand. This type of pricing scheme fits applications which require guaranteed Quality-of-Service, for example, real-time voice and video (Wang, 2006). By using admission-based pricing, the pricing decision can be isolated from the instantaneous network states. A unifying characteristic of admission-based pricing is that the committed price stays valid for the entire duration of the session.

Pricing on admission level is closely related to the technical admission control function (see Section 3.2.3 for a short description of CAC). In addition to the technical decision to admit or reject a new service request, economic variables allow a provider to select customers according to their willingness-to-pay. In case demand exceeds supply, and the network link operates near its capacity constraint, a provider may set a price so that only users with sufficient utility are admitted.

Compared to the previously reviewed time-scales, pricing on the admission level has limited potential to implement distributed mechanisms for the definition of price levels based on demand. While some models assume that the network can pool admission decisions into batches, in which a couple of customers compete for access, it is usually difficult to implement distributed control schemes on admission time-scale. In consequence, many schemes implement a centralised authority setting the price for an access link. One problem with this approach is that the central authority needs to have extensive knowledge about the demand structure to fulfill the defined objectives such as revenue maximisation. Tâtonnement and learning mechanisms are common tools in this setting to improve the accuracy of the price setting process.

A practical example for admission-based pricing in communication networks is a "hot-spot" WLAN environment, in which the customer is regularly faced with different tariffing options once he connects to the network. Prices may vary with each connection request depending on the time-of-day or even on the demand situation in the WLAN.

One advantage of admission-based pricing is that it is relatively straightforward to understand from a user perspective. Some user groups may also accept varying admission prices as a chance to lower their usage-based costs by selecting times of lower demand for accessing the network. In addition, from a technical perspective, admission-based pricing is usually easy to implement. On the negative side, admission-based pricing is usually unable to achieve economic efficiency due to the problems of implementing distributed pricing mechanisms. Also, the ability of a provider to control demand is limited to the admission decision.

### The subscription time-scale

Under the subscription time-scale we can subsume all tariffing schemes in which prices are defined in terms of months and years. The vast majority of currently implemented tariffing schemes fall into this category. The most common tariffing schemes in wireless networks are flat-rate pricing and resource-based pricing. Voice service providers mostly apply usage-based pricing schemes based on the service duration, the destination, and the time-of-the-day. For data services, volume-based tariffs are common practice, which have been recently complemented by capped flat-rate plans.

A characteristic of tariffing schemes on subscription time-scale is the definition of prices at contract closure. Price changes, if at all possible, can only be realised by revising the contractual agreement between the customer and the provider. Subscription contracts are usually closed over a period of months to years. Many agreements for cell phone contracts run over a defined period of 12 or 24 months to bind the customer to the particular provider and to subsidise a suitable end device.

Pricing on the subscription level mainly serves the purpose of cost recovery and profit maximisation. Since revenues on this level are highly predictable, a provider can calculate the subscriber numbers to break-even network investments in a defined time period. According to such analyses it can also decide on the level of coverage and the capacity in each area to cover peak demand.

Besides the main objectives of cost recovery and profit maximisation, subscription-based pricing can also be used for traffic shaping on a very coarse level. The adjustment of prices emerges in an iterative manner, such as a tâtonnement process (Courcoubetis and Weber, 2003). When a supplier posts its prices user demand adjusts to the new setting. The provider may then reconsider his tariffing plans, which leads to a further adjustment.

While pricing plans in the residential customer segment are usually equal for all customers, individual price negotiation has become a common tool in the business customer segment. Providers usually offer individual discounts for certain network services when customers demand larger volumes. This form of price adaption can be seen as a form of dynamic pricing at which prices can be individually negotiated at the time of subscription. However, prices are static over the contractual period.

The main advantage of subscription-based pricing from a provider perspective is the simplicity of the charging and billing process. Uncapped flat-rate tariffing schemes do not require any measurement on the network management level. For volume-based or usage-based tariffing schemes, a provider has to keep track of the resources that have been consumed by the customer over the billing period. Disadvantages of subscription-based pricing are the inflexibility to react to rapidly changing demand and to influence demand on a lower time-scale such as hours or minutes.

## ▶ 3.4 Literature Review

In this section we provide a detailed review of the relevant literature on dynamic pricing in wireless communication networks. Because the large volume of the existing material prohibits a complete review we select the studies most closely related to our work. For an overview of general network pricing models we refer the reader to Reichl et al. (1999b), DaSilva (2000), Falkner et al. (2000) and a recently published book chapter by Wang (Wang, 2006).

The focus of this literature review is on dynamic pricing models that provide specific support for wireless networks. In addition, we review studies that focus on pricing in multiple-access network structures in which competing providers attract customers by setting prices accordingly. We use the time-scale framework developed in the previous section as our classification guideline. In addition to work about wireless network pricing we also review selected material from fixed networks, which are relevant to our context.

### ▶ 3.4.1 The physical channel time-scale

The vast majority of research contributions which are specific to wireless networks use the physical channel time-scale as the base for pricing. Table 3.2 provides an overview of all research papers, which have been identified to fall into this category. In the following we review the papers most relevant to our research.

**Table 3.2:** Studies dealing with network pricing in wireless networks on the physical channel time-scale.

| Author(s) | Title | Traffic Direction | Network Technology | Number of Considered Cells | Service Type | Resource Unit | Pricing Mechanism | Characterisation of Prices | Optimisation Criteria |
|---|---|---|---|---|---|---|---|---|---|
| Falomari et al. (1998) | A new framework for power control In wireless data networks: Games, utility, and pricing | uplink | CDMA | single cell | elastic data | power | non-cooperative game model | linear pricing function with transmitted power | total utility |
| Ji and Huang (1998) | Non-cooperative uplink power control in cellular radio systems | uplink | CDMA | single cell | elastic data | power | non-cooperative N-person game | shadow prices for power usage | technical efficiency |
| Shah et al. (1998) | Power control for wireless data based on utility and pricing | uplink | CDMA | single cell | elastic data | power | non-cooperative game model | linear price function | total utility, fairness |
| Saraydar et al. (1999) | Pareto efficiency of pricing-based power control in wireless data networks | uplink | CDMA | single cell | elastic data | power | non-cooperative game model | linear pricing function with transmitted power | total utility |
| Saraydar et al. (2000) | Power control in a multicell CDMA data system using pricing | uplink | CDMA | m cells | data | power | non-cooperative game model | global and local prices | technical efficiency |
| Liu et al. (2000) | Forward-link CDMA resource allocation based on pricing | uplink | CDMA | single cell | voice | power, codes | constrained maximisation | optimal price function | total utility, revenue |
| Goodman and Mandayam (2000) | Power control for wireless data | uplink | CDMA | single cell | data, voice | power | non-cooperative game model | linear price function | total utility |
| Feng et al. (2004) | Pricing and power control for joint user-centric and network-centric resource allocation | uplink | CDMA | single cell | data | power | Stackelberg game | central price setting | individual utility, revenue |
| Xiao et al. (2001) | Utility-based power control in cellular wireless systems | uplink | CDMA, TDMA | multiple cells | voice and data | power | central price setting | linear price function | technical efficiency, fairness |
| Heikkinen (2001) | Congestion based pricing in a dynamic wireless network | uplink | CDMA | single cell | elastic data | power | non-cooperative game model | congestion pricing | total utility, revenue |
| Heikkinen (2002) | On congestion pricing in a wireless network | uplink | CDMA | single cell | elastic data | power | non-cooperative game model | linear congestion pricing | total utility, revenue |
| Lee et al. (2002) | Downlink power allocation for multi-class CDMA wireless networks | downlink | CDMA | single cell | elastic data | power | non-cooperative game model | linear price function | total utility |
| Saraydar et al. (2001) | Pricing and power control in a multicell wireless data network | uplink | CDMA | m cells | data | power | non-cooperative game model | centralised price setting | total utility |
| Saraydar et al. (2002) | Efficient power control via pricing in wireless data networks | uplink | CDMA | single cell | elastic data | power | non-cooperative game model | linear price function | total utility, revenue |
| Alpcan et al. (2002) | CDMA uplink power control as a noncooperative game | uplink | CDMA | single cell | elastic data | power | non-cooperative game model | linear price function | total utility |
| Marbach and Berry (2002) | Downlink resource allocation and pricing for wireless networks | downlink | CDMA, TDMA | single cell | elastic data | time-slots, power | central pricing vs. auction | fixed pricing, discriminatory pricing | revenue, total utility |
| Siris (2002) | Resource control for elastic traffic in CDMA networks | both | CDMA, TDMA | single cell | elastic data | power, rate | iterative optimisation by pricing | shadow prices for resource usage | user satisfaction |
| Zhou and Honig (2001) | Two-cell utility-based resource allocation for a CDMA voice service | downlink | CDMA | two cells | voice (multiple classes) | power and code | optimisation model | central price setting | total utility |
| Zhou et al. (2002) | Utility-based resource allocation for wireless networks with mixed voice and data services | downlink | CDMA | two cells | data, voice | power, rate | constrained maximisation | shadow prices for resource usage | total utility |
| Zhou et al. (2002) | Forward-link resource allocation for a two-cell voice network with multiple service classes | uplink | CDMA | two cells | voice (multiple classes) | power and code | optimisation model | central price setting | individual utility, revenue |
| Zhou et al. (2003) | Forward-link resource allocation for a two-cell voice network with multiple service classes | downlink | CDMA | two cells | voice (multiple classes) | power and codes | constrained maximisation | shadow prices for resource usage | total utility, revenue |
| Zhou et al. (2004a) | Resource allocation based on pricing for wireless multimedia networks | downlink | CDMA | two cells | data, voice | power | constrained maximisation | shadow prices for resource usage | revenue, total utility |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Zhou et al. (2004b) | Two-cell power allocation for downlink CDMA | downlink | CDMA | two cells | voice | power and code | optimisation model | central price setting | total utility |
| Zhou et al. (2005) | Utility-based power control for a two-cell CDMA data network | downlink | CDMA | two cells | elastic data | power | constrained maximisation | shadow prices for resource usage | total utility, revenue |
| Basar and Alpcan (2003) | A hybrid systems model for power control in multicell wireless data networks | uplink | CDMA | m cells | elastic data | power | non-cooperative game model | differentiable price function set centrally | stability |
| Sun et al. (2003) | Wireless channel allocation using an auction algorithm | both | TDMA | single cell | elastic data | rate | all-pay auction format | identical prices paid by all users | fairness, total utility |
| Sun et al. (2004) | A novel auction algorithm for fair allocation of a wireless fading channel | both | TDMA | single cell | elastic data | rate | all-pay auction format | identical prices paid by all users | fairness, total utility |
| Dramitinos et al. (2004) | Auction-based resource reservation in 2.5/3G Networks | downlink | CDMA | single cell | constant bit rate | codes, frames | generalised vickrey auction | n/a | user satisfaction |
| Dramitinos et al. (2005) | Auction-based resource allocation in UMTS High Speed Downlink Packet Access (HSDPA) | downlink | CDMA | single cell | constant bit rate | codes, frames | generalised vickrey auction | n/a | user satisfaction |
| Huang et al. (2004) | Auction-based spectrum sharing | downlink | CDMA | single cell | elastic data | SINR and power | Vickrey-Clarke-Groves (VCG) auction | second-highest price paid by winners | total utility, revenue |
| Liu et al. (2004) | Single-cell forward link power allocation using pricing in wireless networks | downlink | CDMA | single cell | voice | power and code | optimisation model | central price setting | individual utility, revenue |
| Chen and Niu (2005) | Joint power and rate control for Wireless ad hoc networks: A game-theoretical approach with step-up pricing | uplink | CDMA | single cell | elastic data | power and rate | auction-like pricing game | uniform price in equilibrium | technical efficiency |
| Badia and Zorzi (2005) | Radio resource management with utility and pricing for wireless LAN hot-spots | both | CSMA/CA | single cell | data | rate | constrained maximisation | linear prices according to rate | revenue, total utility |

In an early paper Shah et al. (1998) develop a distributed power control framework for wireless data services using an economic approach. The distribution of power on the uplink is modelled as a non-cooperative game in which all users maximise their utility gained from the allocation of transmission power. Formally, the optimisation problem can be written as $max_{p_i} u_i(p_1, ..., p_N)$, for all $i = 1..N$, with $u_i$ being the utility obtained by user $i$ and $p_j$ the power level of user $j$. A user needs to consider the power allocated to other users when calculating his own optimal power level since each user is causing interference. The main objective of the study is to understand the influence of pricing on the overall efficiency of the allocation. In a first step, a specific utility function is developed, which allows the translation of the signal-to-interference ratio (SIR) to the bit-error-ratio (BER) on the IP level, depending on the channel coding technique and the cost of battery power needed for the transmission. Using this utility function each user can determine the optimal power level at which the individual utility is optimised. The authors show that the optimisation problem has a unique Nash equilibrium in which the received powers at the base station are equal for all users. However, it is also shown that this equilibrium is Pareto inefficient. In a second step the authors introduce a simple linear pricing function, which is proportional with the transmitted power of each user, to analyse the effects on allocation efficiency. The power control game can now be reformulated as $max_{p_i} u_i - F_i, \forall i = 1..N$, with $F_i$ being the price per power unit. It is shown by simulation that, using the pricing approach, an improvement in the overall efficiency can be achieved. One shortcoming of the approach is that it is not a dominant strategy for users to truthfully reveal their true valuation. Also, the model is limited to constant bitrate services. The model is further developed in subsequent papers such as (Saraydar et al., 2002), which elaborates on model extensions and additional factors included in the analysis. One particular concern of the study is to better understand the degree of inefficiency created in equilibrium.

Goodman and Mandayam (2000) present an uplink power-control scheme based on pricing for data services in CDMA networks. This highly educational paper introduces the reader to the basic properties of the wireless air interface in CDMA systems and contrasts the power allocation objectives of voice and data services. For voice-centric services the utility function can be modelled as a step function; after a SIR threshold value the utility a user derives from the network service remains constant. Therefore it is the most efficient solution to provide all terminals with the same target SIR level measured at the base station and to determine the minimal feasible power allocation for all users. In contrast, the utility function for data services is assumed to be an increasing concave function of the SIR level. Therefore, the same power allocation method does not apply for data users. This leads to the insight that a different power allocation method is needed to efficiently allocate power between data users. By introducing a pricing function for uplink transmission power the behaviour of mobile terminals for choosing the power level can be influenced. Instead of asking for the maximum feasible power level, users maximise their net utility as a difference between their utility for the received service and the price to be paid. If a linear pricing model is assumed and the pricing factor is increased from zero to positive values, the equilibrium shifts toward a point at which users attain lower SIR but higher utility due to reduced intra-cell interference. Additionally, the SIR levels at the base station are no longer equal for all users because users maximise their utility at different values of SIR. At this point all terminals operate at lower power, lower SIR levels, lower efficiency and higher utility than in the absence of a pricing function. An equilibrium price factor can be found at which the total utility in the system is maximised.

The paper by Alpcan et al. (2002) describes a price-based power control scheme for data users in a CDMA uplink. The model considers a single cell with $M$ users requesting network resources for the transmission between the mobile device and the base station. The problem is modelled as a non-cooperative game between $M$ users, which individually maximise their utility from receiving a share of transmission power they are allowed to use to transmit information. To define the problem to be solved by each user $i$ a cost function $J_i$ is defined with $J_i = P_i - U_i$; the utility function $U_i$ is chosen as a logarithmic function of the users' SIR received at the base station and the pricing function $P_i$ defines the instantaneous price $\lambda_i$ a user $i$ pays for the received channel gain. The price function in this model is assumed to be linear. The user's goal is to minimise his cost given the sum of powers from other users (which cause interference with his own signal) and his own nonnegative power $p_i \geq 0$. Based on the individual minimisation problem the optimal power allocation for all users can be identified. However, if the solution results in an infeasible power allocation, some users need to be excluded to bring the overall demand to a feasible level. In a second step the authors discuss two pricing strategies to be implemented by the provider. The centralised pricing scheme assumes that users with equal SIR requirements are divided into classes. The role of the base station is to set prices for all base stations so that the SIR targets of all users are met. In contrast, in the decentralised, market-based pricing model the base station sets a single price for all users and users choose the power level, which maximises their utility. It is shown for different update algorithms that the system converges to a stable Nash equilibrium. The authors propose to implement a simple admission scheme to limit the number of users in the system in order to achieve a feasible power allocation.

The study described in Marbach and Berry (2002) considers the downlink in a single cell of a wireless network with varying channel quality for different users and develops a pricing scheme for the allocation of power-limited radio resources. Two different types of wireless systems are described; a time-slotted system in which users compete for channel time and a CDMA system, which is limited in the overall power to be allocated between mobile terminals. In a first step the authors compare the two optimisation approaches, namely revenue versus social welfare. They conclude that optimising the allocation according to social welfare may be unrealistic for two reasons: first, it requires central knowledge of the users utility functions, which is usually private information to the users. Second, providers are usually not interested in the maximisation of utility but only in the revenue they can obtain. In the second step a price-driven auction mechanism for allocation transmission times is proposed. In this scheme, users bid in each frame by submitting a price bid and the base station allocates resources to maximise its revenue. After each round, bidders optimise their bids to maximise net benefits. In this setting providers search for the allocation strategy which maximises revenues in equilibrium. If providers have perfect knowledge about the users' demand functions $D_i$ they can derive such strategy by $\max \sum_{i=1}^{M} u_i D_i(u_i)$, with $u_i$ being the unit price charged to user $i$. In this case the optimal solution would allow a provider to skim all user surplus. Alternatively, the authors propose an allocation strategy for the case of imperfect information (no knowledge of the utility function). In this case the obtained revenue will be lower but converges to the optimum when many users are active in the system and demand is high compared to the available resources.

Zhou et al. (2002) present a price-based allocation approach for power allocation in the downlink of a CDMA cell with mixed voice and data services in a one-dimensional two-cell scenario.[6] The objective is to find the optimal power allocation for both voice and data users, which maximises the total utility per code over a cell subject to rate and power constraints in each cell. The utility of voice users is assumed to be a step function with a certain power threshold, after which transmission is possible; for data users the utility function depends on the received data rate and is therefore increasing and concave with the received SIR. With the proposed pricing scheme each base station announces a price per transmitted power unit $\alpha_p$ and a price per unit data rate $\alpha_r$. Three scenarios can be distinguished for identifying the optimal solution: the cell can be either rate-limited, power-limited or rate-power-limited. Which constraint is binding first is determined by the prices charged for the two resource components. In the model each voice user receives a constant data rate corresponding to the target SIR while data users are more flexible in their demand and can therefore be supplied with variable data rates. The optimisation problem consists of two objectives: first, to find a feasible power allocation for data users and and second to find a radius within which voice users are active. The optimal solution maximises summed utilities over the two-cell scenario. By simulation it is shown how the overall utility changes for different prices for power and data rates. It is also shown how resources can be shifted between user groups by changing the parameters of both utility functions for the two service types. In several subsequent papers ((Zhou et al., 2003), (Zhou et al., 2004a), (Zhou et al., 2005)) the authors further explore the basic model with changing points of focus and providing various extensions to the basic model. For example, in Zhou et al. (2003) the model is extended to multiple service classes, in which user groups receive different surplus from the same service. The paper also elaborates on

[6] The paper extends the ideas of Liu et al. (2000), who have established the basic model for voice users only.

the optimal solution for setting prices when a provider aims for revenue maximisation instead of efficiency maximisation in the economic sense.

Dramitinos et al. (2004) present a channel-based pricing mechanism for CDMA-based wireless networks using an auction. In the model each user chooses a utility function according to the service he is running. In contrast to other approaches, in this model the user can select from different allocation patterns in the cases where perfectly consistent allocation cannot be attained. For example, a user may have a preference for continuous transmission of smaller data units while another user prefers (or is indifferent about) larger data bursts. The predefined utility functions are matched with the existing service classes. The allocation is performed centrally at the base station by using a Generalized Vickrey Auction. The auction is held for each available code slot and bidding is performed automatically on behalf of the users. The experimental results show that users are either served with few violations of the requested scheduling scheme or are not served at all.

In a subsequent paper (Dramitinos et al., 2005), the authors adapt the model to the High-Speed Downlink Packet Access (HSDPA) extension of the WCDMA standard.[7] Each auction spans a time period of 1 sec and its outcome specifies the allocation of 500 2msec HSDPA frames. Again, different predefined utility functions have been defined, which allow users to set different priorities for the scheduling of the packets.

A pricing approach for a multi-cell wireless data network is presented in Saraydar et al. (2001). The system model is concerned with the optimal allocation of downlink resources for $N$ user terminals, which can be supplied by one of $K$ base stations. Users are randomly spread throughout the service area. The path gain of each terminal is determined by the distance between the base station and the terminal. Each base station allocates power autonomously without central control. The utility of data users is modelled as an increasing function of the received Quality-of-Service. Two base station assignment models are analysed in the paper: the assignment based on maximum received signal strength and the assignment based on maximum SIR. While the first model divides the optimisation problem into two steps (identifying the closest base station and maximising the sum of utilities per base station), in the latter the optimisation problem is two-diemnsional. In a first step the authors analyse the optimal solution without pricing for both scenarios. In a second step pricing is introduced in the model, charging a user for the share of powers consumed by the data transmission. Two pricing models are distinguished: global pricing denotes the situation in which all base stations charge the same price $c_j = c$. Local pricing, in contrast, lets each base station set its individual price based on the congestion level in the cell. For all models the authors prove the existence of a unique equilibrium solution. The authors show the consequences of introducing a pricing factor to the allocation of users to the base stations by using simulation. With pricing, cells with low congestion are

[7] One important change with HSDPA is the decrease of the frame size, which allows for lower delay through reduced retransmission.

able to take over some load of adjacent cells. They can also show that, in general, pricing increases the overall utility in equilibrium regardless of the base station assignment rule.

In Siris (2002) a resource control framework for elastic traffic in CDMA networks is presented. The author considers a single CDMA cell in which uplink and downlink resources are allocated separately based on the user's utility derived from the average data throughput. Two different utility functions are considered to be used by users depending on the service type; a concave utility function and a sigmoid utility function, which allows the setting of a minimum bandwidth requirement.

As a preparation for the core work the underlying traffic models for the uplink and downlink are developed. The constraint derived for the uplink is the available chips $W$ per time unit used in the channel. The constraint can be written as $\sum_i r_i \gamma_i < W$, with $r_i$ being the transmission rate of user $i$ and $\gamma_i$ being the target-bit-energy-to-noise-density ratio. In the downlink, the allocation is constrained by the power available to the base station ($\sum_i p_i \leq \bar{p}$). For the uplink, the price $\lambda$ is therefore in proportion to the resources used by the mobile terminal ($r_i \gamma_i$), while on the downlink, the price $\lambda$ depends on the allocated power $p_i$ to user $i$. One interesting observation is that in the reverse link the price is independent of the distance between the mobile station and the base station. This is because the uplink is interference-limited and and interference depends on the received power at the base station instead of the transmitted power at the mobile station.

The author first considers the uplink and shows that the optimisation problem can be decoupled into two problems: finding the optimal allocation of transmission power to minimise interference and identifying the price, which maximises the provider's revenue or the overall utility of all users. The author proves the existence and uniqueness of a solution to the described optimisation problem.

In the downlink the optimisation problem is limited to finding the optimal power allocation to maximise utility of all users given that a price $\lambda$ is charged per power unit. Unlike the case for the uplink, mobile users that are far from the base station incur a higher charge for the same rate.

Two different allocation mechanisms are presented, which differ in the way the equilibrium prices are identified. With the first, the base station announces a price for the resources after which mobile terminals adapt their transmission rate $r$. After several iterations the stable market price has been found under which supply equals demand. In the second allocation approach each user transmits his willingness-to-pay to the base station, which solves the optimisation problem centrally. With this approach only one step is needed to identify equilibrium allocation but it creates the problem of incentivising users to announce their true valuation.

### ▶ 3.4.2 The packet time-scale

Remarkably, the packet time-scale has not yet attracted many researchers developing specific pricing schemes for wireless networks. This is especially interesting since the correlation between channel quality and bit-error-rate (BER) has been well understood and a clear connection can therefore be made between the channel state and the throughput on the packet level.

Since in wireless networks packets are usually subdivided to fit in the specific frame structure of the underlying air interface, the notion of packet transmission in wireless networks for the purpose of pricing becomes somewhat indistinct. While in wireless data networks such frames can usually capture multiple IP-packets into one frame, frames in mobile networks based on CDMA are usually much smaller and one IP-packet needs to be divided into different subparts. In this scenario, only parts of packets need to be retransmitted if unrecoverable transmission errors occur.

One pricing proposal in this category has been presented by Musacchio and Walrand (2003), which develops a pricing scheme for a one-cell WiFi network. The interaction between the provider and the user is modelled as a dynamic game in which the players have asymmetric information and which proceeds in time-slots. Players choose between two service types: web browsing and file transfer. While web browsing users receive a continuous utility from using the service, file transferring only receive positive utility from finishing a file transfer. The authors show that the web browsing formulation leads to a constant price Nash equilibrium but that for file transfer users the outcome is inefficient. In a mixed model, and if file transfer length is bounded, and the probability of file transferring users is small, a Bayesian Nash equilibrium is reached.

For packet-based fixed networks, MacKie-Mason and Varian (1995) took a pioneering step by defining the concept of the *smart market*. It is probably the best-known pricing scheme for Internet pricing on the packet level. With the smart market, a user attaches a bid with each packet he transmits over the network, which expresses his valuation for the end-to-end transmission process. To motivate users to report truthfully, a second-price auction at each router is used to determine the price a user has to pay for the forwarding action. The price is determined by identifying the packet with the lowest clearing price, which is where the market gets ”smart”; this price equals the marginal cost of the externality caused by the congestion. If there is more than one link congested on the path from source to destination a user pays the highest threshold value that the packet passed through, which is called the market-clearing price.

The smart market is incentive compatible and efficient in the allocation of the scheduling slots. However, the concept has been criticised for various reasons. First, it can only guarantee relative priority but cannot give guarantees for absolute QoS since packets sent are not interrelated (Reichl et al., 1999a). Second, the model is not scalable for larger IP-based networks. This is because the auction mechanism would need to be implemented in each router of the network. Also, the additional bidding information represents a large overhead especially for packets with small message content. Third, the approach cannot guarantee network-wide stability.

### ▶ 3.4.3 The flow time-scale

Flow-based pricing approaches differ from packet-based models in the fact that not every packet needs to be priced separately. Instead, the user commits to a flow-based charge, which stays constant for a certain period of time. If conditions in the network change, unit prices may be adapted to reflect the change in demand. Flow-based pricing models have been popular in fixed networks because they can guarantee efficient resource allocation while at the same time producing less overhead than packet-based approaches. The idea has been transferred to resource allocation in wireless networks, often taking into account additional parameters that determine the signal quality.

Maillé (2004) proposes a modified second-price auction for the allocation of downlink power between users. The power allocation stays valid until at least one user changes his bid (or valuation) in the auction. The mechanism, called the *multi-bid auction*, has been developed in previous papers of the same authors (see Maillé and Tuffin (2004a) and Maillé and Tuffin (2006) for details) for the allocation of flows among users. The multi-bid auction is based on the progressive-second-price (PSP) auction developed by Semret (1999), which is incentive compatible, efficient, and leads to a stable allocation of resources in equilibrium. While in PSP the player submits a bid in the form $(p, q)$, where $p$ denotes the unit price, and $q$ the resource share, in multi-bid the user submits a vector of such bids. Those bids approximate his demand function (also called marginal valuation function). When the auctioneer has received all approximated demand functions, he centrally calculates the equilibrium allocation and the corresponding cost per user. Maillé uses transmission power as the object to be sold in the auction. He derives the correlation between the transmission power $p_i$ and the average throughput $a_i$ by using Shannon's Second Theorem, $a_i = r_i \times F(BER(\lambda_i))$, where $r_i$ is the original transmission rate, and $1/F(BER)$ represents the minimum redundancy factor to recover the message without any error; $\lambda_i$ is the signal-to-interference-to-noise-ratio (SINR). With the assumption that with an efficient solution all the transmission power is used, the maximum average throughput only depends on the power allocation $p_i$ for user $i$. With this model he uses the construct of the multi-bid auction to allocate power between mobile terminals. As in the fixed network scenario, users still value the average throughput of the flow but bid for the corresponding power level which maximises their net utility. It is clear that users far from the base station will gain less value from the same power level as a user who is located close to the base station. Therefore, users with the same valuation for average throughput but different distance to the base station will be supplied at different power levels. Maillé shows a simple example of the allocation process but does not further explore the topic by simulation.

In Shah and Nahrstedt (2004), a centralised price-based allocation scheme for time-slots in wireless LANs (IEEE 802.11) is presented. The main objective of the pricing mechanism is the maximisation of provider revenue and the match between supply and demand. In their model the authors consider a single WLAN subnet with one access point and a set of mobile terminals. Mobile terminals state their minimum and maximum bandwidth requirements and a maximum price for a defined share of the channel. The price stays valid for the entire time of the service.

At startup time the bandwidth manager situated at the access point defines a reserve price per slot share which covers the operating costs of the wireless link. If the total demand is smaller than the available time slots the reserve price times the allocated channel share is charged to the users. If congestion occurs the bandwidth manager allocates the slot proportions so that revenues are maximised. This is done by an iterative process which sorts all user requests according to the stated unit price per slot share and searches for the feasible allocation by repeatedly removing the user with the lowest unit price from the set of users served with their maximum demand. All users for which minimum requirements cannot be fulfilled at the current threshold price are deleted from the set of active users. This process is repeated until a feasible allocation has been found. All users pay a uniform price times the allocated channel share, which is determined by the lowest clearing price in equilibrium. The allocation process is restarted as soon as new users join the network or existing users become inactive. The authors test their allocation algorithm through simulation and compare it with a fixed-price model. They show that the algorithm performs significantly better when network demand is high. Furthermore, the model can guarantee some minimum bandwidth during congestion. The paper does not elaborate on an analytical model of the proposed pricing scheme nor does it test the consequences of alternative bidding strategies if users deviate from the proposed bidding strategy.

The aim of the paper by Soursos et al. (2003) is to extend the DiffServ framework to the wireless link in a General Packet Radio Service (GPRS) access link. To regulate the demand in such a system a flow-based pricing system is implemented, which incentivises users to choose the suitable service class for the intended service type.

The proposed model differentiates between three service classes denoted by $J$, which are Premium Service, Assured Service, and Best-Effort Service. Users in the system are modelled as price-takers, with no influence on the price or the service quality. Each user sends with a certain rate $x^i_j$ in either of the service classes. QoS is modelled as delay $d^i_j$, experienced by the user. It is assumed that the delay of the next-lower service class is influenced by the load in the higher service classes. Additionally, each user has a cost from the delay given by a function $\gamma i_j$. Given this model and a user utility function $u_i$, each user maximises

$$V_i(x_1^i, x_2^i, x_3^i) = u_i(x_1^i, x_2^i, x_3^i) - \gamma_1^i d_1 x_1^i - \gamma_2^i d_2 x_2^i - \gamma_3^i d_3 x_3^i.$$

With this function and the overall objective to maximise social welfare the socially optimal demands $\{x_1^{i*}, x_2^{i*}, x_3^{i*}\}$ are derived. To incentivise users to choose the welfare-maximising service combination the authors introduce differentiated prices per service class and determine that such prices are equal to the marginal delay cost suffered by all users in the service class. The main problem of the approach is the determination of the delay function. The authors propose a tâtonnement process in which the provider approximates the delay function through which prices can be updated.

### ▶ 3.4.4 The admission time-scale

On the admission level, pricing can serve as additional criteria for deciding on the admission of a particular request. In wireless networks admission control has mainly been used with voice services, where service guarantees and a low dropping ratio are essential. Increasingly, connection admission control is used in wireless networks to provide certain quality guarantees for packet-based connections. In the following we present a selection of admission-based pricing approaches, which correspond with the work done in Chapter 5.

An admission-based resource allocation model with different pricing schemes for W-CDMA networks is presented by Elayoubi et al. (2005). The focus of the study is on comparing different CAC models and on understanding the influence of of pricing. Two different service types, voice and data, are assumed. While voice users need to be supplied with a fixed SIR, the SIR value of data users can possibly be varied to adjust to the current market demand.

The authors present three different CAC models: *preventive CAC* assigns fixed resource reservations such as a constant SIR level for each admitted call. In this model resources are reserved most conservatively. *Measurement-based CAC* uses a prediction method for newly arriving requests together with measuring the traffic characteristics of the already admitted flows to decide on the admission. With the second model resources can be used more efficiently because multiplexing effects of the different traffic types can be considered. The most sophisticated CAC scheme, *the squeezing CAC approach*, makes use of the elasticity of data flows in resource management. An ongoing data connection may tolerate a certain downgrading to allow for the admission of additional users to the system.

In the next step the authors propose a utility function for data users, which is modelled as an increasing function of the SIR level and depends on the channel gain and the sending rate of the user. The data user uses his utility function to determine if the current received rate is still acceptable or if he will decide to drop the connection.

With the squeezing CAC approach, three cases at admission time can be distinguished. First, some users will be rejected because of insufficient resources. Second, some data users will reject the admission offer because their demand cannot be fully fulfilled and the offered price does not provide them with a positive surplus. Third, some users are accepted by the network and also accept the price offer. Additionally, it is possible that some data users quit the service upon experiencing a bad utility-price pair.

Two different pricing strategies are tested together with the CAC schemes: flat-rate pricing, in which the overall price for the resource stays the same even during times with service degradation, and resource-based pricing, where users only pay for the received resources. The authors explore the changes in provider revenue and connection blocking with the three admission schemes by using a simulation approach. As expected, the squeezing CAC approach performs best by generating the highest revenue at the lowest blocking rate. However, the authors argue that the measurement-based CAC scheme is to be preferred because of its simple implementation.

An alternative approach on the admission level, developed by Acemoglu et al. (2004), uses the channel gain of different users for identifying a feasible power allocation and a revenue-maximising admission price. The game is modelled as a Stackelberg game, in which the service provider is the lead player, announcing an admission price and a feasible power allocation and users are the followers, deciding to either join the network or stay inactive. The user's utility derived from joining the network is modelled to be dependent on the individual channel gain $g_i$ and background noise $\sigma$. The rate at which the base station transmits to user $i$, denoted by $x_i$, is given by $x_i = log\left\{1 + \frac{h_i p_i}{\sigma}\right\}$, where $p_i$ is the power allocated to user $i$. A user always joins the network if his utility is greater than the cost of joining. However, the received utility depends on the decision of the other users and the final power allocation. Therefore, a user needs to consider the behaviour of other users when making his decision. The maximisation problem to be solved by the provider is to set an access price $q$ and a feasible power allocation , which maximises overall revenues. The revenue maximising solution is found by maximising the utility of the so-called marginal user, which is indifferent of joining or staying inactive. The authors present the equilibrium solution of the dynamic game for different utility functions. Several issues are not touched on by the approach. For example, it remains open how long the power allocation stays valid. Also, the model assumes that the channel gain is static and does not change over time.

In Yaipairoj and Harmantzis (2006), an auction-based admission control scheme using dynamic pricing for GPRS-based mobile networks is presented. Since in GPRS, priority is given to voice calls, only a few channels are usually available for data traffic. In contrast to scheduling models, which operate on a lower level and allow efficiency improvements within a certain range of incoming traffic, the proposed scheme installs an admission-control mechanism to prevent new traffic from entering the network and to preserve certain Quality-of-Service commitments to existing users. The auction model used in the approach is a multi-unit second-price sealed bid auction (multi-unit Vickrey auction) with a reservation price $v_*$. Users trying to gain access to the network have to submit a bid together with the actual service request. Based on the generalisation of the Vickerey auction the $K$ highest bidders are admitted to the network and pay the value of the highest loosing bid or the reservation price, whatever is higher. The reservation price in the model serves two purposes: first, it allows network provider to set a minimum revenue, which covers for the operating costs of the network. Second, in time of congestion, the reservation price can be increased to $v_c$ so that the overall demand is reduced. The authors analyse the correlation between mean system delay and reservation price and present the objective function for determining $v_c$ if certain delay targets have been committed to users already admitted to the network. In several simulation experiments the authors show how the proposed admission-control scheme can improve the experienced mean system delay in a GPRS network depending on the defined reservation price. Furthermore, the trade-off between network performance and provider revenue is analysed by varying the reservation price. It is shown by simulation that there exists an optimal point which maximises revenues and keeps the blocking probability low.

Wang et al. (1997) have developed a pricing approach to maximise provider revenue for integrated-services in fixed networks, which has been summarised in Wang (2006) (see Figure 3.5). The model defines $N$ service classes with guaranteed QoS and one best-effort (BE) service class. All services are provided by a network with total capacity $C$. Users in the QoS classes are charged a time-based price $p_i(t)$. Users in the BE class are charged a variable per-packet price $p_b(t)$. In the QoS classes, the user arrival process is modelled by a Poisson process $\lambda(p_i(t), t)$, which is time-dependent and controllable by the price while the service duration is defined independent from the price level. The pricing problem is formulated as an optimal control model with several resource constraints, which ensure that the chosen allocation is feasible in terms if QoS, target blocking ratio and average throughput. The authors provide a general solution by using the price elasticities of demand in the various classes and by using the Lagrange multipliers as shadow prices for reserving and using bandwidth. Additionally, the Hamiltonian associated with the state variables of the model denotes the opportunity cost of accepting a QoS service at time $t$.

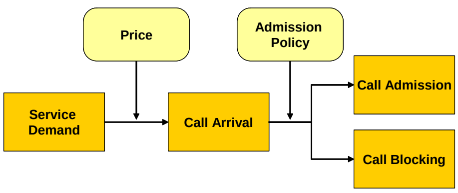



**Figure 3.5:** Call admission process for Guaranteed Services. Source: Wang et al. (1997).

While the analysis provides the general pricing principles, an explicit solution requires extensive knowledge of the network states and demand functions in dependence of the price $p_i(t)$. Therefore, the authors propose a three stage procedure to obtain a near-optimal solution by keeping certain variables fixed in each step. The first stage solves the optimal investment problem of choosing the optimal capacity. The second stage derives the optimal pricing policy for guaranteed services and the third stage develops the optimal spot price for BE services.

### ► 3.4.5 The subscription time-scale

Dynamic pricing models on the subscription time-scale are very limited. This may find its reason in the fact that such models are more interesting for providers themselves than for the open research community. For example, developing a sophisticated rebate and discount system for business customers may be very relevant for the strategy department of a large wireless provider targeting new customer segments. However, to develop suitable models and to understand the demand implications does require detailed knowledge of past usage data and user reactions to price changes. Since such material is not generally available outside the provider companies for competitive reasons, academic work limits its focus to general pricing models on the subscription time-scale. Since we could not identify any relevant work specific to wireless networks we present two approaches developed for fixed networks.

The only dynamic subscription-based pricing model identified in the literature is the Cumulus Pricing Scheme (CPS) (Reichl et al., 2001a). The approach aims at overcoming the "NUT trilemma" described in Section 3.2.5 by integrating all time-scales into one pricing scheme and by balancing between the technical, user and provider requirements.

The model defines three time-scales. On the long-term time-scale the user enters a contract with a service provider with a flat-rate tariffing scheme. However, this flat-rate is not defined as static but can be reviewed and adapted long-term (such as months or years) according to the user's behaviour. On the short-term time-scale the provider measures the customer's behaviour and collects a user pattern in the medium time-scale. This pattern is communicated back to the customer over so-called "Cumulus points". By overusing the network compared to the agreed terms a user collects red cumulus points whereas he earns green points for underusing the network resources. If a certain reaction threshold is reached the provider has the right to renegotiate the contractual terms and to change the subscription pricing. Figure 3.6 shows the basic principle. The actual usage is measured against some expected bandwidth requirement. In each billing period the customer receives feedback on his usage behaviour. When the threshold is reached the terms are renegotiated.

The model has the main advantage that user charges are predictable and only indirectly linked to the user actions. On the other hand the provider has the possibility to balance network load in long-term. On the other hand it does not allow to manage congestion nor does it enable the implementation of differentiated services in its basic form.

In a subsequent paper (Reichl et al., 2001b) the authors extend the Cumulus approach

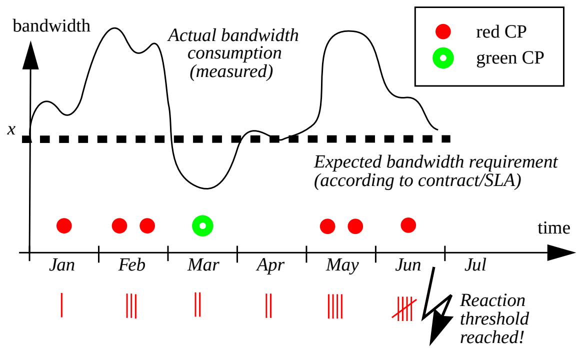



**Figure 3.6:** The Cumulus Pricing Scheme. Source: Reichl et al. (2001b)

to support multidimensional QoS parameters. Three alternative approaches are evaluated. With the first model a single parameter expressing the bandwidth and experienced QoS is assumed. In this model excessive use of resources leads to negative flags while bad QoS performance leads to positive flags to "compensate" the customer for service degradation. The second case treats bandwidth and quality separately and adds up flags from both dimensions into a common evaluation system. However, quality parameters are merged into a common indicator. The third case enables the monitoring of each quality parameter separately and allows the most detailed monitoring of the service quality at higher complexity. The presented approach can only guarantee relative QoS experienced by users over a certain timeperiod but is unable to manage QoS on the time-scale of actual congestion. It is instead assumed that the adaption of user behaviour will lead to stable network conditions in the long-term.

In a subsequent paper, the Cumulus pricing model is implemented in a simulation platform to understand the implications of varying the threshold value for collecting Cumulus points and to understand the properties in a practical setting (Ma et al., 2002). A simulation platform is presented with which consumption data can be collected and analysed on different time-scales.

In recent work Hayel and Tuffin (2005) analyse the mathematical properties of the Cumulus Pricing Scheme. The main motivation for this is to understand the optimisation problem posed by setting the threshold value. In a first step it is shown that the original definition of CPS does not incentivise a user to truthfully declare his expected traffic volume; instead, he usually tend to underestimate his consumption. Second, the authors slightly modify CPS by introducing a penalty fee for each negative Cumulus point collected, which allows a more direct penalisation of user behaviour. With this modification the authors can show that CPS is incentive compatible and motivates users to truthfully state their consumption.

The second part of the paper elaborates on the optimisation problem given by choosing the threshold value (the number of positive and negative cumulus points collected) in order to maximise provider revenue. By simulation, special cases are analysed and it is shown how the optimisation problem can be solved numerically.

Another study situated on the subscription time-scale is Odlyzko (1999). The authors compare two different tariffing strategies, namely flat-rate and usage-based pricing, when marginal production costs are close to zero and multiple firms compete for customers. In the model, firms are allowed to review their pricing at every period (for example, after each billing period) and to adapt prices in order to maximise their individual revenues.

Two different analytical models are presented. In the first model a customer simply acts as a cost minimiser based on the expected usage intensity in the next period. The second model introduces additional features: first, it models a budget constraint, which restricts the choice of a customer for a potential pricing scheme. Second, it allows for restricted usage of resources when the overall fees would exceed the budget constraint.

The main conclusion of the work is that in the absence of collusion, the direct price competition usually leads to ruinous price wars, which drive revenues close to zero. However, some stable equilibria could be identified that result in nonzero revenues for all firms. In those cases the revenue is well below the monopolist's revenues and the firm with the flat-rate pricing model does slightly better than the firm using a usage-based tariffing scheme.

The assumption that customers can switch in each period may often not be realistic depending on the duration of the intervals between price changes. For example, high switching costs may make the market more static with only a few customers switching to the more attractive offer in the next period.

### ► 3.4.6 Literature on pricing for wireless multiple access

The number of studies, which have their direct focus on the pricing in heterogeneous network environments with multiple-access, is limited (Ormond et al., 2005). While the concept of seamless handovers between heterogeneous wireless access points has drawn the attention of research from a technical viewpoint, the economical aspects have not yet been in the center of much in-depth work. One reason for this is that many researchers assume a cooperative relationship between different providers, which internalises the pricing problem to defining interconnection fees.

Table 3.3 provides an overview of the work. Due to the large variation of the covered topics, it is not possible to define general characterisation criteria. In the following we briefly review such papers.

In Ormond et al. (2005, 2006), an algorithm that selects the best available network for transferring non real-time data is presented. It is assumed that all available access points, which potentially use different wireless air interfaces, employ a fixed price-per-byte pricing scheme and that the transfer is done on a best-effort basis. In consequence, networks with available capacity will charge a higher price for the expected higher throughput while networks experiencing congestion will offer a lower price since throughput can be

| Author(s) | Title |
|---|---|
| Le Bodic et al. (2000) | Dynamic 3G network selection for increasing the competition in the mobile communications market |
| Azouzi et al. (2003) | Telecommunications network equilibrium with price and quality-of-service characteristics |
| Bircher and Braun (2004) | An agent-based architecture for service discovery and negotiations in wireless networks |
| Lin et al. (2005) | ARC: An integrated admission and rate control framework for competitive wireless CDMA data networks using non-cooperative games |
| Zhang (2005) | Bearer service allocation and pricing in heterogeneous wireless networks |
| Shin and Weiss (2005) | Optimal pricing for broadband wireless Internet access service |
| Ileri et al. (2005) | Demand responsive pricing and competitive spectrum allocation via a spectrum server |
| Ormond et al. (2005) | Network selection decision in wireless heterogeneous networks |
| Ormond et al. (2006) | Economic model for cost effective network selection strategy in service oriented heterogeneous wireless network environment |

**Table 3.3:** Overview of research with focus on a competitive multi-provider environment in network pricing.

expected to be smaller.

Mobile terminals with an upcoming data transfer may employ different network selection strategies to choose the network at the start of a transfer. For example, the terminal could use an *always-cheapest-strategy*, which selects the network with the lowest available price. However, the particular network may be overloaded and transfer times may become long. The authors therefore introduce a piece-wise linear user utility function to describe the trade-off between the user's time and budget limitations. Simulations are used to show the performance of the different strategies from a user's viewpoint.

Zhang (2005) presents a resource allocation model using pricing when multiple network technologies have been deployed. While he assumes that all networks belong to the same provider he also refers to a situation in which the network environments may be owned by competing carriers. The main research problem addressed in this paper is how to distribute new service requests among the different networks. Since each network technology has different efficiency for different service types, such as voice, real-time data or best-effort, efficiency can be increased by properly allocating bearer services in different networks. Figure 3.7 shows the minimum and maximum combined capacity in GSM and UMTS, assuming linear capacity regions.

In a second step, pricing is introduced to incentivise users to select the network yielding the highest efficiency in terms of resource consumption. However, it is assumed that the price charged for a service is identical in all networks. The result of the constrained maximisation problem indicates that network revenue is maximised when the marginal revenues of all services are the same.

In Cao et al. (2002), a game-theoretical analysis of a two-provider scenario is presented. The game is assumed to be non-cooperative and both providers set prices indepen-

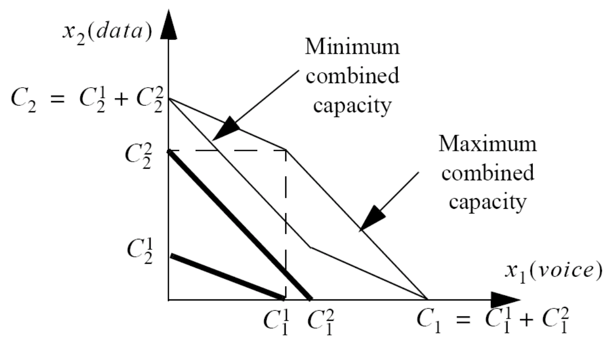



**Figure 3.7:** Minimum and maximum combined capacity in GSM and UMTS, assuming linear capacity regions. Source: Zhang (2005)

dently. Users request network services with certain Quality-of-Service requirements. Each provider $i$ offers a bandwidth $\mu_i$ at a price $c_i$ to the customer. Users, in turn decide for one of the service providers with probability $\alpha$ and $\beta$, respectively, in order to maximise their utility. The goal of the study is to find the Nash equilibrium points of the game in which both providers cannot change their prices without cooperation to improve their utility. Instead of an analytical discussion of the game the authors provide a numerical example, how the equilibrium points can be identified. In this example the equilibrium is stable but inefficient. It is also shown that the proposed approach is computationally feasible if the Quality-of-Service functions are discrete.

Das et al. (2004) illuminate provider competition from another interesting perspective. While they do not allow for direct competition between providers they include long-term customer churn in an admission control scheme to decide if a new user should be admitted to the network or not.[8] The consequences of the admission are modelled in two dimensions. First, a customer can be motivated to switch provider if he experiences frequent rejections by the network. Second, already admitted customers can be negatively affected by the admission of a new user and may decide to change provider. A two-player game with one provider and one customer is presented. The provider's options are to admit or reject a new customer, while the customer can either stay with the provider or leave him. With corresponding payoffs for each field in the $2x2$ matrix a provider can identify the utility maximising strategy to provide satisfactory service quality by controlling the admission to the network.

In Azouzi et al. (2003) a game-theoretic model of competing service providers is presented. In the game, service providers compete over prices and a Quality-of-Service (QoS) parameter, which they announce to the market. It is assumed that service providers do not own the network but have to buy resources from a network provider, which deter-

[8] The paper has been the base for the work presented in Lin et al. (2005). It provides a broader perspective on the implications of customer churn.

mines the minimal price on the market. It is further assumed that overall demand for one provider does not only depend on the price and quality level offered by one service provider, but is influenced by the decisions of all other providers about price and quality level. Two types of complete-information games are presented. In the first model the quality levels of all providers are fixed while prices can be set freely. In the second model both parameters can be set by the provider. It is shown for both cases that under some stringent conditions a unique Nash equilibrium exists.

In Le Bodic et al. (2000) a dynamic market for network services is developed in which providers can offer resources at a price and users select the offer best matching their preferences according to the price for the service and the reputation of the provider. Since in the mobile environment the available radio capacity is often highly variable, providers need to employ several methods for providing Quality-of-Service guarantees in order to fulfill their service contracts. The success rate of contract fulfillment depends on how aggressively providers load their networks. To provide customers with feedback, market agents measure the level of compliance and provide this information together with the price offered. The market mechanism used for selling resources is a sealed-bid, first-price auction in which one buyer receives offers from multiple sellers and decides for one according to a predefined decision procedure. The sealed-bid format has been selected because of its one-shot character (with only one auction round) since the admission decision is time critical and does not allow for a a multi-step bidding process. The market exchange has been developed as middleware handling the resource requests between software agents. Additional to the three roles of the main software agents, namely, user agents, network agents and service agents, the role of the market agent has been introduced to manage the negotiation process. The authors propose the formation of separate marketplaces for a certain geographical area in which the usage pattern is homogeneous. In simulations it is shown that for two network operators equipped with the same network and connection-admission strategies, the price offered by both operators is similar and reaches equilibrium. It is also shown how a third provider entering the market place evokes a permanent decrease in market prices.

The paper by Bircher and Braun (2004) develops a comprehensive market-place-based agent-based architecture for the negotiation of network resources. The focus of the paper is on the development of the communication setup and procedures between the different actors. The authors use the FIPA[9] contract net protocol for the core negotiation process. The entire setup involves service discovery, the service negotiation phase, and the application adaption according to the negotiated resources. Instead of creating one marketplace for a large geographical area the market is segmented by the different hot spot areas. This allows a user agent to contact such market places without previous contractual relationships. The proposed approach has been implemented in the JAVA

[9]Foundation for Intelligent Physical Agents

programming language and tested in a real environment using an agent-based middleware. Performance measurements show that procedures such as service discovery and service negotiation can be performed in less than one second despite the significant overhead introduced by the FIPA-OS platform. Despite delivering useful information about the implementation the paper does not elaborate further on the possible pricing strategies for a provider but instead assumes static pricing and no price updates according to user demand.

## ► 3.5 Selection criteria for in-depth analysis

After having introduced the time-scale framework and having explained the different layers of granularity for pricing in wireless networks, we now turn our attention to the definition of viable scenarios for competitive pricing. In principle, every layer could be interesting to look at and to understand the opportunities and potential problems when thinking about distributing resources under competition. In the follwing we very briefly discuss each layer and depict the main properties.

**The subscription layer** This level depicts today's situation in most mobile networks. Providers set prices on a long-term basis and try to maximize their market share by binding customers into long-term contracts. Customer churn is the main instrument of moving customers between providers. Usually, high switching costs exist to prohibit customers of doing so. In terms of pricing this layer is well understood from a practical perspective.

**The access layer** Customers decide to join a network on an access basis. This is the case with current wireless computer networks. If local competition exists customers may be able to compare prices and decide for the best offer. Theory about optimal provider behaviour in such a setting has not been strongly developed so far and thus, may be a worthwile field of study.

**The flow layer** On the flow layer multiple data sinks compete for gaining a share of resources to establish or maintain a data flow. The existing literature has a strong focus on developing models of optimizing resource distribution among entities under different optimality criteria such as technical efficiency (e.g., througput) of economic aspects (revenue, social welfare). Only a few studies so far have looked into the idea of competitive access. An interesting aspect on this layer is how entities behave when faced with the option of multiple access at a certain location and the effect on the overall economic situation in such a market.

**The packet layer** Pricing approaches on this level are usually strongly influenced by the smart-market model and it has been generally understood that the information overhead is very high. However, a scenario could be developed in which each user in a network decides ad-hoc over which network he prefers to send a certain data packet. It is arguable if competition on such low technical level contributes to economic efficiency aspects of resource distribution and would influence behaviour of the involved actors (sellers and customers). Thus, we discard a more detailed discussion of competitive pricing on this level.

**The physical layer** Most of the research work on wireless networks is focused on physical units of wireless resources and a extensive research body exists. Since resources at the physical level (such as frequency spectrum or transmission power) are usually statically assigned to providers, it seems to make limited sense of discussing competitive pricing on this level. However, there is an increasing trend that advocates for unlicensed allocation of spectrum. Unlicensed spectrum would allow several interested parties to share a given spectrum. Pricing can be used to signal current usage and user valuation and to efficiently allocate available resources on this level. On the other hand, pricing on the physical time-scale would need to be supported by a fine-grained charging approach. As pricing information needs to be exchanged for every allocated unit the information overhead is regarded as very large.

Because of the above arguments we have decided to conduct in analysis in two areas: user behaviour on the flow layer and provider behaviour on the access layer. While many other meaningful choices can be made without any doubt, we feel that we can provide a good cross cut through the field of research with this selection and that the two selected levels of granularity are particularly meaningful to the research community.

## ▸ 3.6 Chapter Summary

The objective of this chapter was to introduce the reader to the basic concepts and aspects of network pricing with a focus on wireless communication networks and to provide a comprehensive overview of the relevant research. We have presented the pricing time-scale framework, which describes various levels of granularity for pricing in wireless networks. We have used this framework to classify the existing work and to present selected studies most closely related to our work.

We have seen that one main focus of network pricing in wireless networks is on the modelling of the physical layer. This observation is consistent with the observation made in previous work (see e.g., Arabas et al. (2003, p. 154)). Pricing concepts, developed on this time-scale complement the technical radio resource management function by adding economic concepts to the network control function of power control and bandwidth allocation. The reason for this can be seen in the enormous challenges to control the random properties of the wireless air interface. Economic theory supplements this challenge by adding aspects of user utility, which can be used for deciding on the most effective resource allocation based on user preferences. For this reason, many studies concentrate on the efficiency aspects instead of looking into pricing with the purpose of cost recovery and profit maximisation. Prices are usually not intended to result in real charges imposed on the user but only serve as virtual prices to be used as a way to communicate network states and user valuations.

Since the wireless air interface differs in the forward and reverse direction, separate approaches have been developed for each direction. Often, studies concentrate on either direction to reduce the complexity of the solution. Because of the fast-changing channel conditions, central pricing is preferred on this time-scale, since it allows for fast convergence, whereas with distributed pricing mechanisms, several iterative steps are needed to reach a stable operating point.

We could also identify substantial work done on higher pricing time-scales. While no studies on a packet-level could be found that directly refer to a wireless setting, a growing interest can be identified for pricing on the admission time-scale. Admission-control has become a standard element of radio resource management in wireless networks with Quality-of-Service guarantees. Prices can serve as decision variables for admitting or rejecting new user requests if the network is already congested. Additionally, providers can only admit users with higher valuation during congestion time by varying the admission price for a service.

Since admission-control is not as time-critical as power control, advanced allocation mechanisms can be used on this time-scale. Auction formats play an important role in which users compete for access and the price is formed in a distributed way. However, due to the asynchronous arrival of customers in the network, traditional auction formats need to be modified or auctions need to be conducted in regular time intervals.

Studies with a direct focus on provider competition are just emerging. One reason for this is that many researchers still assume cooperative network structures, in which providers jointly optimise their network operations and avoid competition on the access time-scale or flow time-scale. However, with the expected deployment of high-speed wireless networks and the increasing independence of services from network transport, provider competition is expected to increase (Ormond et al., 2006). This is especially true when considering the opportunities for smaller players offering low-cost network transport services by using native IP solutions such as WiMAX or IEEE802.11n standards. Such standards by far surpass the capacities achievable by current 3G technology, even when considering 3.5G standards such as HSDPA or EV-DV. By offering high network capacities at flexible and adaptable pricing in high-density areas, this competition may also increase the motivation by the large mobile providers to explore additional business models to sell network transport services.

With this review and the substantiation of the scenario selection in the previous section we are now ready to proceed with presenting our work on dynamic pricing in wireless networks with the focus on direct competition between network providers.

## ▶ 3.7 Chapter Appendix: Technical Background

This section provides a non-technical overview of the engineering concepts relevant to this thesis. In the first section we provide a short introduction into multiplexing over the air interface. We then present the main concepts for providing Quality-of-Service in cable-based networks and describe the additional complexity introduced by the air interface. Due to the analytical density of the concepts and the large body of literature, we are only able to provide a general overview about these topics. For a more detailed technical description of the presented topics the reader is referred to the respective literature such as Sheldon (2001); Wisely et al. (2002); Walke et al. (2003); Sheikh (2004).

### ▶ 3.7.1 Multiplexing and data transmission over the wireless channel

Multiplexing techniques in wireless communication are used to allow for the simultaneous use of the air interface by more than one mobile station (MS). Traditionally, the radio band was subdivided into several narrow-band channels, each of which was assigned to one transmission source. The technology, called *Frequency-Division Multiple Access* (FDMA) is commonly used for radio and TV transmission and has also been used for the first generation of mobile telephony (Walke et al., 2003). A second approach, called *Time Division Multiple Access* (TDMA), allows several MS to time-share the same frequency by subdividing the channel into different time slots. TDMA has been shown to be more efficient with bursty traffic sources since time slots can be dynamically allocated depending on the current bandwidth requirements of the MS. Both multiplexing schemes have in common that MS make exclusive use of the air interface in the assigned frequency or time slots. Therefore, intra-cell interference is less of a problem in FDMA and TDMA, while in TDMA, the main problem is to manage the access to the channel. The transmission power in such multiplexing schemes plays a secondary role to conserve battery power of the MS and to limit the interference between bordering cells. Because of the static frame structure of TDMA, the capacity and transmission radius is fixed on the physical level.[10]

In a search for more efficient multiplexing schemes, the *Code-Division Multiple Access* (CDMA) scheme has emerged as the most successful technique for future wireless network technologies and has found widespread use with current 3G implementations. CDMA encodes data by orthogonal *spreading codes* associated with a channel and uses the constructive interference properties of the signal medium to perform the multiplexing. Thus, data from several MS can be transmitted simultaneously and can be reconstructed by the receiver by using the same spreading code. Since essentially, an infinite number of such codes is available, CDMA systems can supply more MS compared to other multiplexing schemes. The use of CDMA also avoids the overhead of continually allocating and deallocating a limited number of time slots or frequency channels between MS. In

[10]The transmission rates on higher levels may be lower due to transmission errors, which depend on factors such as distance from the base station, channel fading, or noise.

addition, CDMA offers increased support for bursty packet traffic since MS can stay active without actually transmitting in the channel and can directly start transmission without requesting dedicated resources from the base station.

CDMA systems have many other advantages compared with TDMA and FDMA. The most important one is the higher spectral efficiency of the wireless channel. Additionally, due to the spreading of the signal over a larger frequency band, CDMA is robust against interference and noise. However, CDMA also introduces additional complexity, which needs to be managed by the radio resource management (RRM) layer.

On the uplink, CDMA systems are interference-limited due to the simultaneous transmission of multiple mobile stations (MS) over the same channel. To find a feasible power allocation for all active MS, the base station (BS) needs to continuously monitor the signal-to-interference ratio (SIR) of the received signal and has to adapt power allocation of each MS. Factors such as fast and slow fading, external interference and shadowing additionally require a continuous power adjustment by all active MS. Mobility is another reason for continiously adjusting power to minimise interference caused on all other MS.

On the downlink, CDMA systems are limited by the transmit power of the BS to supply all MS with a sufficient SIR. Since the spreading codes are orthogonal, intra-cell interference is minimal and only inter-cell interference and background noise needs to be considered. The fact that power assignments are variable is also one reason for the soft capacity of the CDMA system. Therefore, the maximum cell size is determined by the channel gain and the SIR requirements of the active MS and can vary with the actual load of the cell.

### ► 3.7.2 General Quality-of-Service architectures

When analysing pricing and differentiated services in cable-based networks, two basic concepts play an important role in providing service guarantees: *Integrated Services* (IntServ) and *Differentiated Services* (DiffServ). While both architectures do not explicitly include pricing in their core assumptions and do not propose any pricing related models for resource prioritisation, they provide the required basis for implementing a pricing scheme for services requiring guaranteed Quality-of-Service. We briefly present the core concepts of both architectures and describe the relevance for the pricing in communication networks.

#### The Integrated Service architecture

The *Integrated Service* (IntServ) architecture has been developed with the intention to provide customised support for different service classes. It enables an explicit flow-based end-to-end reservation of resources before the actual data transmission is started. To enable the reservation of resources, each router on the network path needs to support the IntServ architecture and needs to keep track of the admitted flows which are actively using resources (Figure 3.8).

To signal new reservation requests across the network and to provide updates on existing flows the *Resource ReSerVation Protocol* (RSVP) has been proposed as protocol. It defines different message types, which are commonly understood by all routers in the network. The routers between the sender and listener have to decide if they can support the reservation being requested.

Two different service classes have been defined within IntServ: *Guaranteed Services* (GS) and *Controlled Load* (CLS). GS supports delay-intolerant applications such as interactive video or voice by making explicit guarantees for maximum delay and minimum bandwidth availability. In contrast, CLS provides an application with approximately the end-to-end service of an unloaded best-effort network. This means that instead of providing strict guarantees on delay and loss, the CLS service simply guarantees relative service quality with a low error rate and a low end-to-end latency. CLS is also often called better-than-best-effort because of its ability to simulate lightly loaded network conditions.

IntServ has quickly been shown to be impractical in providing end-to-end flow control in a distributed network such as the Internet. First of all, the use of per-flow state and per-flow processing raises scalability concerns for large networks (Wu, 2005). IntServ in its basic form does not provide any features for bundling of flows with similar QoS requirements. Therefore, IntServ enabled routers need to keep track of each admitted flow individually. Another disadvantage is the enormous signalling overhead created by the RSVP messages. To ensure that no abandoned flows are kept in the reservation table, applications are required to send keep-alive messages every 30 seconds. Especially for short-living flows, the IntServ architecture means a large burden of additional network traffic.

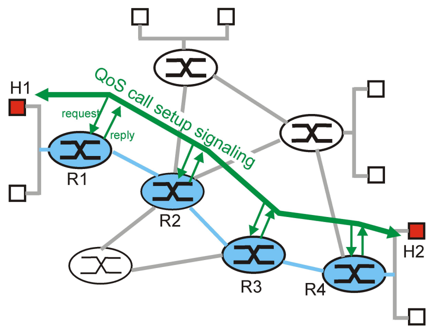



**Figure 3.8:** The IntServ architecture. Source: Kurose and Ross (2005)

## The Differentiated Service architecture

The *Differentiated Services* (DiffServ) architecture emerged as a complement to IntServ and addresses the shortcomings of its counterpart by a simpler differentiation of traffic (Plasser et al., 2002). Instead of making guarantees on a per-flow basis, DiffServ has been designed to offer service guarantees to aggregates of classes, where the number of individual flows are aggregated into different traffic classes. Therefore, no per-flow state needs to be maintained in the routers and no per-flow admission control is required. The classification of the service flows is pushed to the edges of the network and routers of the core network only need to implement the ability to handle different packets with different priorities (Figure 3.9).

Three classes have been defined to allow for differentiated services; *Expedited Forwarding* (EF) defines the highest priority service with highest forwarding priority and low delay. *Assured Forwarding* (AF) guarantees the delivery of the packet but allows for higher variance in overall delay. *Default Forwarding* (DF) corresponds with the best-effort service of today's Internet with variable delay and packet loss in case of congestion.

One disadvantage is that DiffServ cannot give absolute guarantees for single flows. Since no admission control takes place beforehand networks can still experience congestion and service quality degrades. Also, DiffServ provides only a coarse classification and the details of how routers deal with such service classes remains arbitrary. If packets cross two or more different DiffServ clouds, it is difficult to predict end-to-end behaviour (Courcoubetis and Weber, 2003).

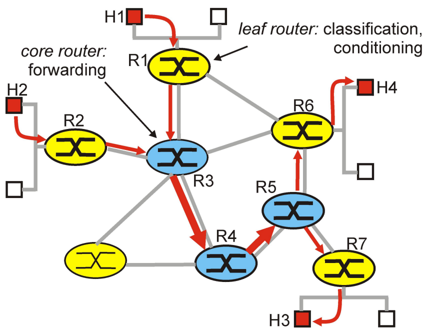



**Figure 3.9:** The DiffServ architecture. Source: Kurose and Ross (2005)

### Pricing with IntServ and DiffServ

Both resource management schemes have been the base for numerous proposals for implementing pricing for differentiated services in fixed networks. Most of the studies concentrate on DiffServ because of the higher chances to be implemented in future network structures such as the Internet (see, for example, Stiller et al. (2001a); Semret et al. (2001); Maglaras and Zeevi (2005)). However, also IntServ has attracted researchers for two main reasons: first, the RSVP protocol is an excellent base for communicating prices between network entities (for a protocol of this type see Wang and Schulzrinne (2000)) and second, because IntServ allows for the stringent guarantee of QoS parameters important for many applications (see, for example, Jin et al. (2003)).

### ► 3.7.3 Particularities for providing Quality-of-Service over the air interface

Compared to resource allocation in fixed networks, in which the quality of the transmission on the physical layer is usually relatively stable, the transmission over the air interface introduces a large degree of variability in terms of data throughput and error ratio. Because of the limitation in radio spectrum, an important engineering task is to increase the efficiency of data transmission by tailoring protocols toward the full support of IP-based protocols. However, providing QoS over a channel with stochastically changing capacity due to fading and interference between mobile terminals remains a complex task (Zhu and Chlamtac, 2005). We summarise the main complications on the physical layer into five points:

- *Shared channel:* In wireless networks all users use the same frequency band for data transmission. Depending on the allocation of channel time (TDMA) or power and code slots (CDMA) between mobile terminals and the channel gain of each mobile station the overall channel capacity varies. A QoS scheme needs to consider such factors to, on the one hand, provide individual service guarantees, and, on the other hand, optimise the overall network efficiency by, for example, favouring customers with large channel gains.

- *Varying channel capacity:* In contrast to wired networks, which provide fairly stable transmission rates at a very low error rate, the capacity and quality of the wireless channel varies randomly with time. Degraded channel quality may be caused by interference between the signals of different users or by noise from other radio emitting devices. The channel quality may also be influenced by physical structures, which cause fading of the radio signal. A QoS scheme need to be able to handle such changing conditions and to maintain a minimum quality level.

- *Mobility:* Wireless networks allow for mobility of cell phones and computing devices within the transmission range of the network. Mobility causes the signal quality to change rapidly as the user changes positions. Also, the speed of movement causes Doppler effects, which can distort the radio signal. The underlying QoS scheme needs to be able to handle this additional complexity. Mobile networks are based on a cellular network infrastructure, which allows users to stay connected even when moving through different cells at high speeds. When switching between cells, users expect the same QoS guarantees in the new network environment and the QoS scheme needs to be capable of pre-negotiating for resources before switching to the new cell.The latest technical developments allow for mobility in a heterogeneous network environment, where active network sessions are serviced by different network technologies during the duration of the session.

- *Limited energy:* Mobile devices are usually powered by battery. The power needed for data transmission over the wireless networks determines the overall operating time of the device. QoS schemes may need to consider such factors in order to minimise energy consumption and battery lifetime of the mobile devices.

- *Multi-hop communication:* Wireless communication opens up the opportunity of dynamically forming ad-hoc networks consisting of several mobile stations. Ad-hoc networks allow to dynamically set up network structures, which extend the reach and capacity of non ad-hoc wireless networks. Providing end-to-end QoS in ad-hoc networks, in which data is forwarded by several nodes before reaching the core network is the subject to current research.

QoS extensions for wireless networks managing the physical transmission layer need to consider the above factors in order to efficiently provide differentiated services with guaranteed service quality. Since such factors influence the service quality on different layers, a resource management approach providing QoS needs to be aware of the capabilities of each functional layer and needs to be informed about the current status in each layer (Berezdivin et al., 2002).

The existing QoS architectures in fixed networks have only limited capabilities for supporting the additional complexity introduced by the wireless channel. For example, using RSVP in a wireless environment is problematic due to the unstable wireless links. Another issue with RSVP is the need for mobility support, which requires to regularly change the established paths (Passas and Merakos, 2003).

Different cross-layer architectures have been proposed to address several of the problems (Nahrstedt et al., 2005). Some prominent schemes are dRSVP (Mirhakkak et al., 2001), INSIGNIA (Lee et al., 2000), and SWAN (Ahn et al., 2002).

# Chapter 4

# The Progressive-Second-Price Auction for Flow-based Resource Allocation in a Competitive Wireless Environment

## ▶ 4.1 Introduction

In this chapter we study a decentralised, flow-based resource allocation model for wireless resources based on congestion pricing. The main idea of congestion pricing is to update prices dynamically over time such that prices increase during congestion periods and cause users to reduce their demand (Yuksel, 2002). Instead of using a central authority controlling the price we use an auction, which lets users interact by submitting bids and learning about competitive bids.

In this chapter wireless resources are seen as a public good. Therefore, the main objective of the allocation process is the maximisation of social welfare, in a setting in which multiple, competing network providers offer resources to customers on-demand. Customers are assumed to have no predefined relationship with any of the providers but can negotiate resources at the time of their demand. With changing network conditions, resource allocation can be adapted through the auction process.

The auction format used in this work is a modified second-price (Vickrey) auction, called the *Progressive-Second-Price* (PSP) auction. It embodies all of the favorable properties of the Vickrey auction format such as efficiency and incentive-compatibility. With PSP a user submits a bandwidth-price pair to express his demand. Assuming elastic demand, users update their bids based on the auction result and the bid profiles of competitive bidders.

While the PSP auction format has been reasonably well understood in a one-auctioneer setting for controlling demand in a single access link, we extend its use to a setting in which multiple, competing network providers use the PSP auction mechanism to allocate resources between users. Besides understanding the implications of such a market, one main focus of this work is the development of bidding strategies which allow users to distribute their demand among multiple providers. Because each auction itself motivates users to reveal their true valuation, a goal is to understand how the existence of a multi-auction setting influences the users' decisions to distribute their demand.

One central assumption of this work is that end devices are able to bundle resources from multiple wireless access links. This method is called *multihoming* and is commonly used in wired networks to increase the reliability of the Internet connection for an IP network. However, a similar technique can be used in wireless networks to bundle bandwidth from multiple, independent connections. While currently not used in practical implementations, multihoming is considered as a future technique to let end devices handle wireless resources more flexibly and stably (Bahl et al., 2003).

In the following we provide a brief overview of the mechanism selection process and the additional assumptions taken for modelling resource allocation in the wireless channel.

### ▶ 4.1.1 Auctions as market institutions for resource allocation

In principle, any market institution can be used to allocate resources among users and to determine the price level so that demand matches supply. In a centralised setting, providers determine the optimal price level from the signals they receive from users. Such signals can be derived from past usage patterns or from market surveys. With additional information, providers are able to update price plans and optimise resource usage (Courcoubetis and Weber, 2003). In such a setting, users usually have no incentives to reveal their true valuation for the traded resources. Rational users will always try to shade their true valuation to maximise their net utility. Providers may not be able to allocate resources so that they are efficiently distributed among users.

Instead of using a centralised approach for price setting, providers can also make use of a distributed mechanism for letting users compete for resources. In such a setting, each user needs to decide on how to behave in order to maximise his utility, given that he receives signals from the market about the actions taken by other users. Their choice of action may depend on the level of information they have about the market situation.

Also, the rules defined by the mechanism may motivate users to behave in certain ways.

One prominent and widely used class of distributed mechanisms are auctions. Auctions have gained a lot of popularity among allocation mechanisms. Under certain conditions auctions can allocate resources efficiently between players. A key feature of auctions is the presence of asymmetric information between the bidders and the auctioneer (Klemperer, 2004). In a private-value model, each bidder only knows his own valuation for a resource, but other parties are unaware of this value. The auction process invokes a partial revelation of this information to other parties. Depending on the auction format, bidders may learn about the bids of other bidders and may adapt their actions accordingly.

PSP has been found to provide optimal features for the task to be accomplished. It allows allocating an arbitrarily divisible resource, such as bandwidth, between multiple users. Through the use of a second-price mechanism, it motivates users to reveal their true valuation (incentive-compatibility). Also, it incentivises users to participate only if they can gain positive utility (individual rationality). Finally, PSP has been shown to drive resource allocation to the economically most efficient allocation. *Economic efficiency* uses the concept of *social welfare*, which is the sum of the individual surplus gained by agents producing or consuming resources. A mechanism is called efficient if it allocates resources to those consumers who value them the most (Krishna, 2004).

PSP strongly builds on the concept of negative network externalities, which occur when scarce resources are shared by selfish users which are not willing to coordinate their actions (Liebowitz and Margolis, 1994). By charging users according to the level of externalities they are causing in the market, users are motivated to adapt their resource usage accordingly. With the PSP auction, we take a game theory approach to this problem by providing a mechanism where the intelligence and decision-making is distributed (Semret, 1999). The objective of an efficient and fair allocation is reached solely by the design of the allocation mechanism itself.

With the basic PSP mechanism invented by Semret (1999), several subsequent model extensions have been developed in the research community. Tuffin (2002) have proposed small changes to the allocation mechanism to eliminate allocation problems when using PSP in a stochastic environment. Maillé (2003) has detailed the equilibrium concept of PSP and has further investigated the role of the auctioneer reserve price on revenue generation. Another PSP "spin-off" is the multi-bid allocation mechanism developed in Maillé and Tuffin (2004a) and Maillé and Tuffin (2004b), which makes use of an approximated demand function submitted in full to the auctioneer. While keeping the principle PSP allocation rule and pricing rule, multi-bid eliminates the convergence process to equilibrium by identifying the equilibrium allocation within one step. Maillé and Tuffin (2004b) have shown that the resulting mechanism retains all important properties such as efficiency and incentive-compatibility while increasing technical efficiency by minimising the message exchange between agents. However, this approach changes some fundamental assumptions of PSP since players reveal their demand in full to the auctioneer.

### ▸ 4.1.2 The PSP auction mechanism in a wireless IP-based environment

The PSP concept has not yet been transfered to resource allocation in a wireless setting.[1] On the other hand, the wireless channel introduces additional complexity to the problem of resource allocation. Since we are interested primarily in the economic aspects of introducing competition in the wireless access link, we need to greatly abstract from such complexity to provide meaningful results. For the research described in this chapter we therefore make the following simplifying assumptions:

- Providers can assure a fixed capacity of the wireless link, which can be arbitrarily distributed among users independent of their position and level of mobility.
- End devices are capable of connecting to more than one network simultaneously.
- Resources from multiple providers can be bundled to increase the capacity of the network link.

As an additional restriction we exclude the problem of cell handovers and assume that users stay within one cell during the time of connection.

While this level of abstraction may not appreciate the complexity of a realistic environment, we are able to understand the consequences of competition without being distracted by the many other drivers of complexity. The described scenario fits the situation for fixed wireless in which mobility is low, connection times are usually longer than with mobile access, and no cell handovers occur. In such a setting customers may find themselves in the coverage area of multiple networks offering network resources on a dynamic basis.

### ▸ 4.1.3 Chapter outline

This chapter is structured into six sections. In the first section we introduce the Progressive-Second-Price auction and describe its basic properties (Section 4.2). Section 4.3 extends the use of the PSP mechanism in a setting of multiple, competing auctions with bidders being able to fulfill their demand from more than one auction simultaneously. The core of this section defines a bidding strategy that resembles the best response of the one-auction case for the multi-auction scenario and lets bidders distribute their demand between several auctions. We provide detailed proofs of the market properties when all bidders implement this bidding strategy. Section 4.4 describes four alternative bidding strategies for agents with different objectives or capabilities. We explore such bidding strategies by simulation and experimentally compare their properties. In the following two sections, we describe the results from the simulation experiments. Section 4.5 focuses on the experimental exploration of the market properties beyond the simple scenarios assumed for the

[1] An exception is the work by (Maillé, 2004), in which he applies the multi-bid auction format, which is closely related of the PSP auction format, for the downlink resource allocation in a CDMA-based cell. Since the multi-bid auction assumes full knowledge of the agents' utility functions, the proposed approach can not be directly compared with our work.

analytical proofs. Section 4.6 presents the results from a multi-cell simulation experiment in which the results from different resource allocation algorithms are compared with a scenario in which all providers allow for the dynamic allocation of resources using the PSP auction mechanism. We summarise the results in Section 4.7.

## ▸ 4.2 The Progressive-Second-Price auction in a single-seller setting

When bidders' values are private, the *Vickrey-Clarke-Groves* (VCG) *mechanism* is a direct, incentive-compatible and efficient mechanism, implying that VCG calls for truthful revelation of bidder's values (Krishna, 2004). When the object being auctioned is divisible, the VCG mechanism requires a bidders to reveal their entire valuation function. Closely related to VCG for divisible goods is the *Progressive Second-Price* (PSP) *auction*, in which the messages are reduced to a price-quantity combination (Semret, 1999). The resulting game becomes iterative with all bidders reacting to the price-quantity combinations submitted by the opponents. In addition, the strong properties of VCG are relaxed in terms of the equilibrium concept.

In its original introduction PSP is a mechanism to allocate bandwidth among competing users of a communication network. On a broader view, PSP is a decentralized and distributed market mechanism for resource sharing in networks. The use of a market mechanism stems from the need to acknowledge that different users may have different valuations for the resource. It is also used as a key part to the provision of Quality of Service (QoS). Selfish users attempting to gain resources, such as bandwidth, from a network provider, will contribute to the state of congestion above which QoS may be compromised. PSP is a mechanism by which prices constitute a dynamic response to (unpredictable) demand, while keeping a trade-off between engineering efficiency and economic efficiency (Semret, 1999).

In the following section, we provide a general introduction of the PSP mechanism and summarise the findings of Semret (1999). Additionally, we use material from subsequent work, which has been motivated by the publication of the PSP auction ( Semret et al. (2001); Tuffin (2002); Maillé and Tuffin (2003); Maillé (2003), Maillé and Tuffin (2004b) and Maillé and Tuffin (2004a)). In many parts of this section we make use of graphical representations of the given analytical expressions and provide simple examples to simplify the understanding of the material.

### ▸ 4.2.1 The PSP auction

The total resource $Q$ to be auctioned by a seller is defined as arbitrarily divisible and can be allocated in any combination to the set of participating players[2] $i \in 1, .., I$. All players submit bids to the seller depending on their actual allocation and the obtained

[2] In the following we use the expression player and bidder interchangeably.

results from the last auction round. The message exchanged between the bidders and the seller consists of a two-dimensional bid $s_i = (q_i, p_i) \in \mathcal{S}_i = [0, Q] \times [0, \infty)$, meaning that an agent wants a quantity share of $q_i$ at an unit price of $p_i$. Consequently, a bid profile is defined as $s = (s_1, ..., s_I)$. In standard game-theory notation, the profile $s_{-i}$ is defined as $(s_1, ..., s_{i-1}, s_{i+1}, ..., s_I)$ with $s = (s_i; s_{-i})$. Therefore, $s$ is a vector consisting of $I$ elements of type $s_i = (q_i, p_i)$.

The PSP auction defines an allocation rule $A$, which maps the bid profile $s$ into an allocation profile $a$:

$$\begin{array}{rrcl} A: & \mathcal{S} & \longrightarrow & \mathcal{S} \\ & s = (p, q) & \longmapsto & A(s) = (a(s), c(s)), \end{array}$$

where $\mathcal{S} = \prod_{i \in \mathcal{I}} \mathcal{S}_i$. The $i$-th row of $A(s)$, $A_i(s)$, gives the allocation for player $i$. Player $i$ gets an allocation of $a_i(s)$ and has to pay a total price of $c_i(s)$. The allocation rule is said to be feasible if for all $s$, $\sum_{i \in \mathcal{I}} a_i(s) \leq Q$, and for all $i$, $a_i \leq q_i$ and $c_i \leq p_i q_i$.

**Definition 4.1.** *The PSP auction is defined by an allocation rule $a_i(s)$ and a price rule $c_i(s)$*

$$a_i(s) = \min\left(q_i, \frac{q_i}{\sum_{k:p_k = p_i} q_k} Q_i(p_i, s_{-i})\right), \tag{4.1}$$

$$c_i(s) = \sum_{j \neq i} p_j \left[a_j(0; s_{-i}) - a_j(s_i; s_{-i})\right], \tag{4.2}$$

*where for $y \geq 0$, $\underline{Q}_i(p_i, s_{-i})$ is defined as*

$$\underline{Q}_i(y, s_{-i}) = \left[Q - \sum_{p_k \geq y, k \neq i} q_k\right]^+$$

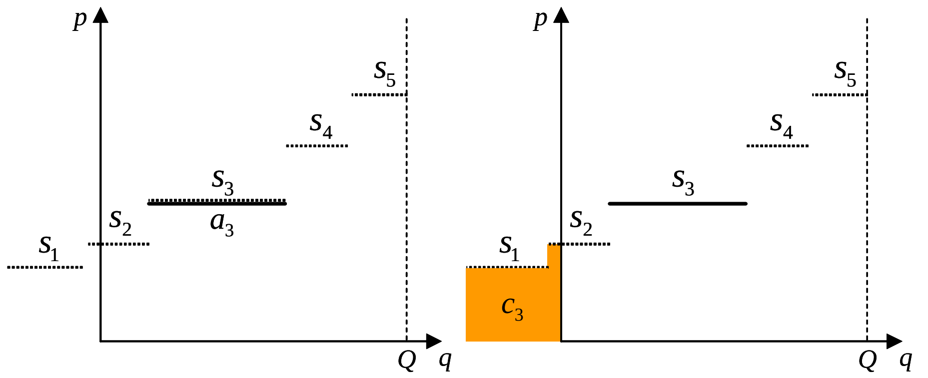
*and*

$$Q_i(y, s_{-i}) = \lim_{\eta \searrow y} \underline{Q}_i(\eta, s_{-i}) = \left[Q - \sum_{p_k > y, k \neq i} q_k\right]^+ .$$

The PSP allocation rule (4.1) assigns the total resources $Q$ according to the corresponding unit prices in descending order until all resources have been allocated. The lowest winning bidder may be allocated only parts of the requested share $q$ while all other winning bidders receive their full request (illustrated in Figure 4.1(a)).

If a player receives a positive amount of resources the costs are determined by the price rule. The intuition behind (4.2) is the *exclusion-compensation principle*, which is based on the negative externality caused by the winning bidder in the auction if the overall demand exceeds the available supply. The costs for player $i$ determined by (4.2) cover exactly the *social opportunity cost*, which is given by the willingness-to-pay declared by players fully or partially excluded from the auction. A bidder's payment can be interpreted as a compensation payment to the seller for lost revenue from other players excluded by the player's presence. Figure 4.1(b) shows the cost allocation $c_3$ for player 3.

The definition of the PSP auction in Definition 4.1 contains a rule modification that




(a) The allocation rule sorts bids by the corresponding unit price and assigns resources up to the total available quantity $Q$.

(b) The pricing rule determines the social opportunity costs a winning bidder causes on the auction by summing up the willingness-to-pay of bidders partially of fully excluded by the winning bidders presence.

**Figure 4.1:** The PSP allocation and pricing rules.

has been proposed in Tuffin (2002). This modification changes the allocation of $a$ for the case that the bid profile contains winning bids with identical unit prices. In the original version of the PSP auction, resources do not get fully allocated if bids with identical unit prices have been submitted by different bidders.[3] With this modification the total quantity is proportionally shared between winning bidders when bidding at identical unit prices. Note that the old and the new allocation rules only differ when two players bet the same amount in bandwidth, but give identical results otherwise.

The allocation rule can also be explained graphically. $Q_i(y, s_{-i})$ is the "staircase" function shown in Figure 4.2(a). All bids in the opponent bid profile are compared with $y$. If demand exceeds supply, $Q_i(0, s_{-i})$ will be zero since all bids will have a unit price larger than zero. With an increasing unit price $y$, the bids of more and more opponent bids will be below this value and $Q_i$ will become stepwise larger. This creates the step function that reaches $Q$ when $y$ is larger than any bid price in the opponent bid profile.

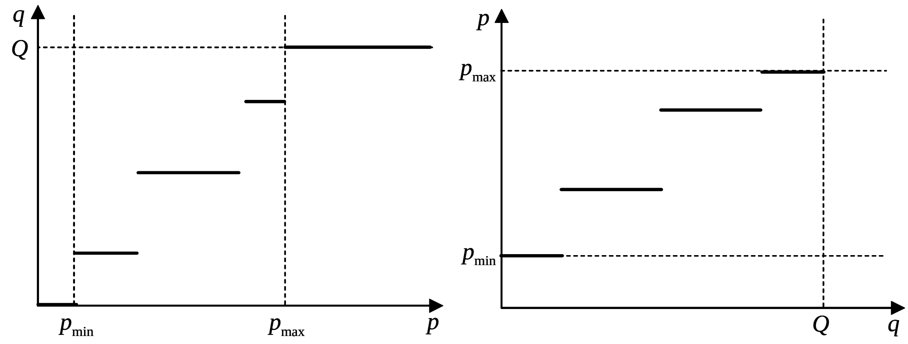
Similarly, the inverse step function of $Q_i$ can be defined

$$P_i(z, s_{-i}) = \inf\{y \geq 0 : Q_i(y, s_{-i}) \geq z\},$$

which gives the unit price required to win a certain quantity $z$ in the auction. Figure 4.2(b) shows and example of $P_i$. With this definition the total payment for each bidder $i$ can be expressed as the integral of $P_i$ over the winning share $a_i$

$$c_i(s) = \int_0^{a_i(s)} P_i(z, s_{-i})dz, \tag{4.3}$$

[3] The original allocation rule defined in Semret (1999) is $a_i(s) = \min(q, Q_i(p_i, s_{-i})$. Example: Image two bidders with the bids $s_1 = (50, 10)$, and $s_2 = (60, 10)$ and $Q = 100$. In this case, PSP allocates $a_1 = 50 \wedge (100 - 60) = 40$ and $a_2 = 60 \wedge (100 - 50) = 50$. Instead of allocating the 100 units only 90 units will be allocated between users. With the modified PSP allocation rule the allocation becomes $a_1 = 45.5$ and $a_2 = 55.5$, respectively.




(a) $Q_i(y, s_{-i})$. (b) $P_i(z, s_{-i})$.

**Figure 4.2:** $Q_i(y, s_{-i})$ and $P_i(z, s_{-i})$ are step functions, which calculate the resulting quantity (a) and the required unit price (b) to acquire a certain quantity with a given opponent bid profile. can be shown as step functions, sorting all bids of a bid profile with the resources available. The picture shows two corresponding functions.

which adds up the willingness-to-pay for all bidders excluded by player $i$.

**Example 4.1.** *To better understand how the PSP rules work consider the following simple example. Five bidders submit their bid to the PSP auction, which has a total capacity of* $Q = 100$ *to allocate between players. The bids are* $s_1 = (10, 25)$, $s_2 = (40, 15)$, $s_3 = (60, 10)$, $s_4 = (45, 8)$, *and* $s_5 = (30, 5)$. *The allocation rule sorts the bids according the unit prices and allocates resources as given in Table 4.1. The payment by each bidder is determined according to the social opportunity costs caused by his presence in the market. Table 4.1 lists the opportunity costs caused by each player in a matrix, and the corresponding source by player. To further explain the calculation of the opportunity costs consider the example of player 2. By his presence, he partially excludes bidder 3 and fully excludes bidder 4 from the market. Bidder 3 obtains only* 50 *of the requested* 60 *units, which he values with a per-unit valuation of* 10. *Bidder 4, with a willingness-to-pay of* 8 *per unit, is fully excluded from the auction and would obtain* 30 *units if bidder 2 were not present.*

| | Bid | Allocated | Total | Opportunity Cost caused on Player | | | | |
|---|---|---|---|---|---|---|---|---|
| | (Quantity, Unit Price) | Quantity | Payment | 1 | 2 | 3 | 4 | 5 |
| Player 1 | (10,25) | 10 | 100 | x | x | 100 | 0 | 0 |
| Player 2 | (40,15) | 40 | 340 | x | x | 100 | 240 | 0 |
| Player 3 | (60,10) | 50 | 385 | x | x | x | 360 | 25 |
| Player 4 | (45,8) | 0 | 0 | x | x | x | x | 0 |
| Player 5 | (30,5) | 0 | 0 | x | x | x | x | x |

**Table 4.1:** Bids, allocation, payment and opportunity cost in the PSP example

## ▶ 4.2.2 Properties of the PSP auction

In this section we briefly discuss the main properties of the PSP auction, which have been proofed by Semret (1999) and subsequent work. While we do not repeat such proofs, which can be found in the corresponding source, we give a short sketch of the properties that have been derived.

### Modelling of users and user preferences

Players of the PSP auction game are assumed as acting rationally, thus maximising their utility with every action played in the game. Since players know the opponent bid profile of the last auction round and are aware of the PSP rules, under which resources are allocated, no uncertainty exists. Players can therefore foresee fully the consequences of all possible strategies.

To determine the utility-maximising action each player $i$ has preferences, which define the valuation $\theta_i$ of a player for an allocated resource $a_i(s)$. The utility $u_i$ is assumed to be quasilinear of the form $u_i(s) = \theta(a_i(s)) - c_i(s)$, where $c_i$ are the cost of player $i$. If the valuation function is fully defined a bidder can assign a utility to each given bid profile $s$. The shape of the valuation function determines the actions of a bidder for a given opponent bid profile $s_{-i}$. Users with elastic demand change their per-unit valuation with a change in obtained resources. Since PSP allows agents to adapt their demand depending on the overall demand situation on the market we assume that all users have elastic demand.

Two features are worth noting. First, the concavity of $\theta_i(\cdot)$ is essential in this formulation. In standard economic applications, this corresponds to the risk aversion of the user. In this case, we can interpret it either as risk aversion, for example, because users dislike potential variability in transmission rates and service quality, or as flexibility. With the latter interpretation, a more concave utility function implies that the user has little flexibility regarding when he or she can transmit whereas a less concave (closer to linear) utility function would capture greater flexibility.

For the remaining chapter we follow the general concavity and regularity assumptions defined by Semret (1999).

**Assumption 4.1.** *As in Semret (1999), the following assumptions on $\theta_i$, $\forall i \in I$, hold:*

- $\theta_i(0) = 0$,
- $\theta_i$ *is differentiable,*
- $\theta_i' \geq 0$, *non-increasing and continuous,*
- $\exists \gamma_i > 0, \forall z \geq 0, \theta_i'(z) > 0 \Rightarrow \forall \eta < z, \theta_i'(z) \leq \theta_i'(\eta) - \gamma_i(z - \eta)$.

Assumption 4.1 gives the general properties of $\theta$. The last item assumes that the valuation function is always concave but can flatten from a certain point on.

**Assumption 4.2.** $\exists \kappa > 0, \forall i \in I$,

- $\forall z, z', z > z' \geq 0, \theta_i'(z) - \theta_i'(z') > -\kappa(z - z')$,
- $\theta_i' < \infty$,

Assumption 4.2 makes some stricter assumptions about the valuation function $\theta$. The first item says that the derivatives of all valuation functions have a minimum slope of $\kappa$, which prevents the valuation function from getting flat before the maximum value. The second item requires that $\theta'$ is a differentiable function for all values of $a_i$.

While the valuation function can have any functional form corresponding to the above assumptions, a specific implementation is the parabolic (second-order) valuation function. In all simulation experiments in this chapter we make use of this type of valuation function, which is given by

$$\theta_i(z) = \begin{cases} -\frac{\overline{p}_i}{2\overline{q}_i} z^2 + \overline{p}_i z & \text{for all } z \leq \overline{q}_i \\ \frac{\overline{p}_i \overline{q}_i}{2} & \text{for all } z > \overline{q}_i \end{cases}$$

Semret (1999) has delivered a detailed substantiation for the use of this function for multimedia traffic in which he shows that the shape corresponds to the quality of most common compression algorithms used with audio and video applications.

### The PSP auction game

Having defined the PSP allocation rule and price rule as well as the user preferences we can now state the normal-form representation of the game repeatedly played between the $I$ bidders.

**Definition 4.2** (The PSP auction game)**.** *The normal-form representation $G$ of the auction game, which is played by the bidders with access to a single PSP auction is given by:*

$$G = (S_1, .., S_I, u_1, .., u_I),$$

*with $S_i$ defining the strategy space of player $i$ as $S_i = [0, Q] \times [0, \infty)$, and $u_i$ being the utility of player $i$, with $u_i(s) = \theta(a_i(s)) - c_i(s)$.*

### Equilibrium concept of PSP

Every time a bidder submits a new bid to the auction he tries to maximise his utility based on the updated opponent bid profile $s_{-i}$. The *best reply* of a player $i$, which gives the utility-maximising strategy, is given with Definition 4.3.

**Definition 4.3.** *The set of best replies to a bid profile $s_{-i}$ of opponents bids is defined as*

$$S_i^*(s_{-i}) = \{s_i \in S_i(s_{-i}) : u_i(s_i, s_{-i}) \geq u_i(\hat{s}_i, s_{-i}),$$

$\forall \hat{s}_i \in S_i(s_{-i})\}$.

In an iterative[4] game, where players recompute their best response based on a modified opponent bid profile, the bid profile can either converge to a Nash equilibrium or not

[4] We use the term *iterative* rather than *dynamic* since players do not devise a contingency plan or strategy over multiple rounds of the game but only react to the given opponent bid profile from the last round.

converge at all (Semret, 1999). With $S^*(s) = \prod_i S_i^*(s_{-i})$, a Nash equilibrium is a bid profile, which lies within $S^*$.

A Nash equilibrium is defined as a strategy vector, or, in terms of PSP, a bid profile $s$, from which no player has a unilateral incentive to deviate (Johari, 2004).

**Definition 4.4.** *A bid profile $s = (s_1, ..., s_I)$ is in Nash equilibrium if every agent maximises its expected utility given its type $\theta_i$, $\forall i$,*

$$u_i(s_i(\theta_i), s_{-i}(\theta_{-i}), \theta_i) \geq u_i(\hat{s}_i(\theta_i), s_{-i}(\theta_{-i}), \theta_i),$$

*$\forall \hat{s}_i \in S_i(s_{-i})$ and $\hat{s}_i \neq s_i$.*

The concept of the Nash equilibrium is fundamental to game theory, but requires very strong assumptions on the agent's information and beliefs (Parkes, 2001). Since PSP is modeled as a game with complete information, in which, after each auction round, all bidders are informed about the updated bids of their opponents, the Nash equilibrium concept can be directly applied.

With the above definition of the best reply, players would bid exactly the amount they evaluate as optimal for a given opponent bid profile $s_{-i}$. In consequence, the auction would not progress (the overall demand is not reduced) as bidders may bid with identical unit prices and may not reduce their demand in the subsequent bidding round. To avoid this, Semret (1999) has introduced a weaker equilibrium concept called the $2\epsilon$-Nash equilibrium.[5]

**Definition 4.5.** *The $2\epsilon$-best reply of agent $i$ can be defined as:*

$$S_i^{\epsilon}(s_{-i}) = \{s_i \in S_i(s_{-i}) : u_i(s_i, s_{-i}) \geq u_i(\hat{s}_i, s_{-i}) - \epsilon, \forall \hat{s}_i \in S_i(s_{-i})\}$$

With $S^{\epsilon}(s) = \prod_i S_i^*(s_{-i})$, a $2\epsilon$-Nash equilibrium is a bid profile, which lies within $S^{\epsilon}$. $2\epsilon$ describes an interval around the Nash equilibrium, within which it is unattractive for bidders to submit a new bid. If the utility improvement is smaller he will stay with the previous bid. When all agents reach the situation in which they cannot improve their utility by at least $\epsilon$ the equilibrium allocation has been found.

$\epsilon$ has been interpreted as a bid fee, which each agent has to pay in addition to the congestion charges when obtaining resources in times of congestion. This interpretation may be misleading as the PSP concept does not include the bid fee as provider revenue nor a transfer of such revenues is described. Rather, $\epsilon$ should be seen as a given threshold value, which has been defined by the designer of the mechanism and which is known by all entities in advance.

Maillé (2003) has pointed out that the definition of the $2\epsilon$ interval creates the possibility that multiple Nash equilibria exist for a fixed set of players. Thus, which specific

[5] Semret has termed this the $\epsilon$-Nash equilibrium. Tuffin (2002) describes it, more precisely, as $2\epsilon$-Nash equilibrium, since the interval has the size of $2\epsilon$.

equilibrium allocation is reached in a particular case depends on the order in which bids are updated. Such a feature may not be desirable in certain circumstances as the equilibrium allocation may differ in different rounds of the game. However, the social welfare achieved in equilibrium can be guaranteed within certain bounds, as we later show.

### Incentive-compatibility of PSP and the truthful best reply

In a Vickrey auction each bidder maximizes his expected utility by revealing his true valuation through the bid submitted in the auction. This property is called *incentive-compatibility* (IC). IC implies that it is a weakly dominant strategy for all players to tell the truth. A bidder cannot do any better by submitting a bid different from his truthful valuation. In this context, truthful behaviour can be interpreted as bidding at a unit price corresponding to the marginal valuation $p_i = \theta'(q_i)$ at quantity $q_i$. By following this strategy, a bidder maximises his utility automatically for the given mechanism. Therefore, the unconstrained set of truthful bids of bidder $i$ can be limited to the set of $\mathcal{T}_i = s_i \in \mathcal{S}_i$ : $p_i = \theta'(q_i)$ and $\mathcal{T} = \prod_i \mathcal{T}_i$.

Under Assumption 4.1, Semret (1999) has shown that a best truthful reply for player $i$, $t_i = (v_i, w_i)$, can be found in the following way: $\forall i \in \mathcal{I}$, $\forall s_{-i} \in \mathcal{S}_{-i}$, such that $Q_i(0, s_{-i}) = 0$ and for any $\epsilon > 0$, there exists a truthful $\epsilon$-best reply $t_i(s_{-i}) \in \mathcal{T}_i \cap S_i^{\epsilon}(s_i)$.

**Definition 4.6.** *A truthful reply in the PSP mechanism can be identified with:*

$$v_i = \left[sup\ G_i(s_{-i}) - \epsilon/\theta_i'(0)\right]^+, \text{ and } w_i = \theta_i'(v_i),$$

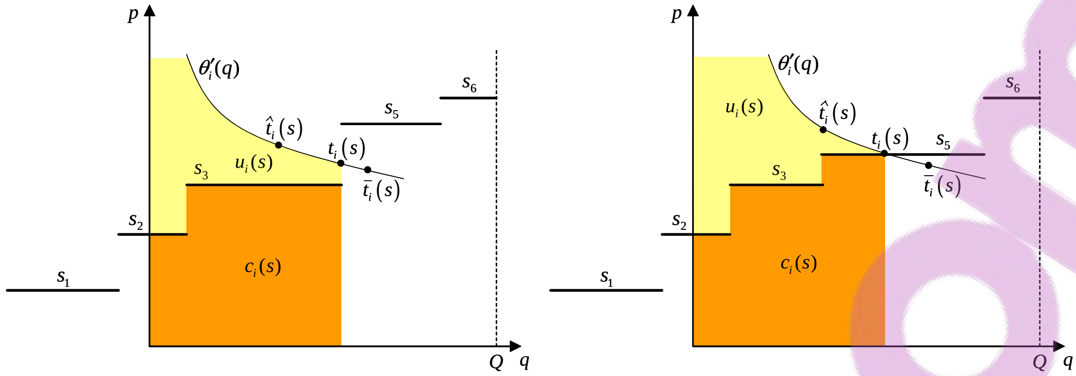
*using*

$$G_i(s_{-i}) = \left\{z \in [0, Q] : z \leq Q_i(\theta_i'(z), s_{-i})\right\}^6$$

Finding the truthful reply for player $i$ with a valuation $\theta_i$ and for a given bid profile $s$ can also be explained graphically (Figure 4.3(a) and Figure 4.3(b)). The truthful bid $t_i$ is found by identifying the quantity at which the marginal valuation $\theta_i'$ equals the value of the increasing bid profile $s$. The marginal valuation determines the truthful unit price $p_i$ to bid; $t_i$ will "beat" all bids below this unit price. These bids determine the charge $c_i$ player $i$ has to pay (using the exclusion compensation principle of the PSP price rule). The difference between the marginal valuation and the bids with unit prices below $t_i$ determine a player's utility $u_i(s_i) = \theta_i(a_i) - c_i$. Two cases can be distinguished. In Case 1, Figure 4.3(a), the truthful reply lies between two opponent bids, while in Case 2, the marginal valuation cuts trough an opponent bid.

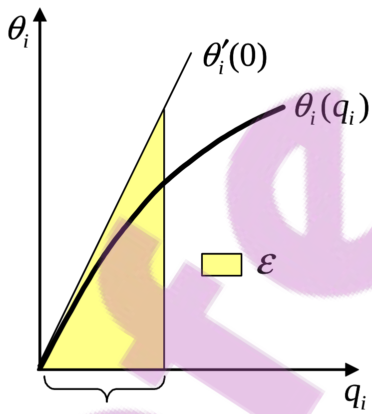
As shown in Figure 4.3(a), any other bid on the marginal valuation function achieves lower utility; $\hat{t}_i$ gains less resources for player $i$ even if his demand would allow him to bid for more resources. $\bar{t}_i$ also gains less resources as the unit price is lower than required to win the shares from the player with the highest unit price the bidder is able to beat with the given demand profile.

[6] We are omitting the condition for a budget limitation for better readability.




(a) Case 1. (b) Case 2.

**Figure 4.3:** The truthful reply is found with $t_i$. Other bids such as $\hat{t}_i$ or $\bar{t}_i$ result in a lower utility for the bidder.




$$= \sqrt{\frac{2\varepsilon}{\theta_i'(0)}} \geq \frac{2\varepsilon}{\theta_i'(0)} \geq \frac{\varepsilon}{\theta_i'(0)}$$

**Figure 4.4:** The reduction of the truthful reply by $\epsilon/\theta'(0)$ ensures that the new bid differs from the last bid by at least $\epsilon$.

The factor $\epsilon/\theta'(0)$, by which each truthful reply is reduced in quantity ensures that the new bid differs from the last bid by at least $\epsilon$ under the assumption of a concave valuation function (Assumption 4.1). Figure 4.4 shows how $\epsilon/\theta'(0)$ has been found as an approximation. For concave valuation functions, the first-order derivative of $\theta$ at point 0 gives the maximum slope of the valuation function. The area of the triangle formed between the x-axis and $\theta'(0)$ must be larger than the area included by the valuation function.

The proof of IC by Semret (1999) formally shows that the utility gained from using the $\epsilon$-truthful reply $t_i = (v_i, w_i)$ is larger than any other bid $s_i = (q_i, p_i)$ if $q_i$ is either smaller than $v_i$ or at least larger by $\epsilon/\theta'(0)$. In the region between, utility may be higher but is bounded by $\epsilon$ as shown in Figure 4.4. Thus, bidding truthfully automatically maximises a bidder's utility and the $\epsilon$-truthful reply is therefore the best reply a bidder can choose.

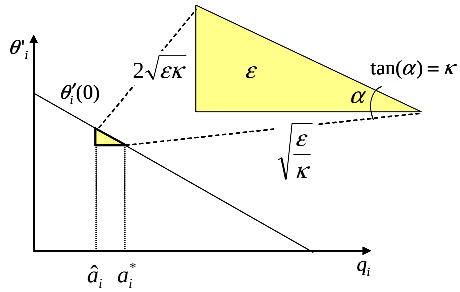



**Figure 4.5:** Boundary conditions in Nash equilibrium for the player $i$ with the smallest slope of $\theta'(z)$ at $a_i^*$.

Note that with the bidding strategy given with Definition 4.6 players only evaluate the bid profile of the last bidding round and maximise their utility for the next round. They do not take into account any information from prior auction rounds nor do they predict future behaviour of their opponents. Semret (1999) has shown that more aggressive bidding strategies can make a player better off if he is the only one using such a strategy. As soon as other players switch their strategy to a more aggressive behaviour this advantage diminishes. The beauty of the truthful revelation approach used by all bidders comes also from the fact that it does not need complex optimisation techniques but can be processed by devices with limited computing power. Since the bidding process is usually conducted by automated agents, the use of a conforming strategy, which cannot be altered by the user, can usually be enforced.[7]

### Nash equilibrium of the PSP auction

In the previous section we have already mentioned that the equilibrium reached by the iterative bidding process must be of Nash type if the auction converges at all. The formal proof by Semret (1999) uses the Karush-Tucker-Kuhn optimality condition to show that the equilibrium allocation of the PSP auction $a^*$ is a Nash equilibrium. By using Assumption 4.2, which requires a minimum slope of $\kappa$ for the marginal valuation functions of all players, it is shown that in equilibrium the difference in marginal valuation between player $i$ with the smallest positive allocation $a_i^*$ and any other player $j$ with an allocation $a_j^* > \sqrt{\frac{\epsilon}{\kappa}}$ is bounded by $2\sqrt{\epsilon\kappa}$. The relationship is shown in Figure 4.5. The bidder with the smallest slope $\kappa$ of the marginal valuation in equilibrium determines if a new truthful reply can be submitted or not. Since, for the submission of the bid the utility improvement must be at least $\epsilon$, the minimum change in the unit price can be determined by the relationship of the slope $\kappa = \tan(\alpha)$ and the area $\epsilon$.

[7]One good example for such an approach is the congestion control in the TCP protocol. Even if users are able to alter the algorithm to their advantage, such behaviour can rarely be observed in the Internet.

### Efficiency of the PSP auction

An auction allocates efficiently if resources are allocated to the bidders who value them the most (Krishna, 2004). In economic terms, efficiency is measured as social welfare that is generated from the pattern of resource allocation and is defined as the sum of the valuations of all agents receiving a positive resource share ($\sum^I \theta_i(s)$). The efficiency measure of social welfare is equivalent to the maximisation of the sum of all players' utility including the utility of the seller received from the payments of the other players. To see this we define the seller by $i = 0$ and define his utility as $u_i = \theta_0(a0) + \sum_{i \neq 0}^{I} c_i$. $a_0$ denotes a possible unallocated share of resources. Then, the total welfare in the system can be written as

$$\sum_{i}^{I} u_i = \sum_{i \neq 0}^{I} (\theta_i - c_i) + u_0 = \sum_{i \neq 0}^{I} \theta_i + \theta_0 = \sum_{i}^{I} \theta_i$$

The efficiency properties of the PSP auction can easily be understood intuitively. Imagine a situation in which two players $i$ and $j$ compete for resources. Whenever player $j$'s marginal valuation is smaller than the marginal valuation of player $i$, and $j$ receives a positive amount of resources, player $i$ can take away a share from $j$ at a price less than his own marginal valuation. The overall social welfare increases with this step because player $i$ values the share higher. In equilibrium no player is able to unilaterally improve his utility and social welfare is maximised.

Semret (1999) shows that social welfare in the $2\epsilon$ Nash equilibrium is close to the absolute optimal value within the bounds $4Q\sqrt{\epsilon\kappa}$ if the smallest non-zero allocation $\underline{a^*} > \sqrt{\epsilon/\kappa}$.

### Optimality of PSP

An optimal auction is a bidding mechanism designed to maximise a seller's expected profit, which consists of the cumulated charges from all bidders (Bulow and Roberts, 1989). The PSP auction has been designed with the main objective of efficiency, namely to maximise the social welfare created by the allocation of resources to different bidders. Therefore, the PSP auction is generally not revenue maximising.

A possibility to increase the sellers revenue is the introduction of a reserve price, under which no resources are sold. This reserve price has to be paid by all winning bidders even in times the market is not congested. To model a reserve price an additional player $i = 0$ can be introduced, which represents the auctioneer, and which bids with a fixed valuation $\theta_0 = p_0 Q$. The bid of this bidder determines the minimum bid needed by other bidders to obtain resources.

With $\epsilon = 0$, the revenue $R$ collected by a seller, the sum of all payments $\sum_{i \in \mathcal{I}} c_i(s)$, tends to be $p_0 Q$, which is the reserve price set by the seller multiplied by the total resources $Q$ to be sold. Maillé (2003) has followed up on this issue and has shown the correlation of revenue in equilibrium with the reserve price $p_0$. He defines the demand function for each bidder as $d_i(p) = \arg\max_q \{\theta_i(q) - pq\}$, which, under Assumption 4.1, can be written

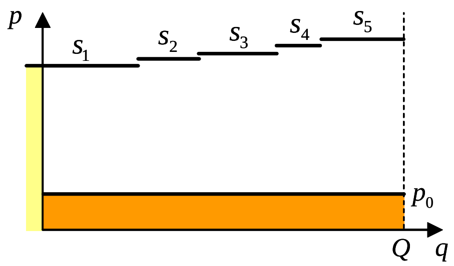



**Figure 4.6:** A typical allocation in equilibrium. There exists only small congestion (yellow area, which is paid by all bidders except $s_1$) but the reserve price $p_0$ ensures minimum revenue to the seller (orange area).

as

$$d_i(p) = \begin{cases} (\theta_i')^{-1}(p) & \text{if } 0 < p \leq \theta_i'(0) \\ 0 & \text{if } p > \theta_i'(0) \end{cases}$$

Under Assumption 4.1 all demand functions $d_i$ are continuous and non-increasing on $(0, +\infty[$. Then, the market clearing price $u$ is a unique price such that $\sum_{i\in\mathcal{I}} d_i(u) = Q$. We can also define $\mathcal{I}^+ = \{i \in \mathcal{I} : d_i(u) > 0\}$ to be the set of players which would buy a positive quantity at the market clearing price $u$. As necessary and sufficient condition for the existence of such a market-clearing price the overall demand at reserve price $p_0$ must exceed the overall supply: $\sum_{i\in\mathcal{I}} d_i(p_0) > Q$. We can conclude that the provider revenue $R$ must be in between $p_0Q > R > uQ$. Especially, when $\mathcal{I}^+ = \mathcal{I}$, the revenue will be exactly $p_0Q$. The reserve price is therefore an effective tool for maximising revenue in the PSP auction as the second objective behind welfare maximisation. It should be set below the market clearing price $u$, but very close to it to collect the maximum revenue.

The result obtained by Maillé (2003) can also be understood intuitively. With $\epsilon = 0$, players reduce their bidding quantity up to the point where no congestion exists. Since the revenue is based on the exclusion-compensation principle, the obtained revenue for the seller will be $p_0Q$ in equilibrium. With $\epsilon > 0$ agents do not fully reduce their bids until the sum of bids is equal to the resource quantity but some congestion remains in equilibrium. The remaining congestion determines the congestion-based revenue earned by the auctioneer. But since the equilibrium is not unique with $\epsilon > 0$ the revenue cannot be guaranteed and depends on the specific progression of the auction. At the same time, economic efficiency may be lower than with $\epsilon$ set to zero because bidders need to be able to increase their utility by at least $\epsilon$ to update their bids. As described by Semret (1999), this situation can be interpreted as a trade-off between economic efficiency and technical efficiency. Faster convergence and less convergence steps are traded against some loss in social welfare. Figure 4.6 shows a typical allocation with a reserve price $p_0$ for a small $\epsilon$.

Another important observation reveals a typical shape of the congestion-based revenues during the iterative bidding process. Figure 4.7 depicts such a graph for a 10 bidder example with parabolic valuation functions and $\epsilon = 0.01$. Congestion-based rev-

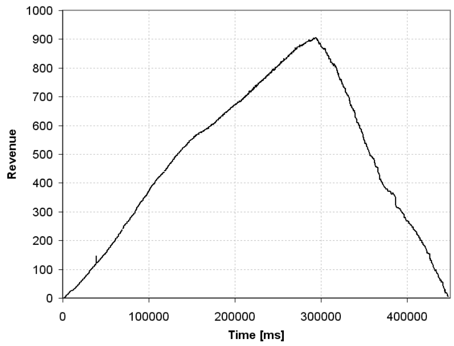



**Figure 4.7:** The typical shape of the cumulated revenue graph over the iterative bidding phase until an efficient allocation is reached ($Q = 100$, $\epsilon = 0.01$, $p_0 = 0$).

enues reach a global maximum after which they decrease to zero in equilibrium.

## ▶ 4.2.3 A simple PSP example

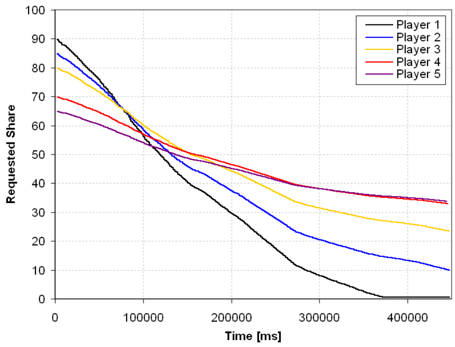
The following example applies the PSP auction rules and the truthful best-reply concept in a simple numerical setup. Five bidders with different valuation functions are competing for resources from one seller. The types of the bidding players are given in Table 4.2. The seller offers a total quantity of $Q = 100$. We set $\epsilon = 0.01$ and the smallest bidding unit $\alpha = 0.1$, where $\alpha$ gives the precision of the bids submitted to the auctioneer. All players become inactive for $t_i = 1s$ before again updating their strategy and potentially submitting a new bid.

| Name | Maximum Resource Share $\overline{q}$ | Maximum Marginal Unit Price $\overline{p}$ |
|---|---|---|
| BidderAgent1 | 90 | 10 |
| BidderAgent2 | 85 | 12 |
| BidderAgent3 | 80 | 15 |
| BidderAgent4 | 70 | 20 |
| BidderAgent5 | 65 | 22 |

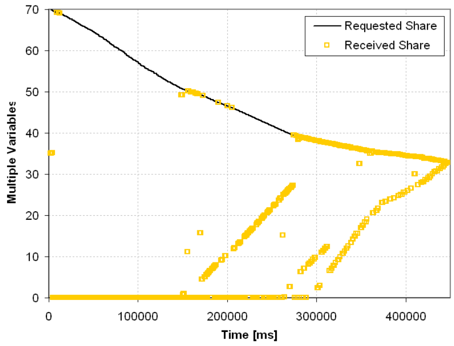
**Table 4.2:** Parameters for the parabolic valuation function of each bidder used in the example.

Figure 4.8(a) shows the bidding for all five players. All players gradually reduce their demand until equilibrium is reached and no bidder can unilaterally increase his net utility. Figure 4.8(b) shows the requested resource share and received resource share for player 4. During the convergence phase, the allocation $a_3$ often jumps from 0 to the demanded quantity. However, the shape of this graph depends on various factors. For example, when reducing the inactivity time to $t = 50ms$, the graph looks considerably different (Figure 4.8(c)). This is because with such a short update interval the players' actions become asynchronous and a player may become active before another player has updated his bid.

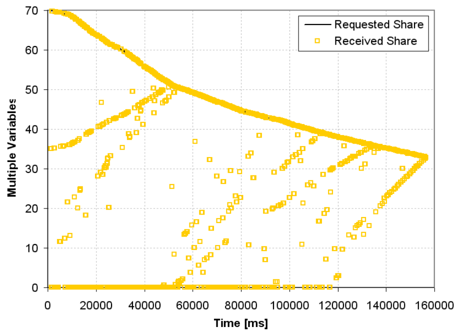
Figure 4.9(a) depicts the revenue generated as the sum of all costs $\sum_i c_i(s)$. Since $\epsilon$ is small, revenue in equilibrium may be close to zero. Before reaching equilibrium the




(a) Requested resource share for for players 1-5.

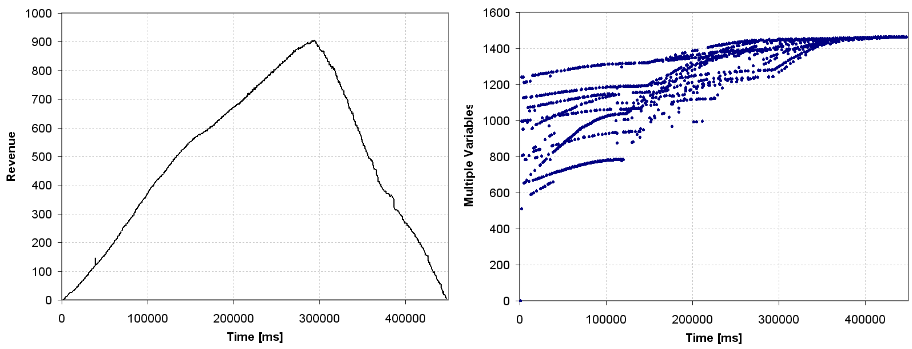



(b) Requested shares and received shares for player 4 ($t = 1000ms$).




(c) Requested shares and received shares for player 4 ($t = 50ms$).

**Figure 4.8:** Requested shares and received shares in the example scenario.




(a) Congestion-based revenue obtained by the seller. (b) Social welfare generated by the allocation of resources between bidders.

**Figure 4.9:** Revenue and Welfare over the convergence phase to equilibrium.

revenue goes through a maximum value after which it quickly drops to a near-zero value. The social welfare is given in Figure 4.9(b) as the sum of all agents' net utility and also includes the revenue of the auctioneer agent ($W(s) = \sum_i u_i(s) + \sum_i c_i$). This equals the sum of the players valuations $\sum_i \theta_i(s)$ since $u_i(s) = \theta_i(s) - c_i(s)$.

For this simple example the optimal allocation for each bidder can also be found centrally by solving the constrained maximisation problem

$$\max_{a_i} \sum_i \theta_i(a_i)$$

subject to

$$\sum_i a_i \leq Q$$

and

$$a_i \leq \bar{q}_i \forall i.$$

Table 4.3 compares the solutions obtained analytically and by simulation with the given values. It can be observed that the difference in social welfare is well within the maximum bound of $4Q\sqrt{\epsilon\kappa} = 13.333$. The main difference between the approaches is that the PSP auction solved the problem without full knowledge of the individual valuation functions while, in the analytical case, all types had to be known.

| | Analytical Solution | | Simulation experiment | |
|---|---|---|---|---|
| | Allocated Resource Share $a_i$ | Social Welfare | Allocated Resource Share $a_i$ | Social Welfare |
| Bidder1 | 0.00 | 0 | 0.00 | 0 |
| Bidder2 | 9.92 | 112.19 | 9.90 | 111.72 |
| Bidder3 | 23.48 | 300.48 | 23.47 | 299.83 |
| Bidder4 | 32.91 | 503.44 | 32.93 | 503.81 |
| Bidder5 | 33.69 | 549.07 | 33.70 | 549.31 |
| TOTAL | 100.00 | 1,465.18 | 100.0 | 1,464.67 |

**Table 4.3:** Comparison of the analytical solution of the constrained maximisation problem with the result derived from the simulation experiment with $\epsilon = 0.01$.

## ▸ 4.3 Truthful Bidding in a Multi-Auction Market

The strong properties of the PSP auction, such as incentive compatibility and efficiency, suggests its use in extended scenarios for resource allocation in communication networks. One direction of research has been the so-called "network case", in which the PSP auction is implemented at each network node to allocate resources for incoming and outgoing flows. Users wishing to reserve an end-to-end connection along a specific path are required to submit bids to each auction to secure resources. The bottleneck node, which either experiences the highest demand or users with high valuation for resources, determines the size of the end-to-end pipe. The basic model for the network case has been introduced by the original work of Lazar and Semret (1998). Recent research has extended such work in multiple directions. First, the basic bidding strategy has been complemented by more sophisticated models (Bitsaki et al., 2005). Second, the multi-bid auction format, which is closely related to the PSP auction, has been extended to the network case (Maillé, 2005; Maillé and Tuffin, 2006).

Another direction for research, which we are following, is the "access case", in which multiple auctioneers offer resources for accessing the core network. Users may be able to either select the auction with lowest demand or to bundle resources from multiple auctions.

A prominent question connected with the access case is to design bidding strategies for bidders with different objectives. While rationality as the main motive can still be assumed the concept of truthful revelation cannot directly be transmitted to the case when multiple access options are available. While each auction itself is incentive-compatible in the sense that truth-telling is best for each bidder, bidders now have to decide how to distribute their demand. In this section we present a bidding strategy that resembles truth-telling for the multi-auctioneer case if resources can be bundled from multiple auctions.

In the following we assume that all auctions run independently and no central coordination exists. All auctioning mechanisms used by the sellers are defined to be identical from their allocation rule $A$ using the PSP allocation mechanism as in the single-seller case. All resources are defined as homogeneous, meaning that all units from all sellers are assumed to belong to the same service class with identical quality parameters.

Most of the material in this section has been published in Roggendorf and Beltran (2006a) and Beltran and Roggendorf (2007).

### ▸ 4.3.1 The application of the multi-auction concept to a scenario of competitive wireless access networks

The access case we are interested in reflects the situation of wireless access in which users can dynamically decide on the network they use to connect to the core. While this is, in principle, also possible in wired networks the application is much more direct in wireless network structures as no physical connection or central authority, which allocates resources on behalf of the providers, is needed.

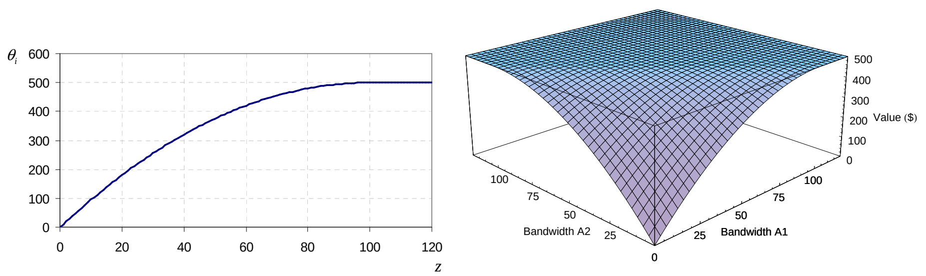
With flow-based resource allocation, as implemented by the PSP auction, users are able to optimise their ”resource portfolio” as soon as new information about the network situation is available. If a user experiences high congestion in one wireless access network he is able to switch his demand to other networks which may be available.

From a practical viewpoint multiple factors can determine the level of competition for a specific location. First, wireless access networks of different providers may differ in reach and capacity. Depending on the location, a user may be able to have access to two or more networks running a continuous PSP auction to allocate their resources. Second, the wireless technology used by the provider needs to match the capabilities of the mobile user terminals. Finally, a user may need a contractual relationship with the provider to gain access to the access network. This may be organised over a broker model, in which a wireless resource broker holds the contractual relationship and offers users flexible access to such network. Alternatively, users may be able to build up short-term billing relationships with providers.

## ▶ 4.3.2 Truthful behaviour with multiple auctions

When players are faced with more than one mechanism, identifying the incentive-compatible strategy becomes more complex. A player $i$ needs to coordinate his bids on all available auctions. In principle, it is possible to source the same amount of resources with an infinite number of combinations. Before we proceed with a formal description of bidding in a multi-auctioneer environment, we use a two-auctioneer scenario for a graphical description.

Figure 4.10(a) shows a player's parabolic valuation function with $\overline{q}_i = 100$ and $\overline{p}_i = 10$. With these parameters the valuation function takes the form $\theta_i(z) = -\frac{1}{20}z^2 + 10z$ for $z \leq 100$ and $\theta_i(z) = 500$ thereafter. Figure 4.10(b) uses this valuation function for a two-auctioneer scenario. The same value level can be achieved with an infinite number of combinations sourced from both auctioneers. Since a bidder has full information about the opponent

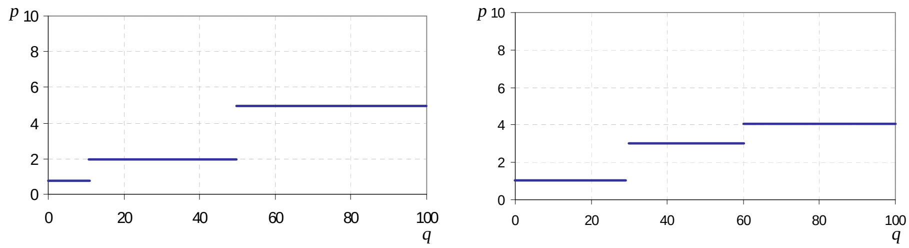



(a) Valuation function $\theta_i(z) = -\frac{1}{20}z^2 + 10z$ for player $i$.

(b) Valuation function for a two-auctioneer scenario.

**Figure 4.10:** User valuation.

bid profiles from the last round at all auctioneers and is aware of the allocation rule of the mechanism, it can calculate a cost function for both auctions by integrating over the bid profiles.[8] The bid profiles used here for the example are shown in Figure 4.11(a) and 4.11(b). The derived cost function for the two-auctioneer scenario is shown in Figure 4.12.




(a) Opponent bid profile for auction 1. (b) Opponent bid profile for auction 2.

**Figure 4.11:** User bid profiles.

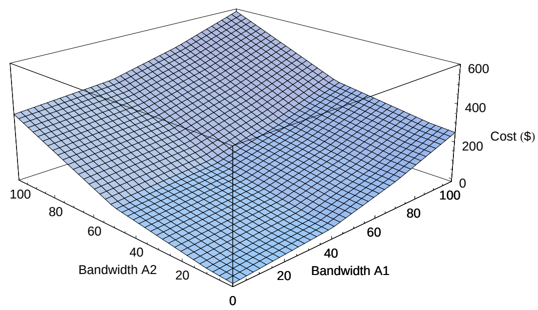



**Figure 4.12:** The cost function derived from the opponent bid profiles.

From these two functions a player can derive his utility function for each combination of resources from both providers by substracting the cost function from the valuation function (Figure 4.13). In this way, a bidder can identify the utility-maximising combination of quantities from both auctions. While the graphical representation helps to understand the principle optimisation problem in the case of multiple auctioneers, a formal description of the bidding strategy is needed. In the following section we describe the optimal bidding strategy and examine its main properties.

### ► 4.3.3 The BalancedBid bidding strategy

We study the formal equilibrium characteristics of multiple independently-managed auctions when bidders are allowed to split their bids in order to aggregate shares of the re-

[8]From the bids of the other players a bidder can learn which other bidders he is going to exclude from the market by his own presence. The cost function will therefore be a stepwise-linear function, which is increasing in slope with each new bidder excluded from the market.

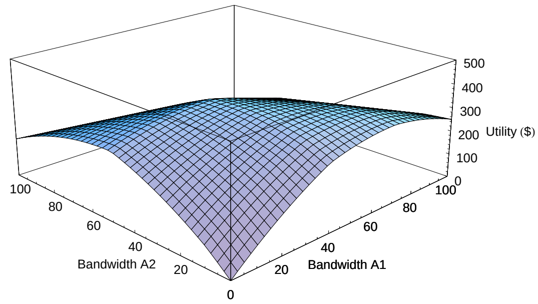



**Figure 4.13:** The utility function derived from the valuation function and the cost function showing the resulting utility for all combinations of resources.

source from several auctions. Assuming PSP is used by each auctioneer, we are concerned with the convergence to equilibrium at each auction when every bidder uses a bidding strategy known as *BalancedBid* (Roggendorf and Beltran, 2006a).

Before presenting the bidding strategy in detail and discussing its main properties some additional notation is introduced.

Let us assume the set of bidders is $1,..,I$, the set of auctions is $1,..,J$ and $Q^{(j)}$ units are being sold at auction $j$. Bidder $i$'s bid on auction $j$ is $s_i^{(j)} = (q_i^{(j)}, p_i^{(j)}) \in S_i^{(j)} = [0, Q^{(j)}] \times [0, \bar{P}]$ with $\bar{P}$ an upper bound on the unit price. Let us also define $S_i = \prod_j S_i^{(j)}$ as the set of all possible bids of bidder $i$ at the auctions. A composite or split bid $s_i = (s_i^{(1)}, .., s_i^{(J)})$ from bidder $i$ is a point in $S_i$. We assume that bidder i submits a bid to all or some of the auctions at the same time.

To model a reserve price in each auction an additional player $i = 0$ is introduced, which bids with a fixed valuation $\theta_0 = p_0 Q^{(j)}$. This bidder is present in each auction and $p_0$ determines the minimum bid needed by other bidders to receive resources.

In the same fashion as $s_i^{(j)}$, bidder $i$'s opponent bid profile at auction $j$ can be written as

$$s_{-i}^{(j)} = [s_0^{(j)}, .., s_{i-1}^{(j)}, s_{i+1}^{(j)}, .., s_I^{(j)}] \in S_{-i}^{(j)} = \prod_{n \neq i} S_n^{(j)},$$

where $S_{-i}^{(j)}$ is a set of all possible opponent bid profiles, consisting of the bids submitted by all bidders $n \neq i$ at the last (most recent) auction round. We summarise all opponent bid profiles from all auctions into a matrix $s_{-i}$, with

$$s_{-i} = \begin{pmatrix} s_{-i}^{(1)} \\ .. \\ s_{-i}^{(J)} \end{pmatrix} = \begin{pmatrix} s_0^{(1)}, & .., & s_{i-1}^{(1)}, & s_{i+1}^{(1)}, & .., & s_I^{(1)} \\ .. & & & & & \\ s_0^{(J)}, & .., & s_{i-1}^{(J)}, & s_{i+1}^{(J)}, & .., & s_I^{(J)} \end{pmatrix}$$

$$= \begin{pmatrix} (q_0^{(1)}, p_0^{(1)}), & .., (q_{i-1}^{(1)}, p_{i-1}^{(1)}), & (q_{i+1}^{(1)}, p_{i+1}^{(1)}), & .., & (q_I^{(1)}, p_I^{(1)}) \\ .. & & & & \\ (q_0^{(J)}, p_0^{(J)}), & .., (q_{i-1}^{(J)}, p_{i-1}^{(J)}), & (q_{i+1}^{(J)}, p_{i+1}^{(J)}), & .., & (q_I^{(J)}, p_I^{(J)}) \end{pmatrix},$$

and define $S_{-i} = \prod_{j=\{1,..,J\}} S_{-i}^{(j)}$.

As in the one-auctioneer case a player $i$ has a valuation function $\theta_i$ for the resource. If, at any given time, $i$ has been allocated shares $a_i^{(1)},..,a_i^{(J)}$ from the $J$ auctions and is supposed to pay $c_i^{(1)},..,c_i^{(J)}$ to every individual auctioneer $j$, his utility is given by $u_i(a_i, c_i) = \theta_i(\sum_{j=1}^{J} a_i^{(j)}) - \sum_{j=1}^{J} c_i^{(j)}$, that is, the valuation for the sum of all resources obtained from all auctions minus the costs from each auction.

With this notation we can define the game played by a bidder with access to multiple PSP auctions, which is identical to the one-auctioneer case except that the strategy space has been expanded by one dimension.[9]

**Definition 4.7** (A game of multiple PSP auctions)**.** *The normal-form representation G of the auction game, which is played by the bidders with access to multiple PSP auctions is given by:*

$$G = (S_1, .., S_I, u_1, .., u_I),$$

*with $S_i$ being the strategy space of player i defined as $S_i = \prod_j S_i^{(j)}$, and $u_i$ being the utility of player i.*

As with the one-auctioneer case we are interested in finding a Nash equilibrium of the game given in Definition 4.4 under complete information. In contrast to the single-auction case the problem is now to identify the strategy of player $i$, consisting of $J$ bids to be submitted to each auction $j$.

To let the auctions converge in finite time we use Semret's notion of an $\epsilon$-Nash equilibrium 4.5, which allows a bidder to stop updating a bid profile once the difference in utility provided by the current composite bid and the next one is less than $\epsilon$. Then, the set of $\epsilon$-best replies is defined as:

$$S^{\epsilon}(s) = \{s_i \in S_i(s_{-i}) : u_i(s_i; s_{-i}) \geq u_i(s_i'; s_{-i}) - \epsilon, \forall s_i' \in S_i(s_{-i})\}.$$

A $2\epsilon$-Nash equilibrium is a fixed point of $S^{\epsilon}$.

### The aggregated market and the $\epsilon$-best reply

To find the utility-maximising bid combination for all auctions we need to introduce an alternative view of the market possibilities any bidder faces. Loosely speaking, an aggregated market "mimics" the behaviour of the individual auctions by defining a resource quantity and an allocation rule. The resource quantity is the sum of all resource quantities offered at all auctions. Our goal is to study the dynamics at each single auctions through our observation of an artificial market that would aggregate quantities as well as bids. In order to understand what the aggregated market is and how it functions, we introduce the utility-optimising $\epsilon$-best reply to such a market and prove its main

[9] All allocation rules $A^{(j)}$ for each individual auction $j$ are identical to the original PSP auction as defined in Semret (1999).

properties. In a second step we propose a way in which a bidder can split his bid into the individual auctions.

To determine the utility-optimising $\epsilon$-best reply to the market, a bidder $i$ has to consider his opponent bid profiles $s_{-i}^{(j)}$ from all auctions $J$. To merge all bid profiles into a common *opponent market bid profile* we propose the following procedure.

- For each auction $j$, create a vector called the *opponent winning bid profile* from auction $j$, with
$$r_{-i}^{(j)} = [(a_0^{(j)}, p_0^{(j)}), .., (a_{i-1}^{(j)}, p_{i-1}^{(j)}), (a_{i+1}^{(j)}, p_{i+1}^{(j)}), ..(a_I^{(j)}, p_I^{j})].$$
The value for all $a_i^{(j)}$'s can be derived with $a_n^{(j)} = \left[Q^{(j)} - \sum_{k \neq n; p_k^{(j)} > p_n^{(j)}}^{I} q_k^{(j)}\right]^+$, which is the PSP allocation rule. Since the bid representing the reserve price is defined as $(Q^{(j)}, p_0^{(j)})$, we can safely assume that $\sum q_i^{(j)} \geq Q^{(j)}\ \forall j$. Therefore, $\sum_{i=0}^{I} a_n^{(j)} = Q^{(j)}$.

- Merge all vectors $r_{-i}^{(j)}$ into a common matrix $r_{-i}$, defined as

$$r_{-i} = \begin{pmatrix} r_{-i}^{(1)} \\ .. \\ r_{-i}^{(J)} \end{pmatrix} = \begin{pmatrix} (a_0^{(1)}, p_0^{(1)}), & .., (a_{i-1}^{(1)}, p_{i-1}^{(1)}), & (a_{i+1}^{(1)}, p_{i+1}^{(1)}), & .., & (a_I^{(1)}, p_I^{(1)}) \\ .. & & & & \\ (a_0^{(J)}, p_0^{(J)}), & .., (a_{i-1}^{(J)}, p_{i-1}^{(J)}), & (a_{i+1}^{(J)}, p_{i+1}^{(J)}), & .., & (a_I^{(J)}, p_I^{(J)}) \end{pmatrix}.$$

In contrast to the matrix $s_{-i}$ the new matrix $r_{-i}$ contains all "winning" shares of the opponents' bids together with the unit-price if player $i$ were not present in the auction. Therefore, $r_{-i}$ can be seen as the opponent bid profile, which includes the capacity constraints in each market. By only including the winning shares a player is able to evaluate how much capacity in total can be obtained from all markets with his individual valuation for the sum of resources.

To gain a better understanding why we need to redefine $s_{-i}$ to create the "opponent market bid profile" a simple example is presented.

**Example 4.2.** *Consider a scenario with two auctions and total resources of $Q^{(1)} = Q^{(2)} = 10$ and three bidders, all of whom have access to both auctions. Now, consider the situation of player 1, which demand function is given by $\theta'(q) = 10 - q$. From each auction the player receives an opponents' bid profile $s_{-1}^{(1)} = [(9, 11), (14, 10)]$ and $s_{-1}^{(2)} = [(8, 2), (14, 1)]$, respectively. To calculate his truthful reply to the market he needs to translate both opponent bid profiles into a common "opponent market bid profile". By just merging both opponent bid profiles he derives $(s_{-1}^{(1)}, s_{-1}^{(2)}) = s_{-1} = [(9, 11), (14, 10), (8, 2), (14, 1)]$. With this opponent bid profile and the overall resources available in the market given by $Q = \sum_{j=1}^{2} Q^{(j)} = 20$, he can derive his truthful reply to be $t_1 = (0, 10)$ (Figure 4.14). However, the merged market bid profile does not consider the constraints given by the distribution of resources between the two auctions.*[10] *Since in auction 1 unit-prices are very high but the total capacity of auction 1 is 10, and in contrast, unit-prices are*

[10]The opponent bid profile $s_{-1}$ implies that the first two bids win a positive amount while the other two bids are loosing bids. This is not what we need since the allocation is constrained by the capacity in both auctions and not by the overall capacity constraint given by the $Q$.

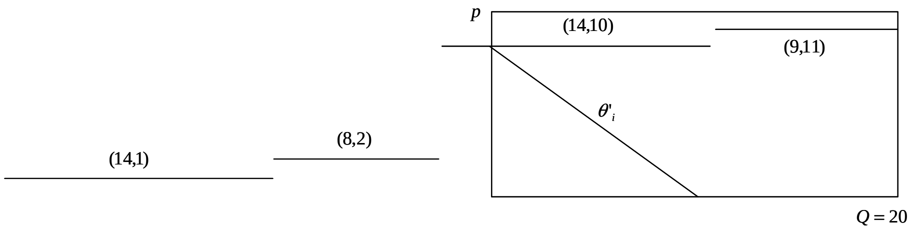



**Figure 4.14:** Graphical representation of $s_{-i}$.

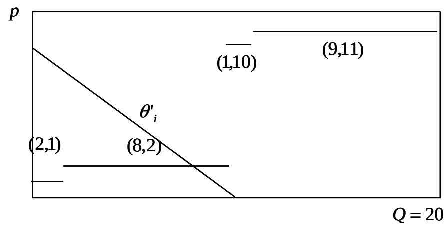



**Figure 4.15:** Graphical representation of $r$.

*very low in auction 2, player 1 will be able to acquire resources from the second market. Therefore, the opponent bid profiles need to be redefined in order to reflect the individual constraints in each auction. By using the above procedure we derive* $r_{-1} = [(9,11),(1,10),(8,2),(2,1)]$. *By using* $r_{-1}$ *for calculating a utility-maximising bid we derive* $t_1 = (8,2)$ *(Figure 4.15). This bid reflects the player's truthful reply to the market because it considers the resource constraints from both auctions.*

With the definition of $r_{-i}$ we can define the aggregated market.

**Definition 4.8** (The aggregated market)**.** *An aggregated market can be defined by a resource Q and an allocation rule A with*

$$Q = \sum_{j=1}^{J} Q^{(j)},$$

*and*

$$\begin{array}{rrcl} A: & S & \longrightarrow & S \\ & s = (q,p) & \longmapsto & A(s) = (a(s), c(s)), \end{array}$$

*with* $S = \prod_{i \in I} S_i$. *The allocation rule follows the PSP allocation rule of a single auction but uses the matrix* $r_{-i}$ *as the opponent bid profile. The i-th row of* $A(s)$ *is the allocation to player i, with* $a_i(s)$ *being the quantity and* $c_i(s)$ *being the overall cost.*

$$a_i(s_i, r_{-i}) = q_i \wedge \underline{Q}_{-i}(p_i, r_{-i}),$$

*with* $\underline{Q}_{-i}(y, r_{-i}) = \left[ Q - \sum_{k=0, j=1, p_k > y}^{I,J} a_k^{(j)} \right]^+.$

*$c_i(s)$ cannot be derived directly from $r_{-i}$ but only once the composite or split bids to the individual auctions have been defined.*

In the following analysis we limit the strategy space to $T_i = \left\{s_i \in S_i : q_i = \sum^{(J)} q_i^{(j)}; p_i = \theta'(q_i)\right\}$ and search for a bid $t_i \in T_i$ as the truthful market reply.

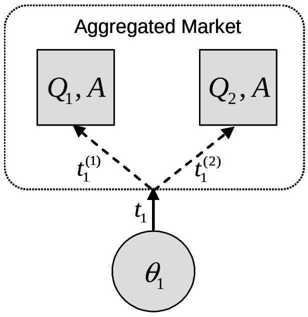



**Figure 4.16:** Graphical representation of the aggregated market bid $t_i$ and the bids $t_i^{(j)}$ to the single auctions.

We define truthful bidding in a way so that the sum of resource quantities a bidder expresses on all auctions at a given unit price corresponds to his demand function. While with this definition the truthful demand is not "visible" to a single auctioneer, the bidder ensures that he does not overbid in the aggregate market. Figure 4.16 expresses the difference between the aggregated market bid $t_i$ and the bids $t_i^{(j)}$ to the single auctions. The aggregated bid is a virtual construct and is only used internally by each bidder to derive the split bids to the auctions to ensure that the overall demand expressed in the market corresponds to the bidder's demand.

**Definition 4.9** (Aggregated market bid)**.** *Under assumption 4.1, a truthful reply to the market $t_i = (v_i, w_i) \in T_i \cap S_i^{\epsilon}(s_{-i})$ is given by*

$$v_i = \left[\sup G_i(r_{-i}) - \frac{\epsilon}{\theta_i'(0)}\right]^+ \quad \textit{and } w_i = \theta_i'(v_i)$$

*with $\epsilon/\theta_i'(0)$ being a factor by which each bid is reduced to avoid ties in bidding price between players and therefore, the utility from the aggregated market increases by at least $\epsilon$,*[11]
*where*

$$G_i(r_{-i}) = \left\{ z \in [0, \sum_{j=1}^{J} Q^{(j)}] : z \le Q_i(\theta'(z), r_{-i}) \right\},$$

*with*

$$Q_i(y, r_{-i}) = \left[\sum_{j=1}^{J} Q^{(j)} - \sum_{p_k^{(j)} > y}^{I,J} a_k^{(j)}\right]^+$$

[11]Note that this value has been interpreted as *bid fee* by Semret (1999). However, in subsequent work (Maillé, 2003), (Maillé and Tuffin, 2004a) it has been noted that this interpretation is potentially misleading.

*Additionally, we define* $\bar{v}_i = [\sup G_i(r_{-i})]$ *and* $\bar{w}_i = \theta_i'(v_i)$ *to form* $\bar{t}_i = (\bar{v}_i, \bar{w}_i)$.

Definition 4.9 is visualised in Figure 4.17.

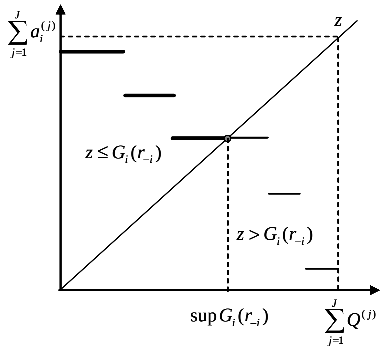



**Figure 4.17:** The graphical representation of $G_i(r_{-i})$ and $\sup G_i(r_{-i})$.

### Bid splitting and the BalancedBid strategy

One key property of the single PSP auction is incentive-compatibility. We aim at understanding if bidders have incentives to reveal their true valuation in a market consisting of multiple PSP auctions. We therefore first present the bidding strategy, *BalancedBid* (Roggendorf and Beltran, 2006a), which mimics truthful bidding when bidders are allowed to split their bids to bundle resources from several auctions. We then show that this bidding strategy is the myopic $\epsilon$-best response for all bidders in such a situation.

Once the truthful reply to the market has been determined, it is split into bids to be submitted to the individual auctions. We refer to such bids as *BalancedBids*. A balanced bid is defined as $t_i^{(j)} = (v_i^{(j)}, w_i)$, with $v_i^{(j)}$ being the quantity and $w_i$ being the unit price, bid at auction $j$, which has been derived from the (aggregated) market bid $t_i = (\sum^J v_i^{(j)}, w_i)$. The intuition behind the allocation rule at an auction $j$ is to sum all bids, received at $j$, which can be "beaten" by the unit price $\bar{w}_i$ determined by the market bid. [12] Two cases have to be distinguished; in Case 1 the demand function $\theta_i'$ does not intersect any of the bid steps. Case 2 describes the situation where the demand function crosses through one of the bid steps. Formally, we can state the two cases by defining

$$\alpha_i \equiv \sum_{m=0, n=1; p_n \leq \bar{w}_i}^{I,J} a_n^{(m)} - \bar{v}_i.$$

Now

$$\text{Case 1:} \quad \alpha_i = 0$$
$$\text{Case 2:} \quad \alpha_i > 0.$$

[12] Note that we use $\bar{t}_i$ for this process as it gives us the truthful reply without the reduction by $\sum^J \epsilon/\theta_i'(0)$. Otherwise, we would need to define a second constraint, which limits the balanced bids to the total quantity $v_i$.

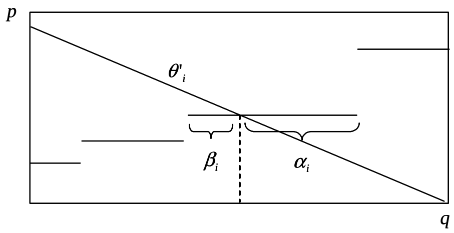



**Figure 4.18:** Visualisation of $\alpha_i$ and $\beta_i$.

We define $\beta_i \equiv \sum_{m=0,n=1;p_n=\bar{w}_i}^{I,J} a_n^{(m)} - \alpha_i$ as the share of the sum of all bids with unit price $\bar{w}_i$, which lies within the valuation of $i$ (see Figure 4.18). This segment can consist of multiple bids from different players and auctions. The following definition for deriving the single bids to the auctions is called *BalancedBid*

**Definition 4.10** (BalancedBid). *(Roggendorf and Beltran, 2006a) Under Assumption 4.1 and Definition 4.9 the balanced bid $t_i^{(j)} = (v_i^{(j)}, w_i)$ for player i on auction j is given by*

$$v_i^{(j)} = \left[ \sum_{n=0,p_n^{(j)}<\bar{w}_i}^{I} a_n^{(j)} + \beta_i \frac{\sum_{n=0,p_n^{(j)}=\bar{w}_i}^{I} a_n^{(j)}}{\sum_{m=0,n=1;p_n^{(m)}=\bar{w}_i}^{I,J} a_n^{(m)}} - \frac{\epsilon}{J\theta_i'(0)} \right]^+ \quad \textit{and } w_i = \theta_i(\sum_{j=1}^{J} v_i^{(j)}) = v_i.$$

*Each balanced bid is reduced by $\epsilon / J\theta_i'(0)$ to ensure that no ties between bidders can occur and that the utility derived from the sum of the bids from the J auctions is increased by at least $\epsilon$.*

One important question to ask is why a bidder should bid with identical unit prices on all auctions and not reduce the unit price to a level where he still wins the amount $q_i^{(j)}$. The reason for this can be found in the pricing rule of the PSP auction. Since a player is only charged with the cost of excluding other players from the market, the unit price does not influence the final charges. Since this unit price reflects the valuation of the total resources gained from the multi-auction market a player uses this price on all auctions.

We can now derive the cost of bid $t_i$, which consists of the sub-bids $t_i^{(j)}$ on each auction. Since we need to consider the resource constraints on each auction we have to calculate partial costs separately. For each auction, costs can be derived by calculating which demand has been excluded from receiving a positive allocation by the presence of bidder $i$. Formally, the costs are given by

$$c_i(t_i, s_{-i}) = \sum_{j=0}^{J} \int_0^{a_i(t_i^{(j)}, s_{-i}^{(j)})} P_i(z, s_{-i}^{(j)}) dz,$$

where

$$P_i(z, s_{-i}^{(j)}) = \inf\{y \geq 0 : Q_i^{(j)}(y, s_{-i}) \geq z\}$$

is the stair-case function $Q_i^{(j)}$ flipped by 90 degrees.

### The BalancedBid bidding algorithm

In the previous part the *BalancedBid* strategy has been analytically defined. To implement the bidding strategy we provide the computational algorithm.

**Algorithm 1** *BalancedBid* Bidding Strategy

Let $s_i^{(j)} = 0$ and $\hat{s}_{-i}^{(j)} = \emptyset$ for all auctions $j$.
**loop**
  Update list of active auctions and receive updates of $\hat{s}_{-i}^{(j)}$ from each auction $j$.
  Compute a truthful aggregated market bid $t_i = (v_i, w_i)$ according to Definition 4.9.
  Compute the balanced bids $t_i^{(j)} = (v_i^{(j)}, w_i)$ for each auction $j$ according to Definition 4.10.
  Compute the overall utility of the aggregated market bid with $u_i\left(s^{(1)}, ...s^{(J)}\right) = \theta_i\left(\sum_{j=1}^{J} a_i^{(j)}(s^{(j)})\right) - \sum_{j=1}^{J} c_i^k(s^{(j)})$.
  **if** $u(t_i; s_{-i}) > u(s_i; s_{-i}) + \epsilon$ **then**
    Send the bids $t_i^{(j)} = (v_i^{(j)}, w_i)$ to each auction $j$.
  **end if**
  Sleep for 1 second.
**end loop**

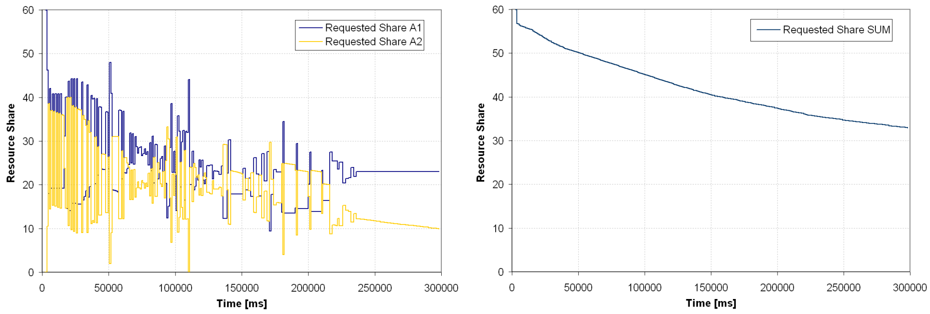
### A simple example using the BalancedBid strategy

To explain the basic workings of the *BalancedBid* bidding strategy we present an example in a simple setting. We use the identical player types as defined in Table 4.2 but introduce an additional auction to which all players have access in addition to the already existing auction. We define $Q^{(1)} = 60$ and $Q^{(2)} = 40$ and set $\epsilon = 0.01$. All bidders have access to both auctions.

We first look into the bidding behaviour of a particular bidder. Figure 4.19(b) depicts the requested shares of player 4 for both auctions. Even if the requested shares for each auction changes rapidly at each bidding step, the added share steadily decreases until an equilibrium is found (Figure 4.19(b)).

Table 4.4 provides the allocation of resources for each player in equilibrium. When adding up the resources players have won on both auctions, the result is very close to the analytically optimal solution (see Table 4.3). This indicates that the *BalancedBid* strategy is the direct extension of the truthful revelation in the one-seller case and is able to efficiently allocate resources from multiple auctions.

Figure 4.20(a) shows the revenue generated by both auctions during the convergence process. Both graphs show the typical shape with increasing revenue up to a certain point and decreasing values afterwards until equilibrium is reached. In equilibrium the revenue generated by both auctions is close to zero. The reason for this has already been explained in Section 4.2.




(a) Requested shares $q_4^{(1)}$ and $q_4^{(2)}$ over the convergence period.

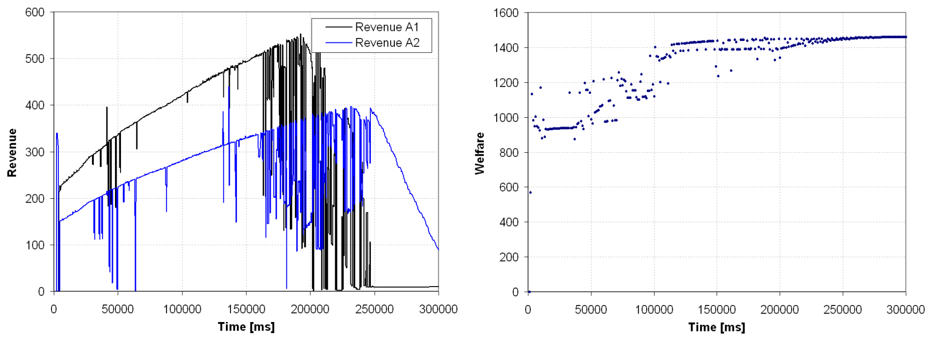
(b) Sum of requested shares $q_4^{(1)} + q_4^{(2)}$ over the convergence period.

**Figure 4.19:** Requested shares over player 4 with the *BalancedBid* bidding strategy.

| | Allocation from A1 | Allocation from A2 | Total allocation |
|---|---|---|---|
| Bidder1 | 0.11 | 0 | 0.11 |
| Bidder2 | 0.72 | 9.17 | 9.89 |
| Bidder3 | 13.98 | 9.51 | 23.49 |
| Bidder4 | 23.07 | 9.84 | 32.91 |
| Bidder5 | 22.22 | 11.38 | 33.60 |
| TOTAL | 60.00 | 40.00 | 100.00 |

**Table 4.4:** Resource allocation in equilibrium for the 5 player example.

The analysis of the social welfare generated by the allocation reveals similar characteristics as in the one-seller case (Figure 4.9(b)). Social welfare quickly increases at the beginning of the convergence process and is maximised in equilibrium. The final allocation efficiency reaches $W = 1,462.78$, which is only 0.3% lower than the maximum social welfare derived analytically in the single-auction case.




(a) Revenue for auction $A1$ and $A2$

(b) Total welfare of the system

**Figure 4.20:** Revenue and Welfare generation with the *BalancedBid* bidding strategy

### Incentive-compatibility in a market consisting of multiple PSP auctions

While it has been shown by Semret (1999) that the PSP auction is incentive-compatible the question remains if bidders still have an incentive for truthful bidding at the aggregated market. The splitting of a bid defined by *BalancedBid* seems to resemble truthful bidding in such a market. We need to show that this strategy is the best strategy for a myopic player compared to all other possible strategies to bid on multiple auctions.

Especially, we need to prove that there does not exist any other $s_i \in S_i(s_{-i})$, which results in higher utility for player $i$. Since $s_i$ has been redefined to contain a vector of bids (to the different auctions) instead of only one bid we need to define $a_i(s) = \sum^J a_i^{(j)}(s_i^{(j)}; s_{-i}^{(j)})$.

**Proposition 4.1** (Incentive compatibility). *Under assumption 4.1, $\forall i \in I$, $\forall j \in J$, $\forall s_{-i}^{(j)} \in S_{-i}^{(j)}$, so that $Q_i(0, r_{-i}) = 0$, for any $\epsilon > 0$, the truthful reply $t_i$ defined in Definition 4.9 is an $\epsilon$-best reply to the market.*

In the following we will show that for any $s_i \in S_i$ the yielded utility is equal or less to the utility gained from bidding $t_i$.

*Proof.* : (Incentive-compatibility)
$\forall s_i \in S_i(s_{-i})$,

$$\begin{aligned} u_i(t_i; s_{-i}) - u_i(s) &= \theta(a_i(t_i; s_{-i})) - c_i(t_i; s_{-i}) - [\theta_i(a_i(s)) - c_i(s)] && (4.4)\\ &= \int_{a_i(t_i; s_{-i})}^{a_i(s)} \left[P_i(z, r_{-i}) - \theta_i'(z)\right] dz. && (4.5) \end{aligned}$$

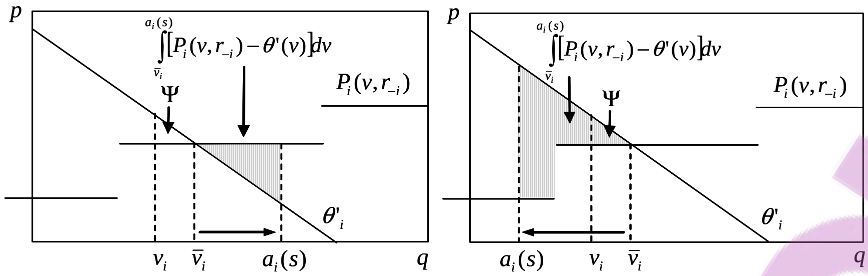
Equation (4.4) is a consequence of the continuity of $\theta$. Note that the integral is always nonnegative because $\theta'(q)$ is non-increasing in $q$ and $P_i(q)$ is non-decreasing in $q$. We can rewrite this property as $a_i(t_i, s_{-i}) = v_i$ and divide the integral in two parts by using $\bar{v}_i$.

$$\begin{aligned} &= \int_{\bar{v}_i}^{a_i(s)} \left[P_i(z, r_{-i}) - \theta_i'(z)\right] dz + \int_{v_i}^{\bar{v}_i} \left[P_i(z, r_{-i}) - \theta_i'(z)\right] dz && (4.6)\\ &\geq \int_{\bar{v}_i}^{a_i(s)} \left[P_i(z, r_{-i}) - \theta_i'(z)\right] dz - \epsilon. && (4.7) \end{aligned}$$

The inequality from (4.6) to (4.7) follows from the upper bound of $(\bar{v}_i - v_i) \leq \epsilon / \theta_i'(0)$ and from the fact that $\theta_i'$ is non-increasing in $q$. Since $\epsilon$ is always positive we now need to show that bidding any value $v \notin [v_i, \bar{v}_i]$ and within the boundaries of $a_i(s)$ yields a utility $< \epsilon$ compared to the utility obtained by bidding $v = \bar{v}_i$. We graphically illustrate that this is the case.

Figure 4.21(a) depicts case 1 with $\bar{v}_i < a_i(s)$ and in Figure 4.21(b) the second case is shown when $\bar{v}_i \geq a_i(s)$. While in the first case the integrand is positive it is negative in the second case. But since the integral is calculated with switched boundaries $\bar{v}_i > v$, the integral turns to be positive.

**Case 1:** ($\bar{v}_i < a_i(s)$): Take any $v \in (\bar{v}_i, a_i(s)]$. By definition of $v_i$, $v$ is not part of $G_i(r_{-i})$. This leads to the conclusion that $c_i(s) = \int_0^{a_i(s)} P_i(\eta, r_{-i}) d\eta \geq \int_0^{v} P_i(\eta, r_{-i}) d\eta$. Therefore, $v$ must be larger than $Q_i(\theta'(v))$. $P_i(v, r_{-i})$ has been defined as $P_i(v, r_{-i}) = \inf\{y \geq 0 : Q_i(y, r_{-i}) \geq v\}$.




(a) Case 1: $\bar{v}_i < a_i(s)$).

(b) Case 2: $\bar{v}_i \geq a_i(s)$).

**Figure 4.21:** The graphical representation of the two cases for any value $z \notin [v_i, \bar{v}_i]$.

Therefore, with a fixed opponent market bid profile $r_{-i}$, $\forall y$ and $v \geq 0$,

$$v \leq Q_i(y, r_{-i}) \Rightarrow y \geq P_i(v, r_{-i}) \tag{4.8}$$

and

$$y > P_i(v, r_{-i}) \Rightarrow v \leq Q_i(y, r_{-i}) \tag{4.9}$$

Because $v > Q_i(\theta'_i(v))$, Equation 4.9 implies that $\theta'(v) \leq P_i(v)$, which proves that the integrand is $\geq 0$.

**Case 2:** ($\bar{v}_i > a_i(s)$): Take any $v \geq a_i(s)$. Because $\theta'_i$ is non-increasing, $Q_i(., r_{-i})$ is non-decreasing and $P_i(., r_{-i}) \geq 0$, any point to the left of $\bar{v}_i$ is in the set of $G_i(r_{-i}), \forall v < v_i, v \in G_i(s_{-i})$. Therefore, we have $v \leq Q_i(\theta'_i(v), r_{-i})$. By Equation 4.9 $\theta'_i(v) \geq P_i(v, r_{-i})$, which shows that the integrand in Equation 4.6 is $\leq 0$. But since the integral is calculated from right to the left, the integral becomes $\geq 0$.

$\square$

## Continuity of the best reply to the market

Another property of the best market reply, which is required to show that the proposed bidding strategy is part of a Nash equilibrium is the continuity of the best reply $t_i$ in opponent bid profiles $r_{-i}$. Semret (1999) has shown the continuity of the $\epsilon$-best reply in $s_{-i}$ for the one-auctioneer case. It is argued that for all $i \in I$, the $\epsilon$-best reply $t_i$ is continuous in $s_{-i}$ on any subset $V_i(\overline{P}, \underline{P}) = \{s_{-i} \in S_{-i} : \forall z > 0, \overline{P} \geq P_i(z, s_{-i}) \geq \underline{P}\}$, with $\infty \geq \overline{P} \geq \underline{P} > 0$. It remains to proof this property for the proposed truthful bidding strategy on the set of opponent market profiles $R_{-i}$.

Before we can proceed we need to prove that $R_{-i}$ is a compact subset of $S_{-i}$. To analyse the properties of $r_{-i}$ we define the mapping $h$, which maps $s_{-i}$ to $r_{-i}$.

$$\begin{aligned} h: \quad & (S_{-i})^J \longrightarrow (S_{-i})^J \\ & s_{-i} \longmapsto h(s_{-i}) = r_{-i} \end{aligned}$$

and define a set $R_{-i}$ as

$$R_{-i} = \{r_{-i} | r_{-i} = h(s_{-i}) \text{ for some } s_{-i}^{(j)} \in S_{-i}\}.$$

**Lemma 4.1.** *$R_{-i}$ is a compact subset of $S_{-i}$, and is therefore bounded and closed.*

*Proof.* In a first step we only consider the first component $a_n^{(j)}$ of the mapping $h$, i.e.,

$$(q_0^{(j)}, \ldots, q_I^{(j)}) \mapsto a_n^{(j)}(q_0^{(j)}, \ldots, q_I^{(j)}) := \left[ Q^{(j)} - \sum_{k \neq n; p_k^{(j)} > p_n^{(j)}}^{I} q_k^{(j)} \right]^+ .$$

This map is continuous as the concatenation of continuous maps. We define

$$A_{-i} := \left\{ a_n^{(j)}(q_0^{(j)}, \ldots, q_I^{(j)}) \mid (q_0^{(j)}, \ldots, q_I^{(j)}) \in \left( [0, \max_j Q^{(j)}] \right)^I \right\}.$$

Since $A_{-i}$ is the continuous image of the compact set $\left([0, \max_j Q^{(j)}]\right)^I$, it is compact itself. Now the set $R_{-i}$ is given by

$$R_{-i} = (A_{-i} \times [0, \bar{P}])^{I \times J} .$$

Thus it is compact as the cartesian product of compact sets. □

**Proposition 4.2** (Continuity of the best reply to the aggregated market)**.** *The truthful bid to the aggregated market $t_i$ is continuous in $r_{-i}$ on a subset $V_i(\underline{P}, \overline{P}) = \{r_{-i} \in R_{-i} : \forall v > 0, \overline{P} \geq P_i(v, r_{-i}) \geq \underline{P}\}$, with $\infty > \overline{P} \geq \underline{P} > 0$.*

*Proof.* Since we have shown in Lemma 4.1 that $R_{-i}$ is a compact subset of $S_{-i}$, $t_i$ must be continuous in $r_{-i}$. We refer to the detailed proof in Semret (1999) for the continuity of $t_i$ in $s_{-i}$. □

The continuity of the best market reply does not necessarily mean that the bids to the single auctions are continuous in the opponent bid profiles $s_{-i}^{(j)}$. While we have first experienced the discontinuity during the first simulation experiments with the *BalancedBid* strategy, we have also been able to understand this analytically. Two main practical reasons can be identified. First, we consider the "symmetrical" case in which all players have access to all auctions. In this setting the reason for discontinuity in bids to single auctions lies in the asynchronism of information when forming the best reply to the market. If an opponent bid profile $s_{-i}^{(j)}$ arrives with a delay, a player may use bid profiles from different time periods to form the new reply. Therefore, he may switch his demand to one of the auctions with the more attractive opponents' bids. Extensive experimentation with an agent-based simulation platform (Roggendorf et al., 2006) has confirmed this behaviour in different kinds of situations. For example, it has been tested how bidders balance their demand when auction access is asymmetric for a large number of bidders present in the system.

The second reason for a large change in single bids can occur in a situation when not all players have access to all markets. In this case opponent bid profiles differ in the number of elements and players and can force a player to switch his demand rapidly from one auction to the other with only a small change in one opponent bid profile. The following example illustrates this case.

**Example 4.3** (Discontinuity in the reply to the single auctions). *Consider two auctions with* $Q = 10$ *and 3 players, all having access to both auctions. We now analyse the best reply of player* 1 *and the bid profiles* $s^{(1)} = s^{(2)} = ([8,2],[6,4],[5,6])$ *and* $s'^{(1)} = ([8,2],[5.9,4.1],[5,6])$*. The resulting opponent market bid profile is* $r_{-1} = ([5,6],[5,6],[5,4],[5,4])$ *and* $r'_{-1} = ([5,6],[5,6],[5,4.1],[5,4])$*. If we assume a demand function* $\theta'_1 = 10 - q$ *for player* 1*, we can derive the best reply to the market, resulting in* $t_1 = (5.9, 4.1)$ *and* $t'_1 = (5.8, 4.2)$.[13] *We can see that a small change in one of the bid profiles does let the best reply to the market change only marginally. However, if we calculate the bids to the auctions we derive* $t_1^{(1)} = (2.95, 4.1)$, $t_1^{(2)} = (2.95, 4.1)$ *and* $t_1'^{(1)} = (5, 4.2)$, $t_1'^{(2)} = (0.8, 4.2)$*, resulting in a large change in how the demand is distributed between auctions. Since now the first auction becomes more attractive to player* 1*, he shifts his demand to this market.*

### Nash equilibrium of the game

With Definition 4.4 we have already established the notion of Nash equilibrium in a general sense. We could also show that a market consisting of multiple PSP auctions is still incentive-compatible, namely incentivises players to reveal their true valuation to the aggregated market. The remaining question is if the equilibrium of the iterative game is of Nash type.

To restrict our attention to truthful bidding to the market as defined in (4.9) we set a reserve price $p_0 > 0$. This implies that for all $i \in I$, $Q_i(y, r_{-i}) = 0$, for all $y < p_0$. Then, Proposition 4.1 is fulfilled and allows us to restrict our attention to truthful bidding, which are still best replies to the aggregated market. As described by Semret (1999), this forms an embedded game within the larger game, with the strategy space being $T \subset S$, the feasible set for player $i$ being $T_i \cap S_i(s_{-i})$, and the best replies $X_i^\epsilon(s) = T_i \cap S_i^\epsilon(s)$. If we can find a fixed point of $X^\epsilon$ in $T$, this must also be a fixed point of $S^\epsilon$ in $S$.

**Proposition 4.3** (Efficient Nash equilibrium of the iterative game). *In the auction game consisting of multiple, independent PSP auctions, a reserve price* $p_0 > 0$*, and players finding their best reply to the aggregate market according to Definition 4.9 and Definition 4.10, if Assumption 4.1 holds, then for any* $\epsilon > 0$*, there exists a* $2\epsilon$*-Nash equilibrium* $s^* \in T$*.*

*Proof.* We now provide a sketch of the proof, which closely follows the work by Semret (1999).

We have shown the continuity of the truthful reply $t_i = (v_i, w_i)$ to the aggregate market in $r_{-i}$ on $R_{-i}$ in Proposition 4.2. Because $\theta'_i$ is continuous (by Assumption 4.1), $v_i(q_i, p_i) = v_i(q_i, \theta'_i(q_i))$ can be viewed as a continuous mapping of $[0, \sum Q^{(j)}]^I$ onto itself (for

[13]Assuming some $\epsilon$

reference, see Semret (1999)). It can now be shown by Brouwer's fixed point theorem that any continuous mapping of a convex compact set into itself has at least one fixed point ($\forall i, \exists q_i^* = v_i(q_i^*) \in [0, \sum Q^{(j)}]^I$). Therefore, an equilibrium in truthful strategies $s^* = t(s^*) \in T$ must exist.

□

In Maillé (2003) it is proved that the difference between the market clearing price and the maximum price bid by any bidder at a given round can be arbitrarily bounded. The bound is a linear function of the square root of $\epsilon$. If $\epsilon$ is sufficiently small, PSP provides an approximation to the market clearing price. In other words, prices set by bids at a $2\epsilon$-Nash equilibrium are arbitrarily close to prices at Nash equilibrium. The latter implies that the aggregated measure of welfare at a $\epsilon$-Nash equilibrium is also arbitrarily close to the aggregated measure of welfare at Nash equilibrium. Because the Nash equilibrium achieved by PSP is efficient, the iterative bidding process is arbitrarly close to the welfare of the efficient equilibrium.

### Properties of the individual PSP auctions in aggregate equilibrium

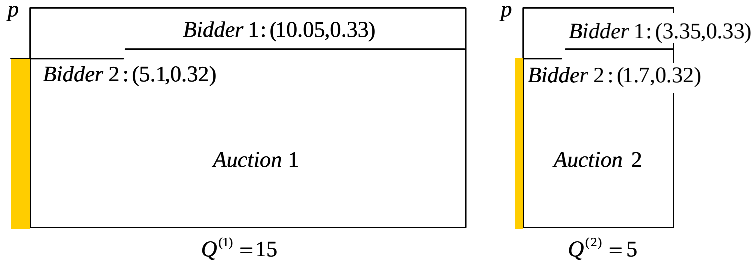
After having shown the properties of the market consisting of multiple PSP auctions we can now proceed in better understanding the implications of the *BalancedBid* strategy (Definition 4.10) to form the bids to the auctions. One property we have already explored is the discontinuity of $t_i^{(j)}$ in $S_{-i}^{(j)}$. This means that small changes in the opponent bid profile can lead to large changes in the balanced to the auctions. Additionally, in equilibrium, it may be possible to shift demand between auctions without disturbing the aggregate equilibrium. We show this by a simple example.

**Example 4.4.** *In this example we define two auctions with* $Q^{(1)} = 15$, $Q^{(2)} = 5$ *and two bidders, both having access to both auctions. The utility functions of both bidders are* $\theta_1' = -0.05q + 1$ *and* $\theta_1' = -0.1q + 1$, *respectively. With* $\epsilon$ *being small we can derive a possible* $\epsilon$*-Nash equilibrium of the aggregated market with the bids* $(13.4, 0.33)$ *for bidder 1 and* $(6.8, 0.32)$ *for bidder 2, resulting in the allocation* $(13.4, 0.064)$ *and* $(6.6, 0)$, *assuming that bidder 1 was the last updating his bid and that bidder 2 cannot further improving his surplus by updating his bid.*

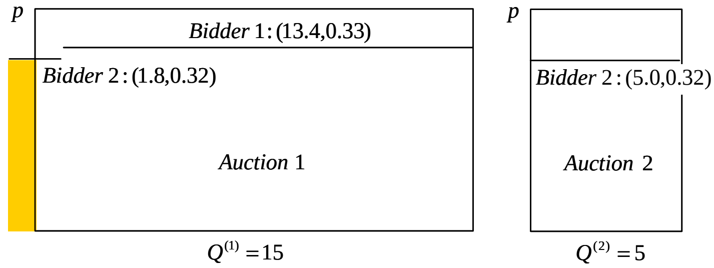
*With the equilibrium solution for the aggregate market being identified, a consequent problem is the definition of the balanced bids for both bidders. The BalancedBid strategy (as defined in Definition 4.10) distributes the bidder's demand proportionally to the bids equal in unit price, which cross the marginal demand of a bidder.*[14] *With our definition of BalancedBid strategy we proportionally allocate the quantity between the auctions, depending on the quantity of the opponent bid. However, other possible combinations of demand allocations lead to the same result if the aggregate market is in equilibrium.*

*Figure 4.22 depicts two possible combinations of demand allocations between the two auctions, leading to the same equilibrium result for the aggregated market as defined before. The*

[14] With only two bidders in the market there exist only two opponent bids coming from the same bidder, which therefore always cross the marginal demand function of the bidder.




(a) Case 1: Demand is equally distributed between auctions, as defined in Definition 4.10.




(b) Case 2: Demand is shifted so that bidder 1 receives resources from auction 1 and bidder 2 receives resources from auction 2.

**Figure 4.22:** Two different allocations of demand between auctions belonging to the same equilibrium solution for the aggregated market.

*first case shows how Definition 4.10 allocates the demand between the auctions (assuming zero delay in the distribution of opponent bid profiles), resulting in balanced bids defined as* $(10.05, 0.33), (3.35, 0.33)$*, and* $(5.1, 0.32), (1.7, 0.32)$*, respectively. Case 2 shows an alternative allocation in which bidder 1 shifts his demand to auction 1, while bidder 2 receives the remaining resources from auction 1 and the full resources from auction 2 (*$(13.4, 0.33), (0.0, 0.33)$*, and* $(1.8, 0.32), (5.0, 0.32)$*). Both combinations result in the same social welfare (total social welfare and individual surplus) and summed revenue for the auctioneers. However, in the first case, auctioneer 1 receives all revenues while in case 2, the revenue is distributed over both auctioneers.*

It needs to be mentioned that a shift in revenue from both providers to only one provider, as shown in the example, is irrelevant in the sense that PSP serves as a congestion-avoidance mechanism and is not intended to generate significant revenues in equilibrium. By choosing a sufficiently small $\epsilon$ the revenue in equilibrium will be small as users reduce their demand in order to avoid any congestion. Instead, by setting an appropriate reserve price, a provider is able to define a fixed charge for each unit of allocated resources in equilibrium.

The above example shows that it is possible to shift the demand of bidders between auctions without changing the fundamental results of the aggregated equilibrium. This also illustrates the difficulty in defining the notion of equilibrium in each single auction if users' preferences are expressed by a one-dimensional demand function. While we can speak of a stable situation on all single auctions due to the possibility of *allocation shifting*, it is not possible to derive a notion of equilibrium.

While with the proposed bidding strategy we could obtain an efficient allocation of resources in aggregate, the possibility of exchanging allocation shares between auctions can be further examined in understanding possible post-market situations. For example, if providers can express a preference for certain allocation patterns, such as a preference for holding the minimum number of active users in a system, a wholesale situation may arise, in which providers swap user allocations to improve their utility in another dimension.

## ▶ 4.4 Alternative Bidding Strategies in a Multi-Auction Market

The *BalancedBid* strategy described in the last section transfers the concept of truthful bidding to situations in which multiple PSP auctions are available to bidders. Every rational bidder would use *BalancedBid* as the utility-maximising strategy in the iterative game. However, depending on the specific objectives, alternative bidding strategies may be required to fulfill the specific needs of users. Many reasons can be given for not using *BalancedBid*. The most important ones can be summarised as follows:

- Bidders may be unable to coordinate their bids for the different auctions either because of missing computational power to compute the optimal result or because bidding on each auction is handled by independent processes.
- Auction results (in the form of opponent bid profiles) may be received unsynchronised. For example, large time gaps between updates may prohibit the coordination of bidding between auctions and motivate agents to submit bids without waiting for the result from other auctions.
- Bidders are unable to bundle resources from different auctions. Reasons for this may be found in either a technical inability of simultaneously using multiple interfaces or in the type of service that may not allow splitting up the demand for bandwidth into multiple sub-streams.

To understand the possibilities of bidders to employ alternative multi-auction bidding strategies we experimentally explore four different bidding models and analyse their properties if used by all bidders. In contrast to Section 4.3 we use simulation as the main methodology to understand the properties of the alternative bidding strategies. Instead of an analytical description of each bidding strategy we provide the bidding algorithm and present selected simulation results showing the system behaviour in a simple setting. The experimental setting used for all bidding strategies is identical to the example in Section 4.3.3.

Before we describe each bidding strategy in detail, we provide an overview of the main properties of all four bidding strategies and compare them with the *BalancedBid* strategy. Most of the results presented in this section has been published in Roggendorf and Beltran (2006b).

### ► 4.4.1 Short description and basic properties of the alternative bidding strategies

Table 4.5 provides a brief description of each alternative bidding strategy.

| Strategy Name | Short Description |
|---|---|
| *BalancedBid* | Identifies the utility-maximising share for the available auctions by creating a virtual market combining the opponent bid profiles from all auctions. Calculates a best truthful aggregated market bid and divides this bid into balanced sub-bids. |
| *BidAll* | Calculates and submits the truthful best-reply for each auction without coordinating between bids. |
| *UtilityBased* | Calculates the truthful best-reply for each auction and submits the bid only to the auction with the highest utility. |
| *OneActive* | Calculates the truthful best-reply for each auction and submits the bid to the auction with the highest utility. Ensures that no positive bid from previous rounds stays valid on other auctions by submitting zero-bids. |
| *ComplementaryUtility* | Calculates and submits a truthful best-reply to the auction with the highest utility and complements the bid with bids on other auctions by forming a truthful best-reply with the remaining demand. |

**Table 4.5:** Short description of the alternative bidding strategies.

In Table 4.6 we compare the main properties of the alternative bidding strategies with the *BalancedBid* strategy. We use three criteria for comparison; does the bidding strategy resemble truthful bidding? If it is used by all bidders, does the auction market reach a Nash equilibrium in finite time? Is the allocation in the Nash equilibrium efficient? All results have been derived experimentally. While in this way we cannot provide a hard proof of the properties, we can preclude them by identifying a counter-example. For example, if an experiment shows that the allocation in equilibrium is inefficient, we can conclude that the bidding strategy cannot lead to an efficient allocation of resources in general.

| Strategy Name | Truthful revelation of user valuation | Convergence | Efficiency in equilibrium |
|---|---|---|---|
| *BalancedBid* | Yes | Yes | Yes |
| *BidAll* | No | Yes | No |
| *UtilityBased* | No | Yes | No |
| *OneActive* | No | No | Yes |
| *ComplementaryUtility* | Yes | No | Yes |

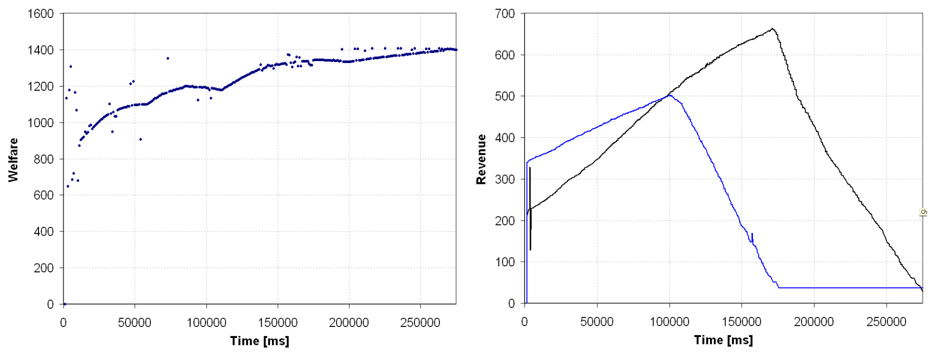
**Table 4.6:** Main properties of the bidding strategies.

### ► 4.4.2 The BidAll bidding strategy

With the *BidAll* bidding strategy players form a truthful best-reply separate by each available auction. This means that each bidder $i$ submits a bid $t_i^{(j)}$ according to Definition 4.3 to each auction $j$ whenever it increases his utility by at least $\epsilon/J$.

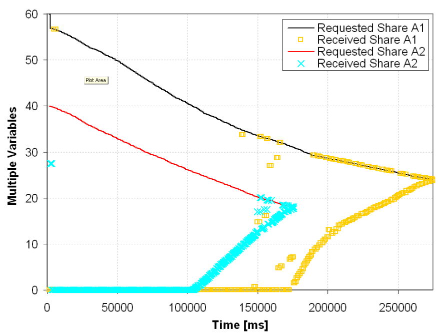
Since bids are not coordinated it is straightforward to conclude that this strategy does not truthfully reveal an agent's preference to the auction market. While each auctioneer receives a truthful reply, the sum of bids expressed by the bidder to the overall market does not reflect his true demand. Players with this bidding strategy usually bid above their valuation since they treat each auction as the only source of resources. If in equilibrium, a bidder holds positive shares from more than one auction he may hold units with zero marginal net utility if the sum of shares is above his maximum quantity $\bar{q}_i$. Additionally, agents may run into the so-called "winner's curse". A player would pay a negative rent for the summed shares and would be better off by not participating in an auction at all. When all agents use the *BidAll* strategy the system converges to Nash equilibrium, which is potentially inefficient, since agents may obtain resources with zero marginal valuation, while other agents may have an unfulfilled demand with positive marginal valuation.

Figure 4.23 provides the results of the simulation experiment with all five bidders using the *BidAll* bidding strategy. Figure 4.23(a) shows the social welfare over the iterative bidding period until Nash equilibrium is reached. Compared to the analytically derived solution the loss in social welfare in equilibrium is 4.61% . The shapes of the revenue graphs for both auctions are identical to the one-auctioneer case (Figure 4.23(b)). Figure 4.23(c) shows the requested and received resource shares for bidder 4, exemplarily. The resulting graphs are identical with the graphs generated in the one-auctioneer model (Figure 4.8(b)) since players use the identical strategy for each auction.




(a) Aggregated social welfare during the iterative bidding phase when all bidders implement the *BidAll* bidding strategy.

(b) Revenue per auction during the iterative bidding phase when all bidders implement the *BidAll* bidding strategy.




(c) Requested and received resource shares for player 4.

**Figure 4.23:** Social welfare, revenue per auction, and bidding behaviour of the *BidAll* bidding strategy.

The bidding algorithm of the *BidAll* strategy is given in Algorithm 4.4.2.

**Algorithm 2** *BidAll* Bidding Strategy

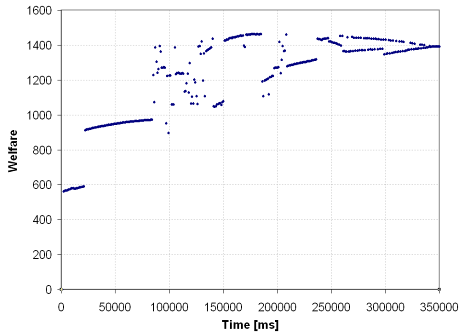
```
Let s_i^(j) = 0 and ŝ_{-i}^(j) = ∅ for each auction j.
loop
   Update list of available auctions and receive updates of ŝ_{-i}^(j) from each auction j.
   for all Available auctions j do
      Compute a truthful reply t_i^(j) for the opponent bid profile s_{-i}^(j) according to Definition 4.6.
      if u_i(t_i^(j); ŝ_i^(j)) > u_i(s_i^(j); ŝ_i^(j)) + ε/J then
         Update bid t_i^(j) on auction j
      end if
   end for
   sleep for 1 second
end loop
```

### ► 4.4.3 The UtilityBased bidding strategy

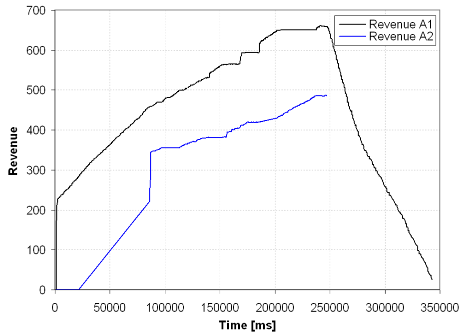
The *UtilityBased* bidding strategy coordinates bidding on several auctions by comparing the utility of the truthful best-reply on each auction and only updating the auction with the highest utility from the updated bid. If a bidder cannot find a new truthful best-reply he uses the allocation $a_i^{(j)}$ obtained from the previous auction round for comparison with the other auctions. If the utility from $a_i^{(j)}$ is higher than from any other auction no bid is submitted. In the remaining auctions bids from previous periods stay valid.

With this bidding strategy a bidder reduces his risk of overbidding since he only sends one truthful best-reply in each period. However, in equilibrium, bidders may potentially end up with resources allocated from more than one auction as bids from previous periods may still be winning bids. As with the *BidAll* strategy bidding is not truthful since the overall demand expressed to the auction market does not correspond to the demand of the bidder. Therefore, the allocation in equilibrium may lead to inefficient outcomes since bidders may hold resources from multiple auctions and resources may be obtained that gain zero marginal value.

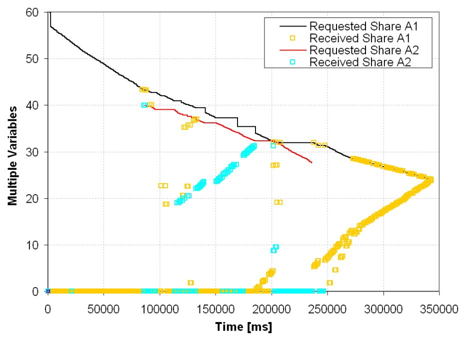
Figures 4.24(a)-4.24(c) provide the simulation results for the *UtilityBased* bidding strategy. The efficiency loss compared to the analytically derived solution equals 5.31%. When comparing the revenue in equilibrium an important difference to the *BidAll* strategy can be observed. Since players stop updating bids at some point during the convergence at auction 2 and continue bidding at 1, revenue for auction 2 does not converge to zero. Instead, auctioneer 2, which offers less resources, can generate significant revenue in equilibrium. Figure 4.24(c) shows the bidding behaviour for player 4. Compared to the *BidAll* strategy the iterative bidding process takes longer and players delay bidding in auction 2 since at start auction 1 is more attractive because it offers more resources. With ongoing bidding, players may find higher utility in auction 2 and start submitting bids to this auction.




(a) Aggregated social welfare during the iterative bidding phase when all bidders implement the *UtilityBased* bidding strategy.




(b) Revenue per auction during the iterative bidding phase when all bidders implement the *UtilityBased* bidding strategy.




(c) Requested shares and received resource shares for player 4.

**Figure 4.24:** Social welfare, revenue per auction, and bidding behaviour of the *UtilityBased* bidding strategy.

The bidding algorithm for *UtilityBased* is shown in Algorithm 3.

### ▶ 4.4.4 The OneActive bidding strategy

*OneActive* explores the possibility of obtaining resources from only one auction while at the same time making use of the available auction market consisting of multiple auctions. It uses the same bidding logic as implemented with the *UtilityBased* strategy but additionally ensures that a bidders does not hold resources from more than one auction at the same time. In each round only one truthful best-reply is submitted to the auction $l$ with the highest utility. Additionally, the bidder determines if he holds resources $a_i^{(j)}$ from any auction $j \neq l$. In this case a zero bid $t_i^{(j)} = (0, 0)$ is sent to this auction.

With the *OneActive* strategy, a player controls his overall demand expressed in his bids and ensures that in equilibrium, does not hold an allocation above his maximum demand $\bar{q}$. However, bids from previous rounds may stay active on other auctions $j \neq l$ as long as the bidder does not receive a positive allocation from those bids (meaning that the bid was outbid by another bidder).

Another question is if the algorithm converges to equilibrium in finite time. Since

**Algorithm 3** *UtilityBased* Bidding Strategy

Let $s_i^{(j)} = 0$ and $\hat{s}_{-i}^{(j)} = \emptyset$ for each auction $j$.<br>
**loop**<br>
  Update list of available auctions and receive updates of $\hat{s}_{-i}^{(j)}$ from each auction $j$.<br>
  **for all** Available auctions $k$ **do**<br>
    Compute a truthful reply $t_i^{(j)}$ for the opponent bid profile $s_{-i}^{(j)}$ according to Definition 4.6.<br>
    **if** $u_i(t_i^{(j)}; \hat{s}_i^{(j)}) > u_i(s_i^{(j)}; \hat{s}_i^{(j)}) + \epsilon/J$ **then**<br>
      Set $x_i^{(j)} = t_i^{(j)}$.<br>
    **else**<br>
      Set $x_i^{(j)} = s_i^{(j)}$.<br>
    **end if**<br>
  **end for**<br>
  Find auction $l$ with $\max u_i(x_i^{(l)}; \hat{s}_i^l)$.<br>
  **if** $x_i^{(l)}$ is a truthful best-reply **then**<br>
    Send a new bid $t_i^{(l)}$ to auction $l$.<br>
  **else**<br>
    Do nothing.<br>
  **end if**<br>
  sleep for 1 second<br>
**end loop**

players cancel their allocation if they receive resources from more than one auction, the convergence process is regularly ”disturbed”. After a zero bid, the allocation of resources may completely change and, in turn, other bidders may switch their demand to this auction. This effect can be observed in simulation experiments. Therefore, the algorithm usually does not converge to a stable operating point after a finite iterative bidding period. For settings with relatively few bidders, however, equilibrium may be reached randomly. Revenue, at this operating point, may be higher than with other bidding strategies as bidders do not remove previous bids from other auctions and therefore create some kind of artificial congestion.

Figure 4.25 provides the simulation results for the *OneActive* bidding strategy. Figure 4.24(a) shows the social welfare over the iterative bidding period, which, in equilibrium, is very close to the analytically derived optimal allocation. However, the revenue graphs of both auctions (Figure 4.24(b)) do not show any asymptotic characteristics but appear to be randomly jumping. In equilibrium, auction 1 generates significant revenue, while for auction 2 the revenue is zero. Figure 4.24(c) shows the bidding behaviour of player 4. It can be observed that the player switches auctions during the convergence process with the vertical lines to zero indicating the submission of zero bids.

The analysis of the simulation results reveals another important distinction between the *OneActive* and previously described strategies. While the order of bid updates has no major influence on the equilibrium allocation as with the other bidding strategies, this factor influences the resulting equilibrium allocation when all bidders implement *One-*

*Active*. Since bid changes with *OneActive* can be drastic within one bidding step, reacting bids of other players may be dramatic as well. Therefore, the timing and distribution of bidding information does have an influence for the final allocation in equilibrium.

**Remark** A variant of *OneActive* is to cancel bids by submitting a zero bid to the active auction as soon as the bidder decides to change to a new auction. In this way, an agent ensures at all times that its expressed demand does not exceed its valuation. Such a bidding strategy would resemble truthful revelation to the multi-auctioneer case. However, such a behaviour leads to a very unstable auction market as the cancellation of resources after each step means that the convergence process is disturbed with each bid. With this variant no asymptotic increase in welfare can be observed over the simulation period. We therefore omit a detailed discussion of this bidding variant.

**Remark** A possible way of limiting the volatility of the bidding behaviour is the introduction of a "switching fee" $\phi$, similar to the bid fee $\epsilon$. Whenever an agent decides to switch from one to another auction the switching fee is charged by the auctioneer. The switching fee prevents auction switching in cases of small utility gains because the gain must be larger than the switching fee. During the comparison process $\phi$ has to be taken into consideration for all non-leading auctions in the following way:

$$\text{Find auction } j \text{ with } \max u_i(x_i^{(j)}; \hat{s}_i^{(j)}) - \phi x$$

with $x = 1 \ \forall j \neq l$.

### ▶ 4.4.5 The ComplementaryUtility bidding strategy

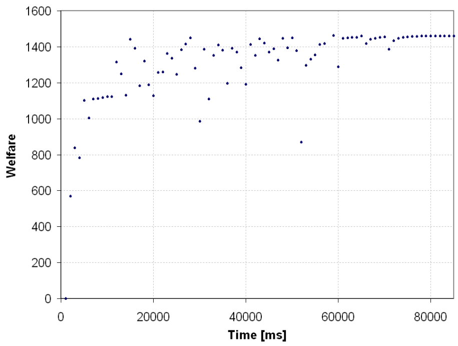
*ComplementaryUtility* implements the idea of "dividing up" the demand for resources between different auctions. As with *OneActive*, a player generates a list of truthful best-replies $t_i^{(j)}$ for each opponent bid profile $s_{-i}^{(j)}$ and determines the auction with the highest utility. If the highest utility is generated by a new bid, the bid is sent to the respective auction. Otherwise, the player proceeds to the next step directly.

In the second step a bidder reduces his demand by the quantity $v_i^{(j)}$ of the first bid. With the remaining demand a truthful best-reply is again generated for all active auctions except for the auction where a bid has already been sent. Again, the auction with the highest utility is selected and the bid is updated. This procedure is repeated until either there is no additional auction available or the demand has been fully allocated among auctions.

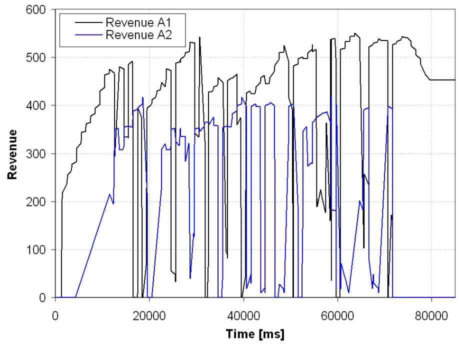
Figure 4.26 depicts how the demand function is reduced in each subsequent bidding step (for a linearly decreasing demand function). If bidder $i$ bids for $v_i^{(j)}$ on auction $j$ he subtracts this amount from his overall demand $\bar{q}_i$ and searches for the auction with the highest utility with the reduced demand function. This algorithm ensures that a

**Algorithm 4** *OneActive* Bidding Strategy

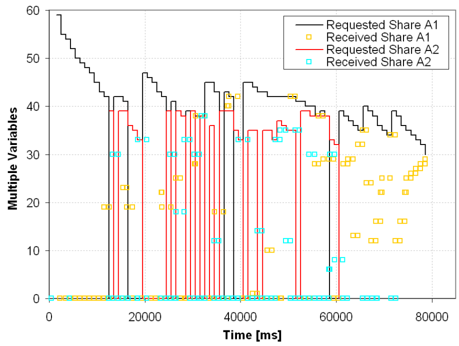
Let $s_i^{(j)} = 0$ and $\hat{s}_{-i}^{(j)} = \emptyset$ for each auctions $j$.
**loop**
  Update list of available auctions and receive updates of $\hat{s}_{-i}^{(j)}$ from each auction $j$.
  **for all** Available auctions $j$ **do**
    Compute a truthful reply $t_i^{(j)}$ for the opponent bid profile $s_{-i}^{(j)}$ according to Definition 4.6.
    **if** $u_i(t_i^{(j)}; \hat{s}_i^{(j)}) > u_i(s_i^{(j)}; \hat{s}_i^{(j)}) + \epsilon/J$ **then**
      Set $x_i^{(j)} = t_i^{(j)}$.
    **else**
      Set $x_i^{(j)} = s_i^{(j)}$.
    **end if**
  **end for**
  Find auction $l$ with max $u_i(x_i^{(l)}; \hat{s}_i^l)$.
  **if** $x_i^{(l)}$ is a truthful best-reply **then**
    Send a new bid $t_i^{(l)}$ to auction $l$.
  **else**
    Do nothing.
  **end if**
  **for all** active auctions $j \neq l$ **do**
    **if** $a_i^{(j)} > 0$ **then**
      Send a zero bid $t_i^{(j)} = (0, 0)$ to auction $j$.
    **end if**
  **end for**
  sleep for 1 second
**end loop**




(a) Aggregated social welfare during the iterative bidding phase when all bidders implement the *OneActive* bidding strategy.




(b) Revenue per auction during the iterative bidding phase when all bidders implement the *OneActive* bidding strategy.




(c) Requested and received resource shares for player 4.

**Figure 4.25:** Social welfare, revenue per auction, and bidding behaviour of the *OneActive* bidding strategy.

player does not receive resources above his valuation in equilibrium. However, as with *OneActive*, previous bids not resulting in a positive allocation stay active.

By using the *ComplementaryUtility* bidding strategy an agent ”price-discriminates” itself since it is willing to pay a higher marginal price for the first units than for subsequent units. Figure 4.27 shows that an agent gives up some if its surplus for receiving resources from multiple sources. The overall cost $\sum_j c_i^{(j)}$ can be higher than in the one-auctioneer case.

In contrast to the *OneActive* strategy the *ComplementaryUtility* strategy does not make extensive use of zero bids since usually non-zero bids are submitted to all auctions. Only in situations in which the competitive situation changes rapidly (e.g., in cases of other players joining or leaving) a player may cancel his allocation in an auction. However, a player may change the order of bids, which, in turn, may lead to large changes in each affected auction. For this reason the algorithm does not converge to an equilibrium in finite time; instead, a stable operating point is only reached randomly. Since, in equilibrium, all bidders express their truthful demand to the auction market, the resulting allocation is efficient.

Figure 4.28(a) shows the social welfare generated over the bidding period for the

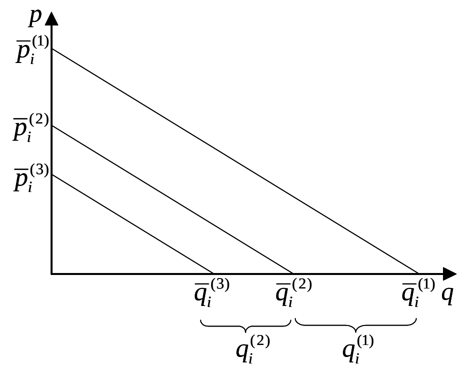


**Figure 4.26:** Demand reduction method used by the *ComplementaryUtility* bidding strategy.

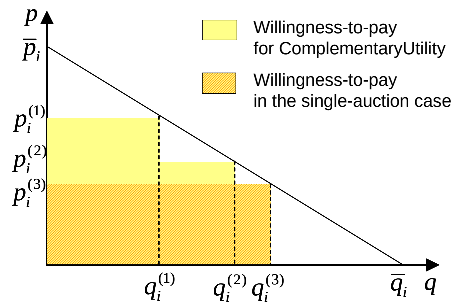



**Figure 4.27:** User price discrimination when submitting multiple bids to different auctions.

*ComplementaryUtility* strategy. The graph shows several local maxima before the stable operating point is found. A repeating pattern can be identified in Figure 4.28(b), which shows the revenue obtained by both auctioneers. Revenue steadily increases until a certain point when the convergence process is disturbed by bidders changing the order of their bids. Nevertheless, no direct comparison with the "typical" PSP revenue graph can be made. In Figure 4.28(c) the bidding behaviour of bidder 4 is shown. The "switching" between the leading auctions can be observed when requested shares suddenly jump to a lower value.

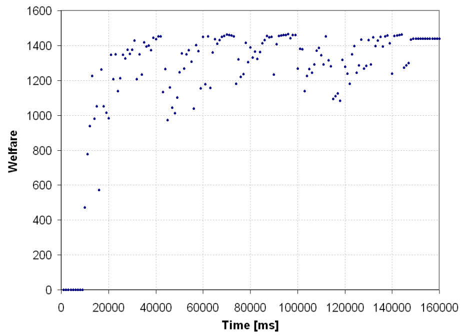
Algorithm 5 presents the formal algorithm to implement the *ComplementaryUtility* bidding strategy.

### ▸ 4.4.6 Comparison of bidding strategies in a stochastic environment

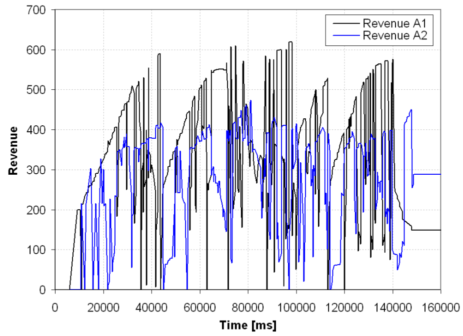
The last four paragraphs have described the alternative bidding strategies. Simulation experiments have complemented the understanding of how the algorithms behave in a simple setting and have provided an intuition about the outcome in equilibrium. In this section we implement the bidding strategies in a stochastic setup to understand the difference in bidding behaviour under different loads.

**Algorithm 5** *ComplementaryUtility* Bidding Strategy

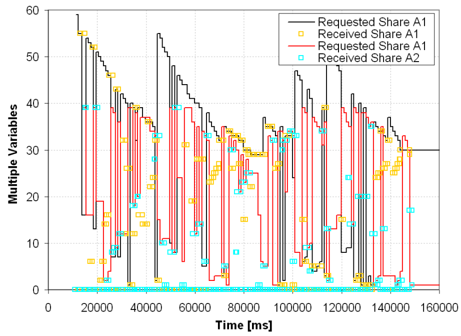
Let $s_i^{(j)} = 0$ and $\hat{s}_{-i}^{(j)} = \emptyset$ for each auctions $j$.
**loop**
  Update list of available auctions and receive updates of $\hat{s}_{-i}^{(j)}$ from each auction $j$.
  **while** $\bar{q}_i > 0$ **do**
    **for all** Available auctions $j$ **do**
      Compute a truthful reply $t_i^{(j)}$ for the opponent bid profile $s_{-i}^{(j)}$ according to Definition 4.6.
      **if** $u_i(t_i^{(j)}; \hat{s}_i^{(j)}) > u_i(s_i^{(j)}; \hat{s}_i^{(j)}) + \epsilon/J$ **then**
        Set $x_i^{(j)} = t_i^{(j)}$.
      **else**
        Set $x_i^{(j)} = s_i^{(j)}$.
      **end if**
    **end for**
    Find auction $l$ with $\max u_i(x_i^{(l)}; \hat{s}_i^l)$ and $l \notin k$.
    Add $l$ to vector $k$
    **if** $x_i^{(l)}$ is a truthful best-reply **then**
      Send a new bid $t_i^{(l)}$ to auction $l$.
      Set $\bar{q}_i = \bar{q}_i - v_i^{(l)}$ and $\bar{p}_i = \theta_i'(v_i^{(l)})$.
    **else**
      Set $\bar{q}_i = \bar{q}_i - a_i^{(l)}$ and $\bar{p}_i = \theta_i'(c_i^{(l)}/a_i^{(l)})$.
    **end if**
    Implement new demand function.
  **end while**
  **for all** Available auctions $j \notin k$ **do**
    **if** $a_i^{(j)} > 0$ **then**
      Send a zero bid $t_i^{(j)} = (0, 0)$ to auction $j$.
    **end if**
  **end for**
  sleep for 1 second
**end loop**




(a) Aggregated social welfare during the iterative bidding phase when all bidders implement the *ComplementaryUtility* bidding strategy.




(b) Revenue generated by the *ComplementaryUtility* bidding strategy for auctions 1 and 2




(c) Requested and received resource shares for player 4.

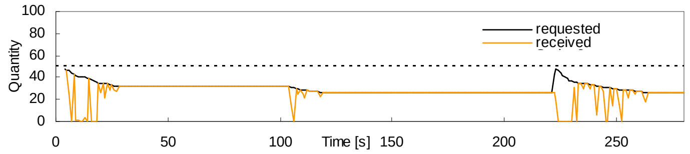
**Figure 4.28:** Social welfare, revenue per auction, and bidding behaviour of the *ComplementaryUtility* bidding strategy.

- Two network providers are running four access points (AP) each to cover an area of 500 by 500 units. Access points are represented by provider agents offering network resources. The entire area is covered by both providers. Each access points offers a capacity of $Q = 300$.

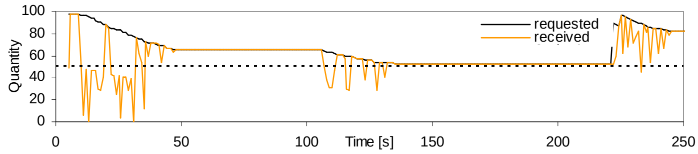
- 100 user agents are randomly distributed over the service area. All users have a constant maximum demand of $\bar{q} = 50$ and a maximum marginal unit price $\bar{p}$ generated from a uniform distribution on the interval $[10, 20]$. All agents have access to only one provider, which has been randomly selected. If a user can access more than one AP it selects the AP closest to him.

- 70 agents initially request service. 30 agents join the market-place at $t = 100sec$. 50 randomly selected agents leave at $t = 220sec$.

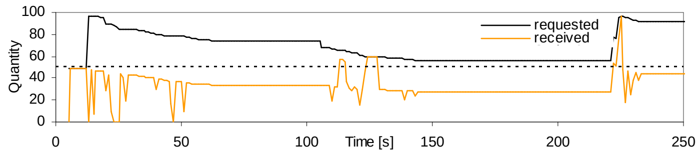
- One agent with $\bar{q} = 50$ and $\bar{p} = 15$, which has access to both providers, is located at position $(200, 200)$. In four different experiments he uses the *BalancedBid*, *BidAll*, *UtilityBased*, and *ComplementaryUtility* strategy, respectively.



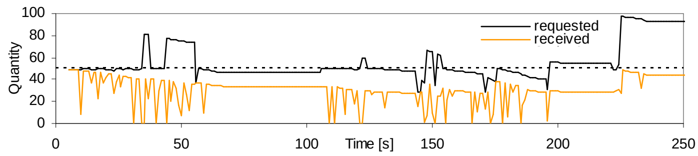

(a) BalancedBid




(b) BidAll




(c) UtilityBased




(d) ComplementaryUtility

**Figure 4.29:** Summed requested and received resource shares for a bidder with access to two auctions under the four bidding strategies.

In all experiments we record the requested and received resource shares for this bidder. Figure 4.29 shows the cumulated requested and received shares. With the *BalancedBid* strategy the requested share is always below the maximum demand $\bar{q}$ and the received share matches the requested share when equilibrium has been reached. We can also observe that the expressed overall demand is lowest at the time of high demand (from $t = 100sec$ to $t = 220sec$).

With the *BidAll* strategy the agent is able to acquire the highest amount of resources. However, since bids to the different auctions are not coordinated, he also receives more than his actual demand for some time periods. While the total price may not be above his willingness-to-pay he captures units of the resource which bring no additional value to him.

By using the *UtilityBased* strategy we observe that overall requested quantity is comparable to the *BidAll* strategy. However, since the player is coordinating his bids, the received quantity is not larger than his maximum demand over longer periods. This is because a bidder may still have a valid bid in an auction but is not updating it any more because resources on other auction places have become more attractive.

With the *ComplementaryUtility* strategy the player bids much more cautiously. The received quantity always stays below the maximum demand. Compared to the other three strategies the identification of equilibrium is erratic and the process to get to a stable allocation takes much longer. This is especially true for the second time period, when a total of 100 players are present in the market.

## ► 4.5 Simulation Results: Efficiency, Revenue and Convergence

We use agent-based simulation to explore the properties of the bidding algorithms developed in the Sections 4.3 and 4.4 in a market consisting of multiple auctions. We design experiments, which reveal properties of the bidding strategies that are potentially beyond the possibilities of an analytical explanation. We can distinguish five fields of interest we cover in this section:

- **Welfare gain from accessing multiple networks:** multiple-access allows agents to fulfill their demand from multiple sources depending on the demand at the different auctions. It is expected that in such a setting the welfare will be higher than in a setting where all users have access to only one network.
- **Revenue generation as additional goal to congestion control and welfare maximisation:** While the PSP mechanism has been developed with the main objectives of controlling congestion and reaching efficient allocations in equilibrium, several modifications allow an increase of revenues obtained by the auctioneer. We are interested if such concepts can be transferred to the setting of multiple auctions.
- **Convergence behaviour:** With the introduction of the multiple-auction market and the proposed *BalancedBid* strategy additional questions about the convergence behaviour arise. What is the influence of multiple-access on auction rounds and convergence time? How does the interconnection between auctions (with bidders having access to different auctions) influences the convergence behaviour of the system? How does the system react to events after convergence?

Our focus is predominantly on the *BalancedBid* bidding strategy since it has been shown that it implements truthful bidding in a multiple-auction setting and is the best reply for a bidder with multiple access independent of the strategies used by the other bidders. However, we will also provide an extended view on the behaviour of some of the alternative bidding strategies described in the previous section.

### ► 4.5.1 Statistical analysis of the simulation results

An important step in the analysis of the simulation results is the selection of the statistical tools to analyse the resulting output variables. In most experiments we are interested in the effects of one input variable on the response or dependent variable, which is the variable of interest to be measured in the experiment. Since we use random input variables, such as the demand profile of each bidder, we need to run multiple replications of the same experiment to derive statistically meaningful results.

In all experiments we use the *Independent Replications (IR)* method to obtain a point estimator for the mean and sample variance of the relevant output variables (Alexopoulos and Goldsman, 2004). Since we can assume that the output variables of the different replications are *independent and identically distributed* (IID), we can use the output variables directly to form such estimators.

Since we are mostly interested in the equilibrium outcomes of a terminating-type simulation, we only record the allocation once the bidding process has been stopped. In experiments we want to analyse convergence time we additionally record the bidding rounds and the time to convergence in seconds.

To obtain statistically relevant results we used a procedure, in which we conducted enough replications so that the half-lengths of the 90% confidence interval divided by the estimator of the output variable of interest is less than 5%. When making comparisons between different scenarios we used the relative output variables for determining the number of replications to reach the target confidence level. On average, at least 20 replication each data point were required to fulfill this goal but in some experiments, more replications were required.

### ▶ 4.5.2 Welfare gain from multiple PSP auctions

We are interested in the social welfare when players have access to more than one network compared to the case when players are restricted to obtain resources from only one provider. Two main questions lead the simulation design. First, we experimentally explore if the *BalancedBid* produces an efficient allocation in equilibrium. Second, we are interested in the gain in social welfare stemming from the multiple-auction case. We measure social welfare as the sum of utility from all players.

#### Efficiency comparison of the BalancedBid strategy

In Section 4.3 we have already proved that in a multiple-auction setting, in which all bidders implement the *BalancedBid* strategy as their truthful best-reply, the resulting equilibrium allocation maximises social welfare. We verify this result by experimentation with a different number of auctions available to the bidders. Since we know that the PSP auction produces an efficient allocation we can use the one-auction case as reference. In additional scenarios we vary the number of auctions from 2 to 5. We then compare social welfare in equilibrium of the first scenario with the other scenarios to analyse the effect of multiple-auctions on the efficiency of resource allocation using the *BalancedBid* strategy.

In the simulation setup we create $N = 10$ players with a constant maximum demand of $\bar{q} = 50$. The valuation for all players is randomised in the interval $[10, 20]$ using a uniform distribution. We set the overall capacity in the system to $Q = 100$. In the first scenario all resources are assigned to one auction. In the other scenarios we distribute resources evenly between $M$ auctions; $\epsilon$ has been set to 0.01 in all experiments.

1,000 replications of the experiment were conducted. Each run consists of six separate runs with identical player profiles. Figure 4.30 shows the distribution of the relative change in welfare for all multiple-auction cases (Scenarios 2-5) compared to the one-auction case (Scenario 1). About 50% of all cases produce an outcome with less than 0.01% difference to the one-auctioneer setting. None of the results shows a larger difference than 0.10%. We can compare this result with the maximum boundary given by $4Q\sqrt{\epsilon\kappa}$. With the given values we derive a maximum loss in social welfare of 17.89, with $\epsilon$ being set to 0.01. With an average absolute value of social welfare of 1,385.65 from all experiments the maximum loss can be up to 1.3%. We conclude that for the particular setup, the use of the *BalancedBid* strategy by all bidders results in an efficient allocation in equilibrium. The result is limited to the case of the specific valuation function used for the experiments and assumes that all players have access to all auctions.

We can also observe a symmetrical accumulation of cases at ±0.09%, which may be connected with the particular parametrisation of the simulation and the precision of the bids of 0.1 being submitted to the auction.

We have also analysed the distributions of the different multiple-auction scenarios, but could not find significant differences to the distribution obtained from averaging over all multi-auction scenarios.

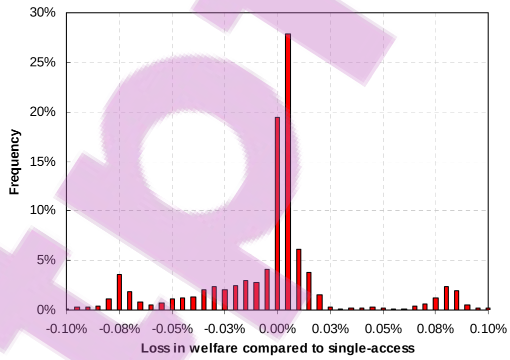



**Figure 4.30:** Distribution of the change in welfare with multiple-access compared to the one-auctioneer scenario.

A second analysis, which could be conducted with the available material is to compare the convergence time of the one-auction case with the multiple-auction cases. In Figure 4.31 we compare the convergence time of the first scenario with all multiple-auction scenarios. The maximum of the distribution is found at about 0%, with an average of −26%. In conclusion we can state that on average the convergence time is reduced if resources are distributed over multiple auctions and players implement the *BalancedBid* strategy. However, this result is limited to the specific parametrisation and setup of the experiment. No direct pattern could be identified that could be responsible for faster convergence in some cases. Additionally, no significantly different patterns could be identified which distinguish between the different scenarios relating to the number of auctions in the market.

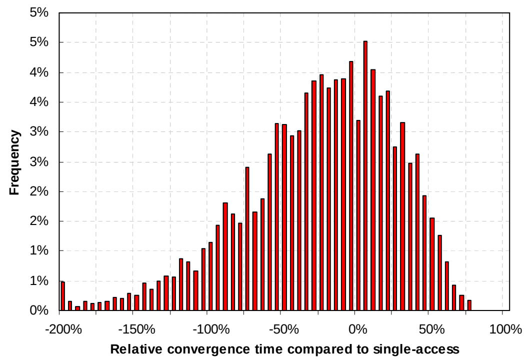



**Figure 4.31:** Distribution of relative convergence time compared with the convergence time in the one-auctioneer case.

In additional experiments we have tested whether the distribution of resources between auctioneers influences the obtained results.[15] With all tested combinations the relative efficiency loss compared to the one-auctioneer setting remains well in the boundary of $4Q\sqrt{\epsilon\kappa}$.

## Welfare gain from multiple-access using BalancedBid

Next, we are interested in the efficiency improvements when players have access to multiple auctions compared to a situation in which players are limited to one auction. Intuitively, the gain should be higher when there is a greater imbalance of supply or demand in the market. This may be the case where the congestion level significantly differs between auctions or where the general resource valuation of one user group is higher compared to another group of users in another auction. In such cases, by letting users have access to multiple auctions, the overall efficiency of the allocation of resources can be improved.

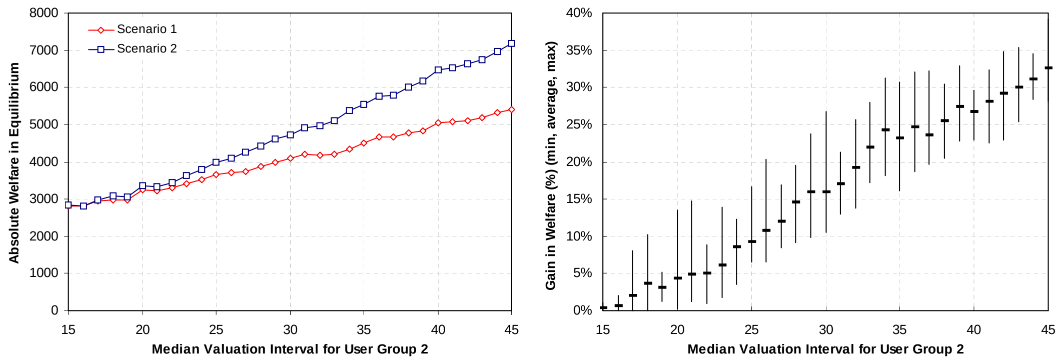
To explore the correlation between different setups and the efficiency improvement we conduct a series of experiments. All experiments make use of two scenarios. In both scenarios two user groups are allocated to two networks. In the first scenario users have access to only one auction. In the second scenario we let users have access to multiple auctions. We analyse the gain in welfare resulting from the multiple-auction option.

**Increasing the user valuation for one user group** Each user group has 5 users. The user profiles of the first group are generated from the interval $[10, 20]$ using a uniform distribution. For group 2 we start with a valuation interval $[10, 20]$ and gradually move

[15] in the previous case we had evenly distributed the resources over all auctions

the interval to [40, 50]. For each data point we initially conduct 20 replications for the two scenarios and record the obtained social welfare in equilibrium.

In Figure 4.32(a) the absolute gain in welfare is shown when letting users have access to both networks. As expected, the welfare gain becomes larger with the increasing average valuation of the one user group since in the second scenario, such users have access to the other auction. Because of their higher valuation of resources they are able to outbid the users from the first user group. Figure 4.32(b) shows the relative welfare gain together with the minimum and maximum values from the 20 replications for each data point. While the interval size between the point estimator for the mean and the minimum and maximum values varies, all data points fulfill the required confidence level as defined in Section 4.5.1.




(a) Point estimator of the mean welfare of scenario 1 and 2 measured at equilibrium from 20 replications.

(b) Relative gain in welfare from multiple-access, showing the minimum, mean, and maximum gain from 20 replications for each data point.

**Figure 4.32:** Experiment 1: Increasing the valuation level for one user group.

The experiment also allows us to analyse the convergence time between the two scenarios. Figure 4.33 shows a comparison of convergence time-to-equilibrium for both scenarios. As additional experiments also do not show any correlation between the definition of the valuation interval and convergence time we can conclude that there is no direct connection between the valuation interval and the convergence time.

To understand how the level of congestion influences the efficiency improvement we conducted three additional experiments and gradually increased the available capacity from $Q^{(1)} = Q^{(2)} = 100$ to $Q^{(1)} = Q^{(2)} = 400$ in steps of 100. Figure 4.34 shows that with decreasing congestion the gain in welfare decreases. Since more of the demand can already be fulfilled with the resources from one network, and the marginal valuation decreases with the quantity of resources obtained, the gain from giving access to another auction is reduced.

**Increasing the number of users in one user group** Experiment 2 aims at understanding the influence of an increased number of bidders in one of the networks on the gain in welfare when access to both auctions is enabled. While keeping the number of players in

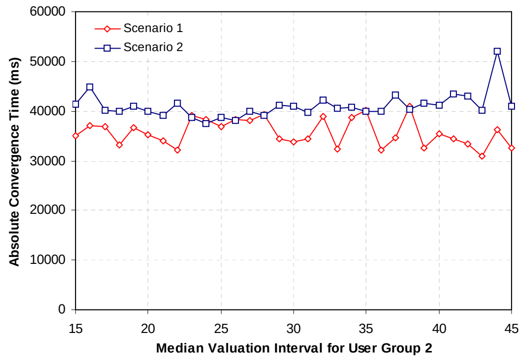



**Figure 4.33:** Comparison of mean convergence time-to-equilibrium between the two scenarios (20 replications per data point) with $Q^{(1)} = Q^{(2)} = 100$.

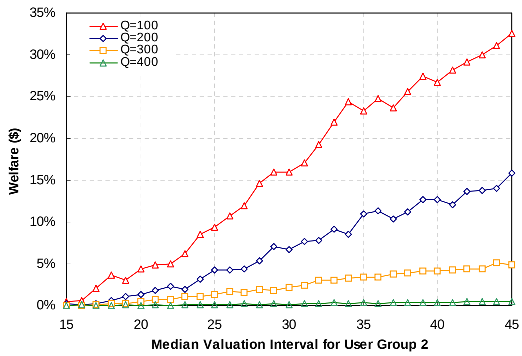
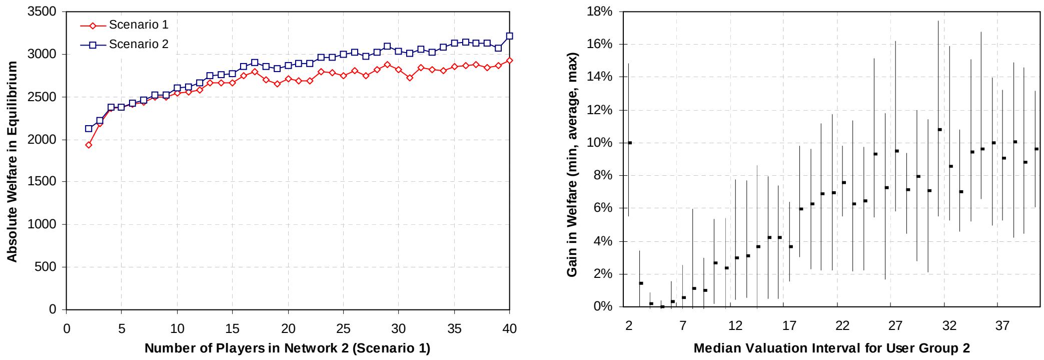



**Figure 4.34:** Comparison of welfare gain with $Q^{(2)} = 100$, $Q^{(2)} = 200$, $Q^{(2)} = 300$, and $Q^{(2)} = 400$.

group 1 fixed to 5, we gradually increase the number of players in auction 2 from 2 to 40 and observe the gain in welfare. The profiles of both user groups are generated with the interval $[10, 20]$. We set $\bar{q} = 50$ for all agents. For data point we conducted 20 replications.

The results of the experiments are shown in Figure 4.35. Figure 4.35(a) shows the absolute welfare in equilibrium from both scenarios. Figure 4.35(b) depicts the relative welfare gain from having players access to both auctions. We can observe that the gain in welfare drops sharply when introducing a third player in group 2. This is because auction 2 becomes more congested and the possibility to gain additional resources from auction 2 is reduced. The gain then slowly increases with an increasing number of players in group 2. However, the gain is limited to about 10% and cannot be further increased by introducing additional bidders. The reason for this can be found in the shape of the valuation function. Since each player gets only a small share of the resource, the valuation function in this area is nearly linear. With the introduction of new users, no additional utility can be generated compared to the situation in which the resource is shared between the already existing users.

We have also analysed the convergence time-to-equilibrium (Figure 4.36). While the




(a) Point estimator of the mean welfare of scenario 1 and 2 measured at equilibrium from 20 replications.

(b) Relative gain in welfare, showing the minimum, mean, and maximum gain from 20 replications for each data point.

**Figure 4.35:** Experiment 2: Increasing the number of users in one group.

time to convergence for scenario 2 is nearly constant for up to 20 users in network 2, it sharply increases when more users are introduced. For scenario 1 (users are restricted to one auction), the convergence time stays relatively constant for all numbers of users. It can be concluded that the complexity of the resource allocation with multiple auctions increases significantly as more players participate in the auctions.

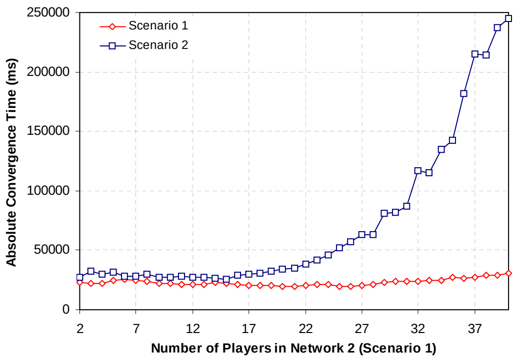
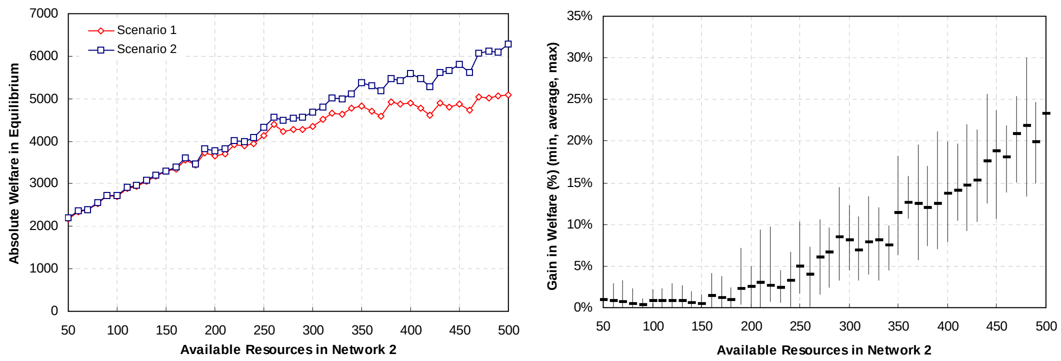



**Figure 4.36:** Comparison of the average convergence time-to-equilibrium between the two scenarios (20 replications per data point) for an increasing number of users in group 2.

**Increasing network resources in one network** The goal of Experiment 3 was to explore the consequences of varying the available capacity in only one of the networks. We gradually increased the capacity of network 2 from $Q^{(1)} = 100$ to $Q^{(2)} = 500$ and observed the gain in welfare between the two simulation sets. All other parameters remain identical to the last experiments ($\bar{q} = 100$, 5 players in each group). For both user groups we used the valuation interval $[10, 20]$ for generating the user profiles. Figure 4.37(a) shows that with increasing capacity in network 2 the welfare gain becomes smaller; this is because players in network 2 can already fulfill most of their demand from network 2 and accessing network 1 does not add additional value to them. Figure 4.37(b) gives the relative gain in welfare with the mean, minimum and maximum values from the replications conducted for each data point.




(a) Point estimator of the mean welfare of scenario 1 and 2 measured at equilibrium from 20 replications.

(b) Relative gain in welfare from multiple-access, showing the minimum, mean, and maximum gain from 20 replications per data point.

**Figure 4.37:** Experiment 3

We were now interested in the results when additionally varying the valuation interval for bidders in group 2. We therefore repeated the same experiment for the valuation intervals [20, 30], [30, 40], [40, 50], and [50, 60]. Figure 4.38 shows the results of relative welfare gain for the 5 experiments. As can be observed the gain in welfare first decreases with increasing capacity in network 2. This effect can be explained with the fact that the advantage of accessing a second network decreases as more of the demand can already be fulfilled by network 1. We can also observe that the welfare gain starts to again increase with higher capacity in network 2. The starting point for the increase depends on the chosen valuation interval for user group 2. This is because bidders in group 1 gradually start to take advantage of the resources in network 2 as capacity grows. Even if their valuation is generally smaller than that of group 2, the decreasing congestion in network 2 does allow them to compete for resource shares, which contribute more to the overall welfare than increasing the resource shares of high-value bidders.

Another analysis that reveals important properties is to look at the convergence times for all 5 experiments. Figure 4.39 indicates that there is a correlation between network capacity (or more precisely, the congestion level in the networks) and the time the market needs to reach equilibrium when all users have access to all networks compared to the situation when users are limited to one network. We see that for a certain range of network capacities the convergence times become close to the case of single access or are even lower.

To understand the absolute change in convergence time we show the absolute convergence times for the valuation interval [10, 20] of user group 2 in Figure 4.40. While the convergence times for single access decrease with increasing capacity, the time to convergence with multiple-access first increases and then stays relatively constant with

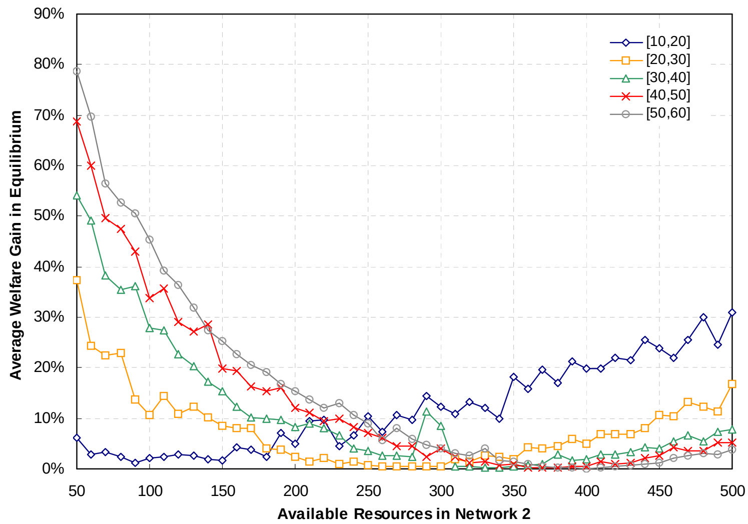



**Figure 4.38:** Point estimator of the mean gain in welfare from multiple-access for different valuation intervals for increasing capacities in network 2 and user group 2.

increasing capacity.

**Comparison of welfare generated by alternative bidding strategies** Another question for analysis is how the other bidding strategies, which have been presented in detail in Section 4.4, perform compared to the *BalancedBid* bidding strategy. In contrast to *BalancedBid*, the alternative strategies have been shown to either be efficient (maximise the social welfare in equilibrium) but not converge, or converge to equilibrium but produce inefficient allocations. Part of the results have been presented in Beltran and Roggendorf (2005).

In the following experiment we compare the outcome of *BalancedBid* and two alternative bidding strategies with the welfare generated by the one-auctioneer case. As explained in Section 4.5.2 we use the one auctioneer case to obtain a solution close to the optimal solution for comparison.

We define the following experiment: In scenario 1, 5 bidders have a constant maximum demand of $\bar{q} = 100$. The maximum unit price is randomly generated from a uniform distribution on $[10, 20]$. In this scenario players have access to only one network with capacity $Q = 200$. All players implement a truthful bidding strategy. In the remaining scenarios, the capacity is divided over two separate networks with initially $Q^{(1)} = Q^{(2)} = 100$. In scenario 2 we implement the *BalancedBid* strategy. In scenario 3, the *BidAll* strategy is used and in scenario 4, we implement the *UtilityBased* bidding strategy.

The other two bidding strategies, namely the *OneActive* and the *ComplementingUtility*, cannot be tested in this context since they do not converge and equilibrium is usually not reached in finite time.

We compare the loss in welfare when using alternative bidding strategies with the

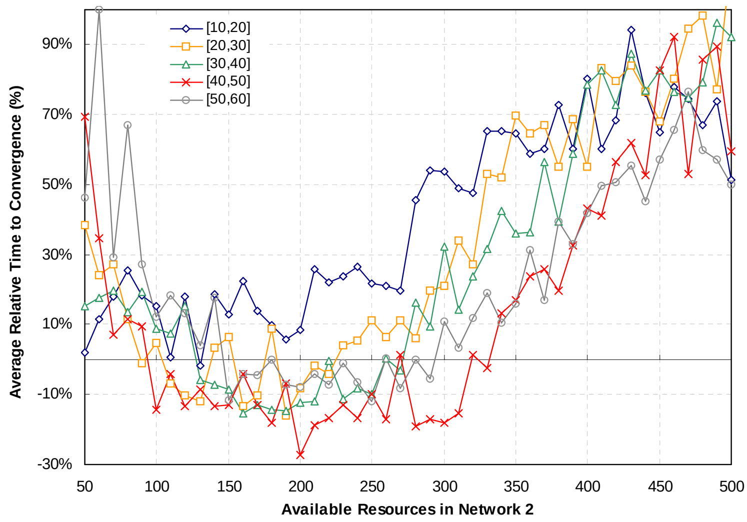



**Figure 4.39:** Comparison of the average convergence time-to-equilibrium between the two scenarios (20 replications per data point) and for the 5 experiments sets with varying user valuation when gradually increasing capacities in network 2.

results obtained from scenario 1. We vary the distribution of resources between the two providers in steps of 10. This means that the initial distribution of $Q^{(1)} = Q^{(2)} = 100$ is shifted stepwise to $Q^{(1)} = 200, Q^{(2)} = 0$. For each setup we initially run 20 replications.

Figure 4.41(a) shows the loss in welfare with the different bidding strategies and the capacity shifted between providers. As expected, in the two-provider setting, the *BalancedBid* strategy achieves nearly efficient allocations.[16] Using the alternative strategies results in a significant welfare loss. The *UtilityBased* bidding strategy performs worst

[16]In some cases we observed a slightly lower welfare of $+/-0.1\%$, which can be explained by the fact that we set $\epsilon = 1$ and therefore allow the equilibrium to be in the $2\epsilon$ range.

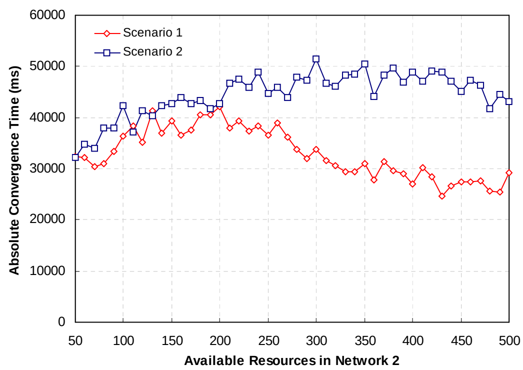
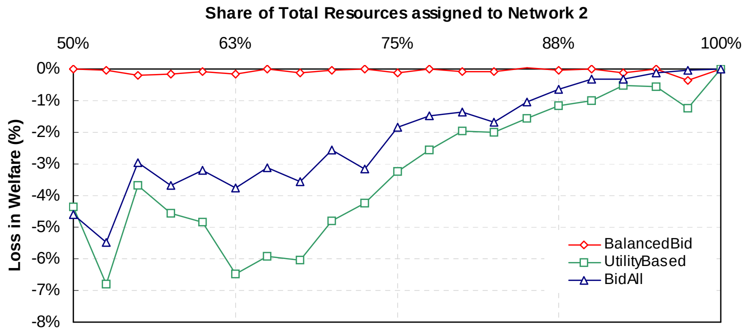


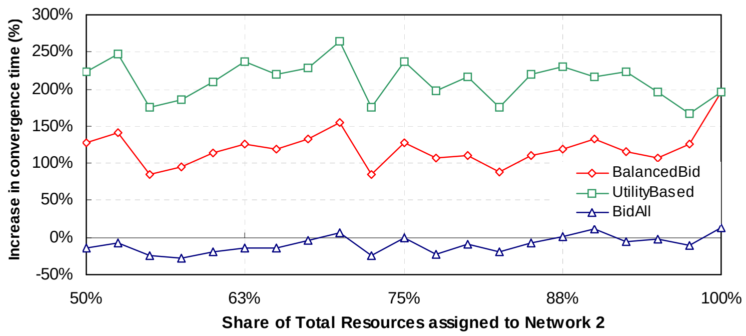

**Figure 4.40:** Comparison of the average convergence time-to-equilibrium between the two scenarios (20 replications per data point) for the user valuation interval [10, 20] when gradually increasing capacities in network 2.

(with a maximum loss of 6.8%). The use of the *BidAll* bidding strategy also resulted in significant losses with a maximum of 5.3%.

When shifting the majority of resources toward only one provider the loss in welfare becomes smaller. Because players get their resources from mainly one provider the influence of multiple-access decreases.




(a) Point estimator of the mean loss in welfare when using alternative bidding strategies and gradually shifting resources between two providers.




(b) Point estimator of the mean convergence time compared with the one auctioneer scenario when using alternative bidding strategies and gradually shifting resources between two providers.

**Figure 4.41:** Comparison of the mean convergence time with different bidding strategies.

When observing the convergence times from each of the strategies we see that efficiency with multiple-access comes at the price of longer convergence times (Figure 4.41(b)). While the *BidAll* bidding strategy finds the equilibrium solution in less time than in the one-auctioneer case, the *BalancedBid* bidding strategy needs between 100-150% more time to converge, on average. The *UtilityBased* bidding strategy performs worst with an average additional time of 200-250% the time of the one-auctioneer scenario. There is no clear correlation between the average convergence time and the distribution of resources between the two providers.

### ▸ 4.5.3 Revenue with access to multiple auctions

The PSP mechanism has been developed to control congestion in a decentralised fashion by using the negative externality for pricing network resources. In the absence of reserve prices, long-term congestion-based revenue is minimal since users adapt their demand so that negative externalities are minimised. The remaining congestion-based revenue depends on the choice of $\epsilon$.

To gain significant revenue a wireless access provider has several possibilities. He may charge a fixed subscription fee covering his variable costs or charge a fixed fee for each connection independent of resource usage. Alternatively, a provider can introduce an auction reserve price under which no resources are sold. As explained in detail in Section 4.2.2 a reserve price can be implemented as a bid $(p_0^{(j)}, Q^{(j)})$, which is kept constant over the iterative bidding process. Whenever a bidder submits a bid $(q_i^{(j)}, p_i^{(j)})$ to auction $j$ resulting in a positive allocation he is charged with at least $(p_0^{(j)}, q_i^{(j)})$. Maillé (2003) has analytically explored the consequences of a reserve price on provider revenue in the one-auctioneer setting. He concludes that the reserve price should be set close to the market clearing price $p^*$, which is reached in equilibrium where $\sum_{i=1}^{I} d_i(p^*) = \sum_{j}^{J} Q^{(j)}$ and $d_i$ expresses the demand of bidder $i$.

We experimentally extend the idea of setting a reserve price to the multi-auction case. In such a setting individual auctions can decide on the level of the reserve price, which in turn influences the decision of bidders to distribute their truthful market reply between auctions. A high reserve price in one auction may deter players from placing bids in this auction if in other auctions resources are available at a lower price. We therefore test several combinations of reserve prices and distribution of players' valuation in a two-auctioneer setting.

### Reserve price in both auctions

As in many of the previous experiments we conduct an experiment with 10 players, which are equally distributed over 2 networks. For all users we set $\bar{q} = 50$ and generate $\bar{p}$ from the interval $[10, 20]$. We set $\epsilon = 0.1$ for all replications.

We define four scenarios for this experiment. In scenario 1 and 2 the reserve price is defined to be zero in both networks. In scenario 1 users have access to only one network while in scenario 2 users have multiple-access. Both scenarios serve as reference point for the maximum welfare with the generated user population. In scenario 3 and 4 a reserve price in network 2 is introduced, which is gradually increased from 0 to 20. While in scenario 3 users have access to only one network, scenario 4 allows for multiple-access. For each data point 20 initial replications were run.[17]

We first compare the welfare produced by scenarios 1 and 3, and scenario 2 and 4, respectively (Figure 4.42). As expected, the two scenarios, single access, and multiple-access do not show any differences. Social welfare starts to decrease as soon as the reserve price is above 6, while revenue is maximised with a reserve price of 9. At this point welfare has already been reduced to 92%. With reserve prices larger than 20, no resources are sold since the valuation of all players is below this value.

[17] In single cases a higher number of replications was required to fulfill the target confidence level.

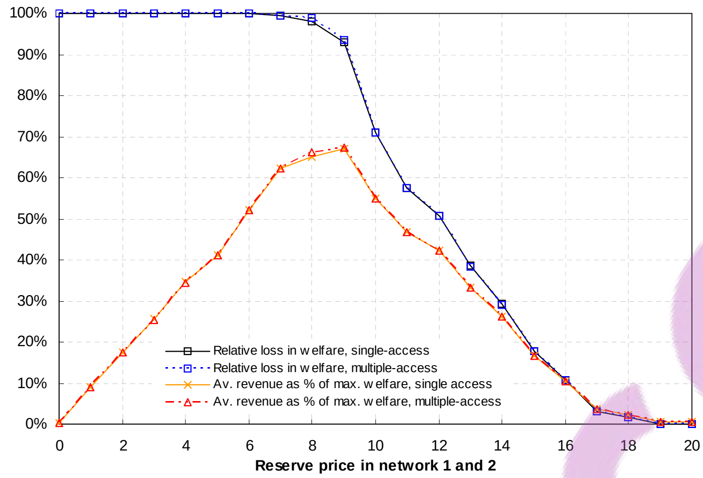



**Figure 4.42:** Comparison of revenue generation and loss in welfare when gradually increasing the reserve price in both networks ($Q^{(1)} = Q^{(2)} = 100$, $\bar{q} = 50$, $\bar{p} \in [10, 20]$)

### Reserve price in one auction

With the second experiment we introduce the reserve price at only one auction (Figure 4.43). With single-access the overall welfare is reduced to 50%. With multiple-access players have the possibility to switch to the other auction, which lets overall efficiency stay at about 60% of the maximum value. What we can also observe is that the point estimator of the mean revenue declines faster with multiple-access as players start to gradually switch to the market without a reserve price.

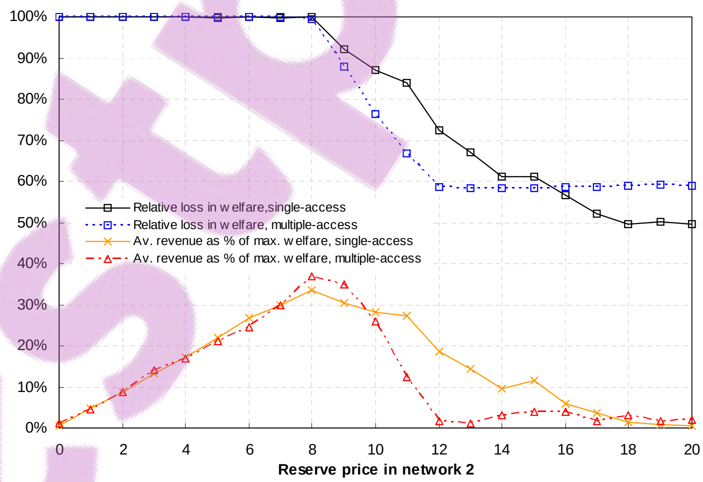



**Figure 4.43:** Comparison of revenue generation and loss in welfare when gradually increasing the reserve price in network 2 ($Q^{(1)} = Q^{(2)} = 100$, $\bar{q} = 50$, $\bar{p} \in [10, 20]$).

We are interested in how the level of congestion influences the results. We therefore lowered the maximum unit request of each player to $\bar{q} = 25$. Figure 4.44 shows that lower congestion in both networks moves the point of maximum revenue to $p_r = 4$. However, the efficiency level starts to decrease earlier.

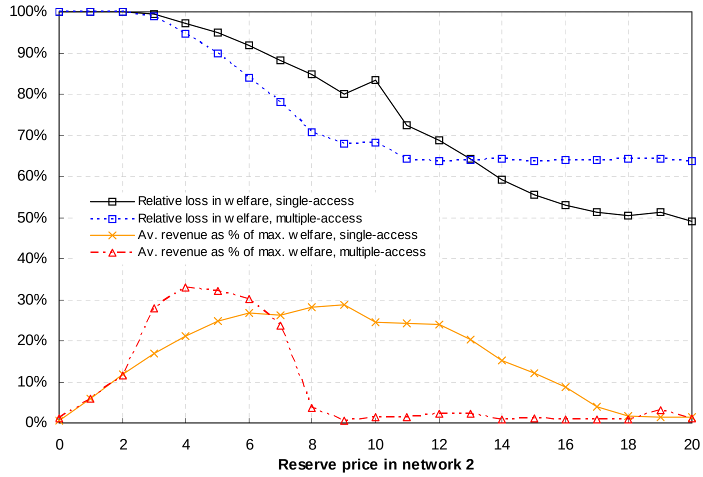



**Figure 4.44:** Comparison of revenue generation and loss in welfare when gradually increasing the reserve price in network 2 ($Q^{(1)} = Q^{(2)} = 100, \bar{q} = 25, \bar{p} \in [10, 20]$).

### Differing user valuation with reserve price in one auction

In the next experiment of this series we changed the interval for generating the user profiles from $\bar{p} \in [10, 20]$ to $[40, 50]$. In this setup agents in one network have a higher valuation for resources than the other group. We again compare the two scenarios of single-access and multiple-access, which are shown in Figure 4.45. It can be observed that revenues with single-access are much higher than in the multi-access case since agents are restricted in their choice and have to buy from the auctioneer with the given reserve price. While in the single-access case the maximum revenue is obtained at a reserve price of 27 (with a mean revenue of about 55% of the total social welfare) the maximum in the multi-access case is already reached with a reserve price of 14 (with a mean revenue of 30%).

We can conclude that introducing a reserve price for reducing customer net utility is difficult from an auctioneer's perspective and creates inefficiencies. First, the auctioneer needs to have exact knowledge about the valuation structure of all users in order to set the revenue-maximising reserve price. Second, and not surprisingly, there is a tradeoff between efficiency and revenue maximisation, which is expressed in the loss of social welfare in equilibrium. The option of bidders with multiple auction access further reduces the auctioneer's possibility of charging a reserve price which significantly differs from the reserve price of other auctioneers as bidders have the possibility to obtain resources from other sources.

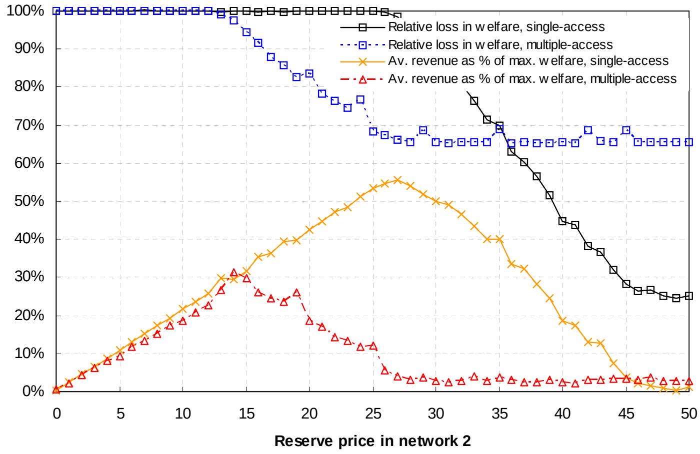



**Figure 4.45:** Comparison of revenue generation and loss in welfare when gradually increasing the reserve price in network 2 ($Q^{(1)} = Q^{(2)} = 100$, $\bar{q} = 25$, $\bar{p} \in [10, 20]$ for user group 1 and $\bar{p} \in [40, 50]$ for user group 2).

## ▸ 4.5.4 Convergence behaviour with mixed access to multiple networks

In this section we are concerned with three main issues. The first one is convergence to equilibrium. While in the one-auctioneer case it has been analytically shown that the system converges to equilibrium in finite time for $\epsilon > 0$, this has not yet been fully understood for the multi-auctioneer case. The second issue to be analysed is the time the market consisting of multiple auctions needs to converge to equilibrium if players implement the *BalancedBid* strategy. While the previous experiments have given a first indication that the system converges, we also need to analyse convergence in more complex setups. The last aspect is the question of convergence when a system in equilibrium is disturbed by an external event. Such an event can be a new user joining the system or an existing user leaving the system.

We cover the first two issues in the first subsection and elaborate on the third issue in the second part of this section.

### Convergence behaviour in interdependent networks

In the previous sections we have found some first indications that with all players implementing the *BalancedBid* strategy in the multiple-access case, the system converges to a stable equilibrium in finite time in which social welfare is maximised. However, we have not yet looked into the issue of convergence in a more complex setup of networks, in which access options vary between users. In the last section we have introduced the notion of ”interdependence” between different auctions. With interdependence we describe the situation in which players have access to multiple auctions and access option varies between users. For example, if player 1 has access to network $A$ and $B$ and player 2 has access to auctions $B$ and $C$, auctions $A$ and $C$ are potentially interdependent if both players decide to bid on both auctions.

We experimentally explore this issue by creating different setups, in which convergence is less straightforward. We first create a simple experimental setup with two auctions. Some users have access to only one auction. Some other users have access to both auctions. We start with an initial setup of 5 single-access users in each auction and 5 users with multi-access. Both auctions offer resources of $Q^{(1)} = Q^{(2)} = 100$ and all agents have a maximum demand of $\bar{q} = 50$. The maximum marginal unit price is again randomly generated from a uniform distribution on the interval $[10, 20]$.

With this setup we conduct two experiments. In the first experiment we gradually increase the single-access agents in both networks. For each data point 20 replications with randomly generated user profiles were run. In the second experiment we gradually increase the number of multi-access users. The results are shown in Figure 4.46(a) and Figure 4.46(b), respectively.

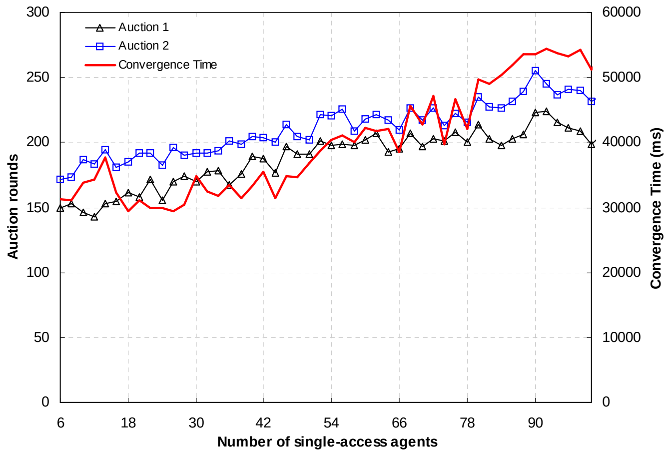
It can be observed that an increasing number of single-access agents does not heavily increase the number of bidding rounds needed to reach the equilibrium. An increase from 6 single-access agents (3 in each network) to 100 agents increases the point estimator of the mean auction rounds by only about 35%. In contrast, an increasing number of multi-access agents has a strong influence of the number of auction rounds. Increasing the number of multiple-access agents from 3 to 50 increases the auction rounds from 150 to 750. With an increasing number of multi-access agents the slope decreases. We can also observe that increasing the number of multi-access agents correlates with a longer time to converge to equilbrium.[18]

Another interesting question about convergence behaviour arises if we create a circular interdependence between auctions. We create a setup with three networks and three user groups. User group 1 has access to networks 1 and 2, user group 2 has access to network 2 and 3, while the last group has access to networks 1 and 3. We start with an initial population of 3 agents in each group and gradually increase the number of agents. As in the last example we set $Q^{(1)} = Q^{(2)} = Q^{(3)} = 100$, $\bar{q} = 50$ and randomly generate the maximum marginal unit price from a uniform distribution on $[10, 20]$.

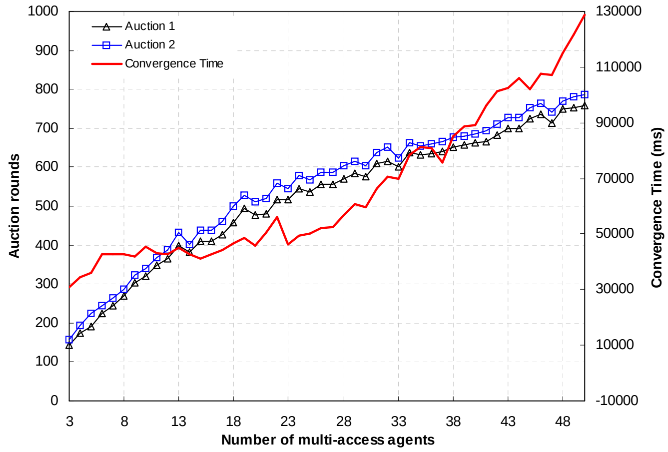
Figure 4.47 shows the results of this experiment. We can observe the same correlation as in the last experiment. The auction rounds of all three auctions increase with an increasing number of multi-access agents in the system. However, even with interdependence created by agents bidding on different auctions, all auctions converge after finite time-to-equilibrium.

In the last experiment of this section we explore the consequences of more complex interdependence between networks through multi-access agents. We create 6 networks, which are all interlinked by 3 users (Figure 4.48). As before, we set $Q^{(1)} = Q^{(2)} = Q^{(3)} = 100$, $\bar{q} = 100$, and randomly generate the maximum marginal unit price $\bar{p}$ from a uniform

[18] In principle there eshould be a linear dependency between the number of auction rounds and convergence time. However, convergence time is also influenced by the system load and the complexity of the overall simulation. Since our lab setting was limited to one machine we could not test the influence of distributing the simulation over several machines (e.g., with seperate machines for bidder agents and auctioneer agents).




(a) Point estimator of the mean auction rounds and mean convergence time-to-equilibrium for an increasing number of single-access users.




(b) Point estimator of the mean auction rounds and mean convergence time-to-equilibrium for an increasing number of multi-access users.

**Figure 4.46:** Convergence with multi-access agents.

distribution on [10, 20]. We gradually increase the number of multi-access agents with access to networks 1 and 2. For each data point we initially run 20 replications.[19] Figure 4.49 shows the number of bidding rounds at each auction for different numbers of agents with access to network 1 and 2. As can be observed the number of agents does only influence the number of auction rounds in these two networks but does not influence the number of auction rounds in the other networks. This means that the *BalancedBid* strategy can be used in a more complex setting in which users have access to different networks and auctions are interdependent.

[19]For some data points a higher number of replications was required to satisfy the target confidence level. The highest number of replications required was 54.

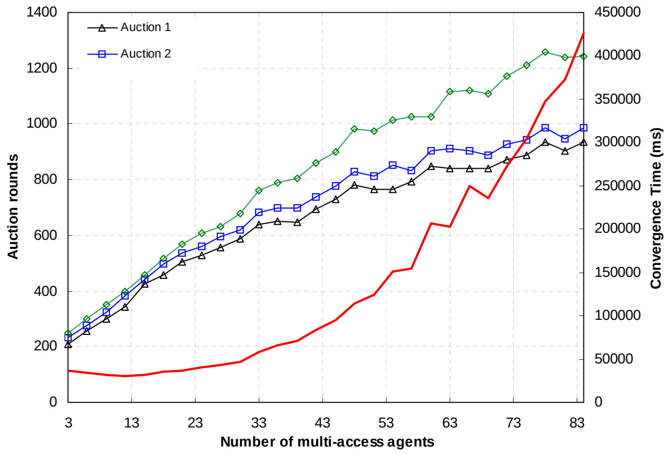



**Figure 4.47:** Point estimator of the mean auction rounds and mean convergence time-to-equilibrium for an increasing number of multi-access users in all three user groups.

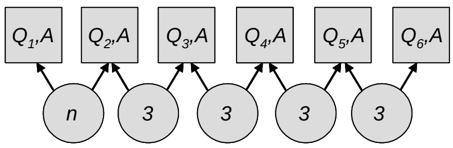



**Figure 4.48:** The experimental setup with 6 auctions and an increasing number of multi-access agents with access to auctions 1 and 2.

### Convergence behaviour after external events

In all previous experiments we have assumed that agents have a continuous demand for network resources over the entire simulation period. In wireless networks this may not be realistic as users join or leave access points by either changing their position or starting or stopping new services on the mobile terminal. Therefore, we are interested in the behaviour of the system with users stochastically entering or leaving the auctions. More precisely, we would like to understand the consequences of agents entering or leaving a market, which is already in equilibrium, and to analyse the correlations between the user setup and convergence time.

In the first experiment of this section we define 2 auctions and initially 3 agents having access to both auctions. After the market has come to equilibrium we introduce an additional multi-access agent entering the market. We measure the additional auction rounds needed to find the new equilibrium. As before, we keep the maximum demand fixed at $\bar{q} = 50$ and generate the maximum marginal unit price from an interval $[10, 20]$. The agent entering in equilibrium has a fixed maximum marginal unit price of $\bar{p} = \{10, 15, 20\}$. We then gradually increase the number of initial agents in the system. To obtain the estimator for each data point we initially run 20 replications with $\epsilon$ being set to 1.

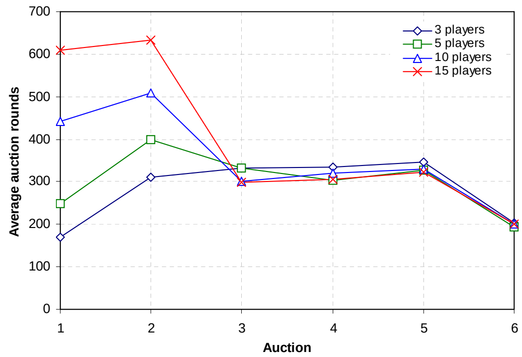
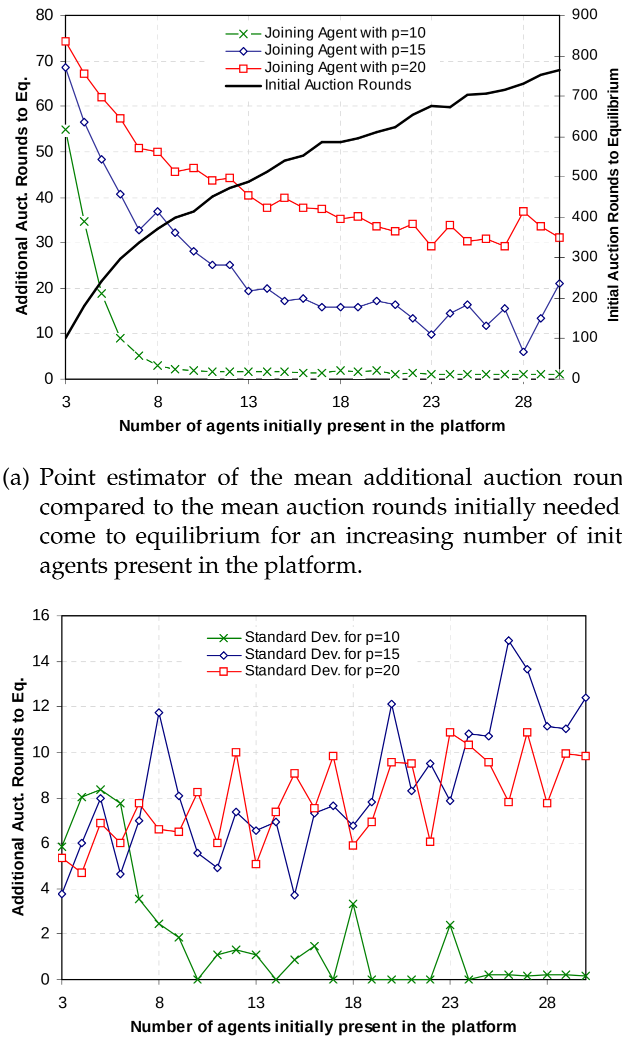



**Figure 4.49:** Point estimator of the mean auction rounds to equilibrium at each auction for an increasing number of multi-access users in networks 1 and 2

Figure 4.50(a) shows the auction rounds to equilibrium and the additional auction rounds needed to find to the new equilibrium after the new agent has entered the market for each of the three values for $\bar{p}$. With a maximum unit valuation of $\bar{p} = 10$ we observe that the additional auction rounds to the new equilibrium quickly drops to zero when more than 10 users are initially present in the platform. For $\bar{p} = 15$ and $\bar{p} = 20$ we also see a decrease in auction rounds to find back to equilibrium. However, the decrease is slower and point estimator of the standard deviation increases (Figure 4.50(b)). Except for $\bar{p} = 10$ we are unable to produce estimators within the target confidence level because the standard deviation strongly increases and the required replications can not be conducted. Besides this, the estimators of the means clearly show the trend that an increasing number of users initially present in the market decreases the number of required auction rounds after a new user joins the market.

In the next experiment we are interested in the consequences of one player leaving the market, which has been in equilibrium. Keeping all parameters of the last experiment fixed we delete one agent from the platform once the initial equilibrium has been reached. Again, we run 20 replications for obtaining the point estimator for the mean and standard deviation. Figure 4.51(a) shows the point estimator of the mean for the auction rounds to the initial equilibrium and to the subsequent equilibrium for all three values of $\bar{p}$ for the user leaving the market. As in the case of a player joining the market, the number of subsequent auction rounds quickly drops, with the quickest drop observed for $\bar{p} = 10$. However, the drop is not as steep as in the case of a joining agent and for $\bar{p} = 20$, the number of auction rounds to the subsequent equilibrium stabilises at around 120 additional rounds. Figure 4.51(b) shows the estimator for the standard deviation, which quickly increases with more than 3 users initially present in the market and then remains on a high level. As in the previous experiment, we are not able to produce the target confidence level of the estimators since the required number of replications is too large for $\bar{p} = 15$ and $\bar{p} = 20$.




(a) Point estimator of the mean additional auction rounds compared to the mean auction rounds initially needed to come to equilibrium for an increasing number of initial agents present in the platform.



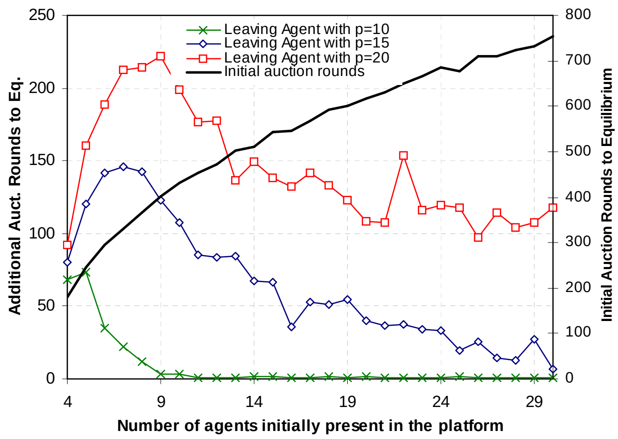
(b) Point estimator of the standard deviation of the initial auction rounds and additional auction rounds after the new agents joins the market.

**Figure 4.50:** Simulation results for a new agent entering a market already in equilibrium.

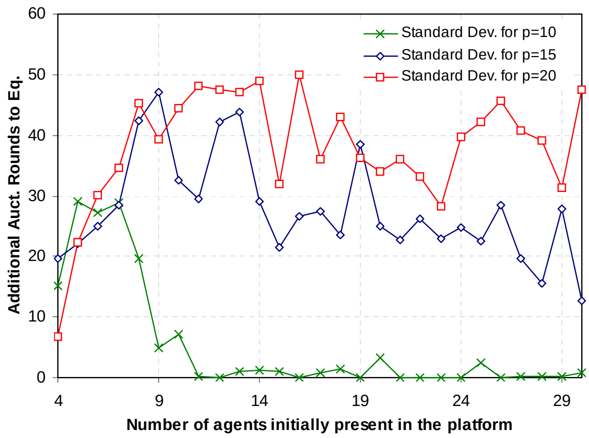
## ▶ 4.6 Simulation Results: Multi-cell case with Three Providers

In this section we analyse how the developed pricing mechanism performs in a complex network setting with a large number of user agents and with competitive providers running multiple access points in a limited geographical region. Beside observing the general performance of PSP in such a setting with multiple access to different networks, we aim at comparing the performance to alternative pricing schemes, which do not include the users' utility for resources in the allocation process. We can summarise our research objectives in three questions:

- How does the PSP mechanism, when implemented at multiple base-stations of competing providers, and users employing resource bundling (via multiple-access),




(a) Point estimator of the mean additional auction rounds compared to the mean auction round initially needed to come to equilibrium for an increasing number of initial agents present in the platform.




(b) Point estimator of the standard deviation of the initial auction rounds and additional auction rounds after the new agents joins the market.

**Figure 4.51:** Simulation results for an agent leaving a market already in equilibrium.

perform in a setting with hundreds of agents randomly requesting resources?

- How does aggregated consumer utility and provider revenue in a multi-access scenario compare to a scenario with single-access, or to simpler pricing schemes?
- How large is the influence of multiple-access on the connection dropping rate under different network loads?

As we have learned from the previous experiments the proposed approach does not equally suit all kind of service types or network setups for two reasons. First, the PSP auction mechanism requires multiple iterations to converge to equilibrium. This makes the system potentially unsuitable for short-lived flows. Second, PSP cannot guarantee a certain bandwidth share but changes allocations according to the overall load and the

| | Market Share | | |
|---|---|---|---|
| | Prov. 1 | Prov. 2 | Prov. 3 |
| Cell 1 | 55% | 35% | 10% |
| Cell 2 | 55% | 35% | 10% |
| Cell 3 | 55% | 35% | 10% |
| Cell 4 | 60% | 40% | - |
| Cell 5 | 60% | 40% | - |
| Cell 6 | 60% | 40% | - |
| Cell 7 | 60% | 40% | - |

**Table 4.7:** Market shares of the three network providers.

distribution of user utility in the system. Therefore, only applications being able to adapt to changing bandwidth allocations go well with the proposed model.

In the following, we draw a scenario for streaming multimedia services, namely audio and video, which are well suited for this purpose. Most multimedia applications can adapt to different network speeds and use a buffer to bridge short periods when less network resources are available.

### ▸ 4.6.1 Provider setup

All users in the defined region are serviced by three competitive network providers, each with a certain coverage and market share. While the first two providers cover the entire region with a separate base station in each cell, provider 3 is only present in the three cells with the highest user density. Figure 4.52 illustrates the provider setup graphically. Table 4.6.1 gives the market shares of the providers for all seven cells. Provider 1 is the market leader with 55% and 60% market share. Provider 3 can be seen as a new market entrant, who concentrates on the areas with high user ratios and has gained 10% market share in those areas.

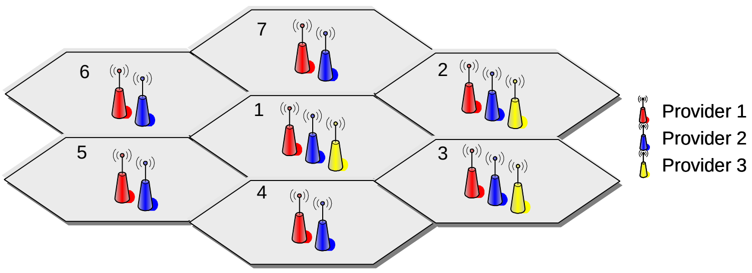



**Figure 4.52:** Provider setup in the seven cells.

We do not define the specific access technology used by each provider to supply terminals with transport services. Instead, we assume that the used access technologies can all support the required QoS level, which allows giving guarantees on bandwidth shares once the stable allocation has been found. Furthermore, we assume that providers assign a fixed capacity on the link layer to the service class in the scope of the simulation, which is set to 1.024kbps for all base stations and providers. This ensures that service guarantees can be made to the active flows in the cell. In TDMA-based access networks this could be realised by assigning a fixed number of time slots to the service class and to ensure that active users in this service class are supplied with sufficient transmission power to guarantee a certain maximum bit error rate (BER).

### ▸ 4.6.2 Service description

We define two different types of service to be requested by user agents, namely audio and video streaming services. The relevant literature on audio and video encoding over wireless access networks has been consulted to set realistic values for guaranteed bandwidth and minimum bandwidth requirements. Using variable bit rate (VBR) video compression techniques, such as proposed in Jiang and Zhuang (2004), enable a managed reduction in video quality when less resources are available. We set the maximum available bit rate for video users to 384kbps, which delivers a satisfactorily video even on larger screens (Sun and Li, 2005). We allow for a degradation of the flow to 32kbps before the connection is dropped by the user. Several studies, such as Sun and Li (2005) or Ries et al. (2005) show that the users' utility is considerably lower for low-bandwidth video quality.

For audio streaming we set the maximum guaranteed bit rate to 144kbps. Together with the currently available and commonly used audio compression standards such as MP3, WMV or AAC, the perceived quality is comparable with near-CD quality. With degrading bandwidth availability the sound quality reduces to radio quality depending on the audio source and complexity of the compression mechanism. We set the minimum acceptable bit rate to 16kbps, under which users will drop the session.

We further assume that applications use a local buffer, which is large enough to bridge periods in which resources are reallocated due to new service arrivals or established flows leaving the cell. Since the convergence durations in the chosen setup are usually small (smaller than 2 seconds), this seems a realistic assumption.

### ▸ 4.6.3 User profiling

The question about user utility for the two service types is difficult from multiple perspectives. First, results from empirical studies about the acceptance and willingness-to-pay for future mobile services cannot be directly translated into the user utility from such services on a time scale of seconds or minutes. Second, the valuation of such services is strongly driven by the application content, and the valuation share for the communication services cannot be distinguished by the user or decoupled from the valuation for the content.

Instead of basing our utility model on empirical data, such as presented in Sun and Li (2005) for content-based streaming services, we develop a hypothetical valuation model for both service types, which takes into account the differences in user utility between audio and video streaming. For the transport services for audio streaming we assume

| | Number of users | | | |
|---|---|---|---|---|
| | Total | Prov. 1 | Prov. 2 | Prov. 3 |
| Cell 1 | 175 | 96 | 61 | 18 |
| Cell 2 | 150 | 82 | 53 | 15 |
| Cell 3 | 125 | 69 | 43 | 13 |
| Cell 4 | 100 | 60 | 40 | - |
| Cell 5 | 75 | 45 | 30 | - |
| Cell 6 | 50 | 30 | 20 | - |
| Cell 7 | 25 | 15 | 10 | - |

**Table 4.8:** Number of users per cell and provider.

an average utility of \$0.10 per time-unit and set the average utility for video streaming services to \$0.15.

With the maximum bandwidth of 144kbps for audio and 384kbps for video we can derive the per-second utility for 1kbps for each service type. Since the highest utility for audio services is already reached at 144kps, the per-unit valuation is higher than with video services. In consequence, users requesting bandwidth for audio streaming services will always receive service as long as sufficient resources are available while video streaming users will reduce their requested flows faster in times of congestion and will drop their services first.

In the simulation each agent requesting services creates a random utility, which can be ±50% around the defined average utility for both user types, using a uniform distribution. In this way a user population with a random utility profile is created.

As with the previous experiments we make use of the parabolic utility function as proposed in Semret (1999). Since this work provides a detailed analysis of the match of such utility functions, we omit a further discussion here.

Different to the use in Semret (1999), we define a minimum quantity under which the user's utility is defined as zero. From this threshold value the utility is positive and defined by the parabolic utility function. To include blocking into the analysis of the following experiments, we assume that users leave the auction whenever the best truthful reply falls below the threshold value.

### ▶ 4.6.4 User setup and service scheduling

The service area shown in Figure 4.52 defines seven areas that differ in network coverage and user density. Table 4.6.4 provides the details about the number of users per area type. We assume that users request video and audio streaming services with the same probability. Therefore, 50% of the connection requests will be video streaming requests and 50% will be audio streaming requests.

We set average inactivity time of each agent to 60min and the average service duration to 10min using an exponential distribution with the given means. To keep the setup reasonably simple we do not distinguish between video and audio streaming services.

While, due to a lack of empirical material, the inactivity and activity times have been arbitrarily chosen to generate sufficient traffic for a partial network congestion, similar values have been the base in comparable simulation studies such as in Murray (2005) or Alwis (2005).

### ▸ 4.6.5 Charging policy

The charging model developed for this simulation consists of two elements: a *volume-based usage charge*, which is imposed independently from the current state of the network, and a *congestion-based charge*, which results from the auctioning process when demand is higher than the overall supply in the service area. While the usage-based component allows the provider to recover variable costs from network operations, the congestion charge serves as additional revenue from the congestion control mechanism to compensate the provider for the negative externalities created by the users excluding other users from the service.

For all experiments we set $\epsilon = 100$. This, at the one hand, ensures a fast convergence at all PSP auctions, and, on the other hand, creates non-zero congestion charges because users reduce their demand only up to a certain threshold value. It needs to be emphasised that the definition of $\epsilon$ critically determines the obtained results in terms of revenue from congestion-based charges. Prior to the actual simulation study we have conducted multiple tests of the simulation experiments and have tried to balance convergence time and efficiency, which resulted in the selection of $\epsilon = 100$.

The usage-based charge is defined as a fixed fee per resource-unit raised by the provider for each time-unit the resource is assigned to the user. To not exclude users from at least obtaining the minimum allocation required to run the streaming service the usage-based charge is defined according to the user with the lowest willingness-to-pay as shown in Figure 4.53. We set the usage charge identical for all providers.

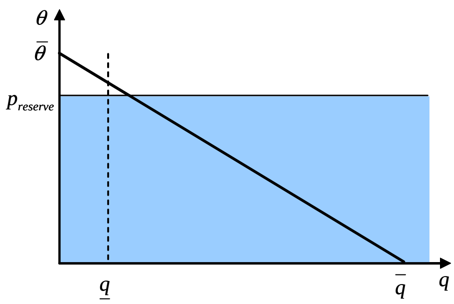



**Figure 4.53:** The reserve price $p_0$ is defined below the marginal valuation for the minimum resource quantity $\underline{q}$ of the user with the smallest willingness-to-pay as to not exclude users from participating in the auction.

### ▸ 4.6.6 Simulation scenarios

We define five scenarios, which differ in the way resources are allocated between users.

S1 In the first scenario resources are allocated according to a simple admission scheme on a first-come first-served basis. Users with a new request are either admitted to the network with the requested guaranteed bandwidth if sufficient capacities are available or are rejected. We assume that users do not retry to get connected once they have been rejected. The provider revenue from the different cells consists solely of the fixed usage-based charges obtained from all active streams.

S2 In the second scenario we implement a simple flow-management scheme in addition to the admission scheme from scenario 1. If the cell is not congested users are admitted with the maximum streaming bandwidth. If congestion occurs (a new request arrives but insufficient capacity is available), the bandwidth of all users is proportionally reduced to accommodate for the new request. This is done until all users only receive the minimum bandwidth for running their service. This procedure is similar to the proposal made in Lataoui et al. (2000), which introduces a *Subscriber Degrade Descriptor* (SDD) to define the priority of bandwidth degradation for each stream. As in the first scenario, the provider revenue only consists of the static usage-based charges.

S3 In the third scenario we employ the PSP auction mechanism at each base station to allocate resources between users. Note that in this scenario no admission control is performed but users are always accepted to connect to the network. However, a user may drop his connection in case he cannot obtain the minimum bandwidth to run the streaming service. It is assumed that users dropping their connection do not try to reconnect. In this scenario the provider collects revenues from two sources, the usage-based charges and the congestion charges as a result of the PSP auction.

S4 The fourth scenario introduces the possibility of multiple-access and resource bundling from different sources. We assume that provider 3 allows users statically subscribed with other providers to connect to its network and to obtain additional resources. We further assume that 25% use this offer and use the *BalancedBid* strategy to manage their bids to the different auctions. As in the third scenario user connections are dropped if the total bandwidth obtained from one or two networks falls below the minimum level. As in scenario 3 providers collect revenues from two sources, the usage-based charges and the congestion charges as a result of the PSP auction.

S5 In addition to provider 3 in scenario 4, in the fifth scenario, provider 2 opens access to its network for non-subscribed users. Again, 25% of all users not originally subscribed with provider 2 make use of this offer.

### ▸ 4.6.7 Simulation output analysis

The type of statistical analysis required for the described experiments differs from the previous section. Since the simulation is now run over an extended period without reaching a steady state we need to apply a different technique to obtain statistically meaningful results.

Instead of recording only the final results of the allocation we now continuously collect data about revenue, welfare, and blocking from all agents in the platform and aggregate this data on a per-cell level. We are then able to obtain an average over the entire time of the simulation run.

For each scenario we run 10 independent replications, each of 11 hours length. To eliminate the influence of the initial transient period, we delete the data obtained in the first hour of the simulation. We then take the average over the 10 replications to obtain the point estimators. While, due to the characteristics of the data, no statement can be made about the level of confidence, with the applied method we can eliminate the effects of specific effects occurring in a single replication.

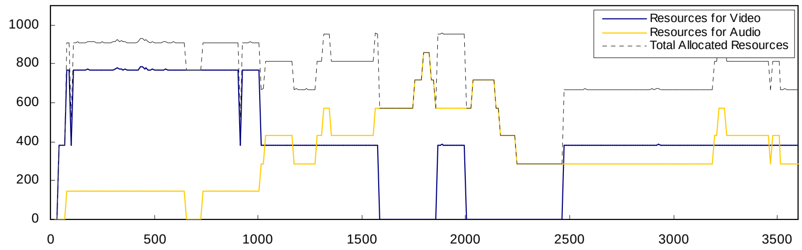
We present the analysis in two parts. The first part provides some insights in the allocation dynamics over time for all five scenarios. Such dynamics let us understand the differences in the allocation behaviour and the reason for blocking incoming streams. In the second part we compare the obtained point estimators from all five scenarios to draw conclusions in regard to welfare, revenue and connection blocking and dropping.

### Analysis of the simulation dynamics

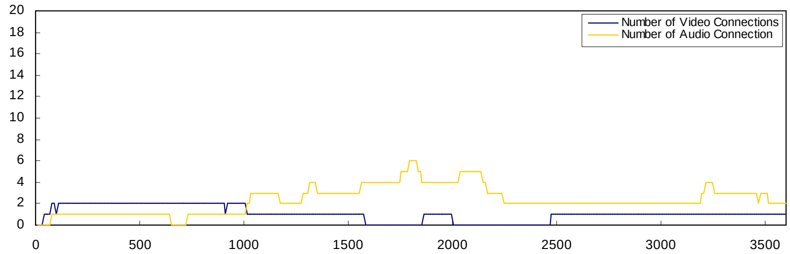
We start with the analysis of the allocation activity of one base station over time. We aim at understanding the influence of the allocation scheme on resource allocation behaviour and the consequences on the number of admitted streams in each service category. To create all graphics we arbitrarily picked one of the simulation runs and visualised the results over time for the first hour of the simulation for each scenario. We selectively look at the allocation of resources to the two service types, the number of admitted streams, the average allocation of resources per stream and service type, and the revenue from usage and congestion.

Figure 4.54 shows the total quantity of resources allocated to video and audio streams (Figure 4.54(a)) and the number of admitted streams (Figure 4.54(b)) for scenario 1. It can be observed that the total capacity is never fully allocated due to the inability of the base station to reduce the allocation to existing flows if new requests arrive. The base station is unable to accommodate more than two video streams simultaneously because of the high resource requirements of such flows in the presence of already admitted audio streams. The maximum number of simultaneous audio streams reaches six around $t = 1.800$. Since usage-based revenues are collected based on a fixed fee per resource unit the revenues over time are directly proportional to the allocation of resources.

In scenario 2 the capacity of the selected base station of provider 1 is almost always fully allocated between active streams. Since the allocation scheme can change the allocation to existing flows if new requests arrive (Figure 4.55(a)), resources can always be fully allocated. However, all streams are admitted regardless of the users' individual utility for the allocated resources. In consequence, all streams of the same type receive an identical allocation, thus making the overall resource allocation inefficient from a social welfare




(a) Total quantity of resources allocated to video and audio streams.



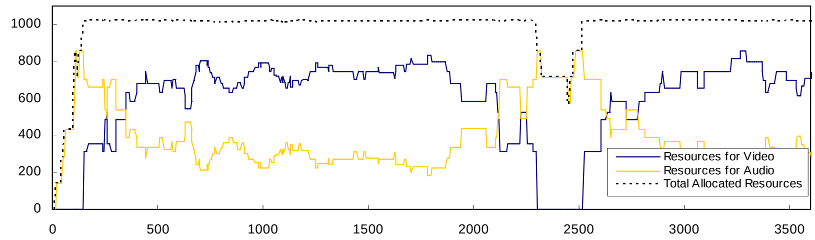

(b) Number of active video and audio streams.

**Figure 4.54:** Scenario 1: Analysis of the simulation dynamics (data shown for provider 1 in cell 2).

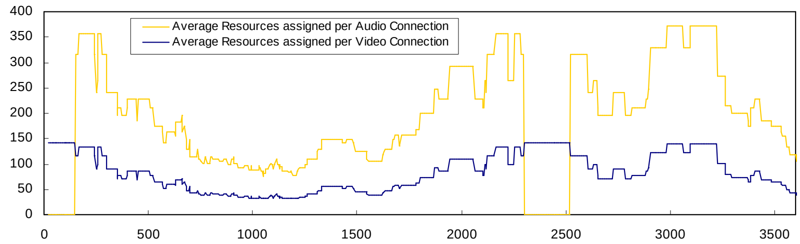
perspective. The number of admitted streams is significantly higher than in scenario 1, reaching 10 audio streams and 8 video streams around $t = 1,000$. Around $t = 2,300$ the number of active video streams drops to zero and all audio streams are admitted with their maximum bandwidth. Figure 4.55(b) shows the average size for each active stream for the two service types. The allocation for audio streaming drops to 34kbps arround $t = 1,000$, while the lowest average allocation for video streaming is 71kbps at around the same time stamp. During times of free supply (e.g. at $t = 2,300$) audio streams receive their maximum allocation of almost 140kbps.

Figure 4.56(a) depicts the resource allocation between service types for scenario 3, in which the base station uses the PSP auction to allocate resources. The total capacity is always fully allocated between active streams as long as demand exceeds supply. It can be observed that the allocation is more "bursty" due to the iterative bidding mechanism when new streams arrive or active streams leave the network. Figure 4.56(b) shows the average allocation per stream for the two service types. During congestion the average allocation for video and audio streams drops to 54kbps and 58kbps, respectively. Compared to scenario 2 the allocation for video streams is further reduced due to the smaller per-unit valuation. It can also be observed that the average allocation usually does not change proportionally with the number of active streams. This is because allocations are based on the user utility. If a new stream with low utility is added it may not take away resources from the same service group but only resources from streams of the other service type.

In Figure 4.56(c) the revenue per time-unit is shown. Since the base station also collects congestion-based fees, four categories are distinguished: usage-based charges




(a) Total quantity of resources allocated to video and audio streams.




(b) Mean quantities of resources allocated to video and audio streams.

**Figure 4.55:** Scenario 2: Analysis of the simulation dynamics (data shown for provider 1 in cell 2).

from video and audio, and congestion-based charges from video and audio. Since the total capacity is allocated and the usage-based charge is fixed, the total usage-based fee adds up to a constant amount. The congestion-based charges are charged during the iterative convergence phase and m,ay remain on a certain level in equilibrium.

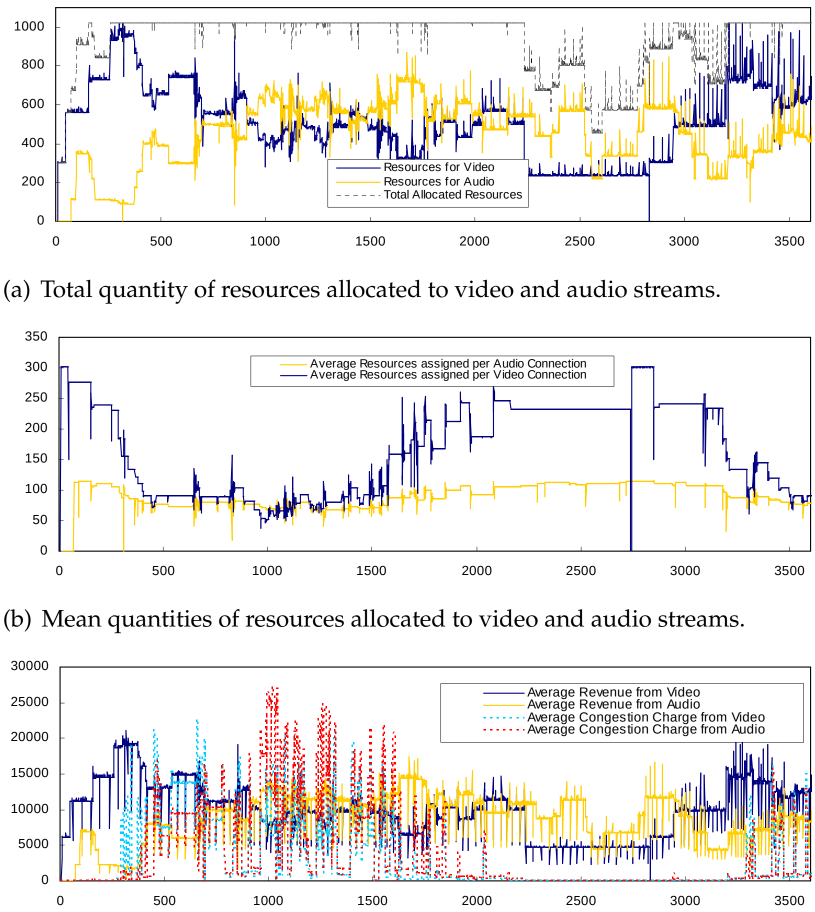
Scenario 4 and 5 allow selected users to bundle resources from multiple networks by using the *BalancedBid* strategy to manage their bids to the single auctions. As in scenario 3 resources of the base station are almost always fully allocated to active streams. The average allocation per stream is lower since users are able to bundle resources from multiple sources (Figure 4.57).

We can also observe that, with the additional demand from the multi-access users, provider 3 is able to collect higher revenues from the additional load as well as revenue from congestion charges due to the increased demand in his cell. Figure 4.58 compares the revenue per time-unit for provider 3 in cell 2 for scenario 3 and 4. While these two runs cannot be directly compared it can be seen that revenues from network usage are significantly higher and congestion-based charges can be collected for a limited time interval.

### Analysis of the total point estimators from multiple replications

In this section we look at the point estimators, such as revenue per provider, social welfare, and blocking and dropping rates, which have been derived with the method described in Section 4.6.7.

Figure 4.59 shows the point estimator of the mean revenue for all three providers and




(a) Total quantity of resources allocated to video and audio streams.

(b) Mean quantities of resources allocated to video and audio streams.

(c) Usage-based and congestion-based revenue for the two service types.

**Figure 4.56:** Scenario 3: Analysis of the simulation dynamics (data shown for provider 1 in cell 2).

all network cells. Changing the allocation scheme from a first-come first-serve basis to a flexible allocation scheme, in which existing streams can be reduced to accommodate new requests, increases provider revenues mainly due to the increased usage of the available capacity. When we compare the results from scenario 2 and 3 we see that revenues from resource usage stay nearly constant since the usage price is static in both scenarios. However, providers obtain additional revenue from the dynamic congestion charges in times of high demand. In scenario 4 and 5 overall provider revenue increases since resources in underused networks from provider 2 and 3 can now be sold to users not subscribed with such providers. Charges from congestion stay relatively constant.

Additional to the total numbers for all cells we analyse the revenue obtained by all providers in cell 1 (Figure 4.60). Figure 4.60(a) shows the revenues in cell 1. The congestion-based revenues make up a larger share of the overall revenues since congestion is more likely due to the high user density. Additionally, no direct influence of multiple access on congestion charges can be observed but usage-based charges increase due to higher resource usage. Figure 4.60(b) shows the revenues only for provider 3. The

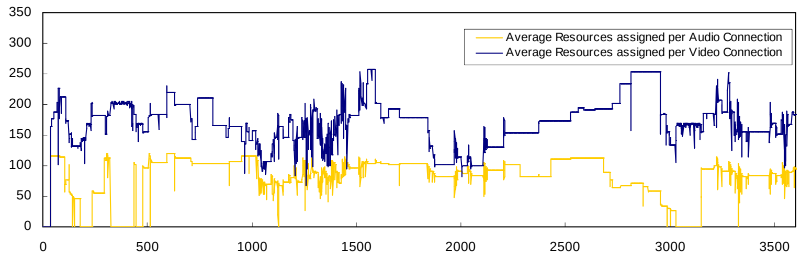
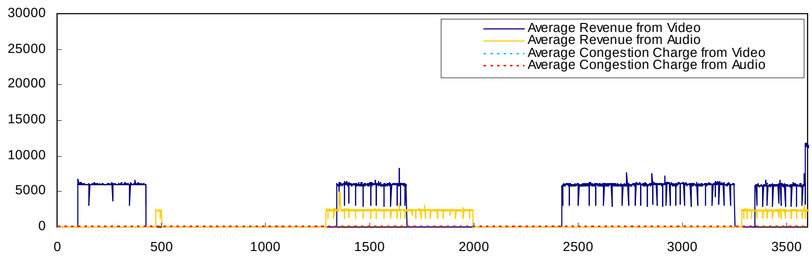
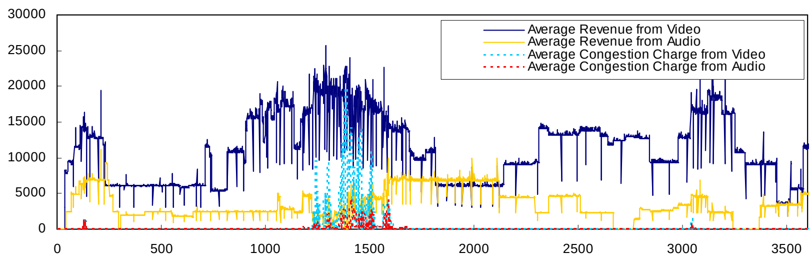



**Figure 4.57:** Mean quantities of resources allocated to video and audio streams (data shown for provider 1 in cell 2).




(a) Scenario 3: Mean revenue per time-unit for provider 3 in cell 2.

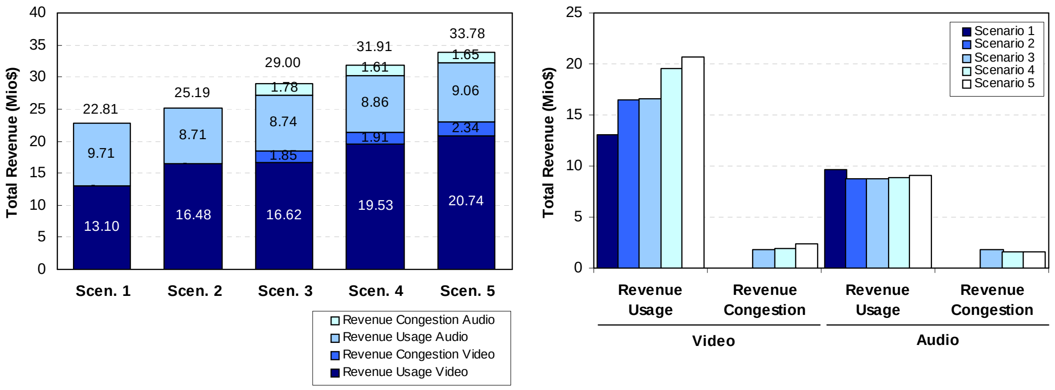



(b) Scenario 4: Mean revenue per time-unit for provider 3 in cell 2.

**Figure 4.58:** Comparison of the mean revenue per time-unit between single-access (scenario 3) and multi-access (scenario 4) for provider 3.

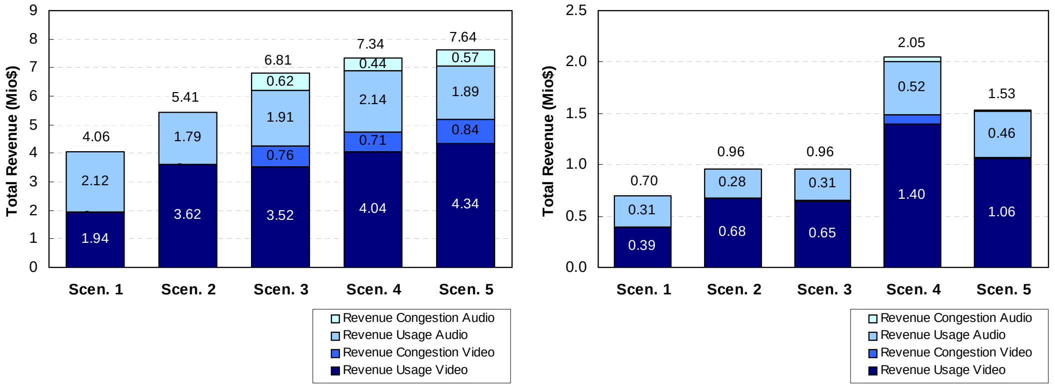
revenue more than doubles when the provider opens up its network for multiple access. The additional move of provider 2 in scenario 5 again takes away about half of the additional revenues. We can also observe that provider 3 only earns congestion-based charges in scenario 4.

Figure 4.61(a) shows the point estimator of the total mean consumer surplus for the entire system consisting of all cells. The use of a dynamic allocation scheme (scenario 2) drastically improves surplus from 31.47 to 51.19 million monetary units. Surprisingly, the use of the PSP auction and the truthful declaration of the users' utility does not result in a significant increase in consumer surplus. The difference in the users' utility, even if it varies by a maximum factor of 4.5, seems not sufficient to create large allocation inefficiencies by using a central allocation mechanism. Another explanation for this is that the iterative process, in which the allocation to individual streams is unstable, causes inefficiencies in the PSP scenarios.




(a) (b)

**Figure 4.59:** Point estimator of the total mean revenue from usage and congestion for all providers and cells.



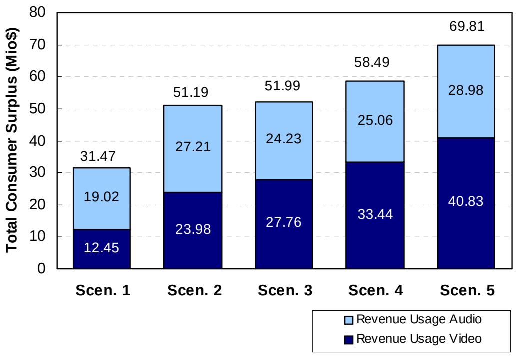

(a) Point estimator of the total mean revenue for all providers in cell 1.
(b) Revenues for provider 3 in cell 1.

**Figure 4.60:** Point estimator of the total mean revenue for cell 1.

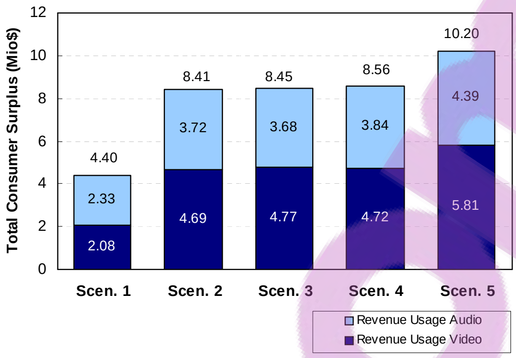
If we allow users access to multiple networks surplus increases to 58.49 million monetary units (scenario 4) and 69.81 million monetary units (scenario 5). The increase in surplus is larger for video streams since such requests had to cut back stronger because of a consistently lower per-unit valuation.

Additionally to the overall mean consumer surplus we have analysed the situation in cell 4, in which only the first two providers are present (Figure 4.61(b)). Again, we can observe the drastic increase of consumer surplus from scenario 1 to scenario 2. Surplus is almost equals for scenario 2-4. Again, the use of the PSP auction does not have a significant influence on the point estimator of total mean consumer surplus. With the possibility for users to obtain additional resources from provider 2, surplus increases by about 19% from 8.56 monetary units to 10.2 monetary units, both for video and audio streams. As in the system case the surplus increase is larger for video streams (+23%).

Figure 4.62 gives the blocking and dropping rates for video streams for cell 1 and the network of provider 1. The ratios have been derived by dividing the number of blocked




(a) Point estimator of the total mean consumer surplus for all users in all cells.




(b) Point estimator of the total mean consumer surplus in cell 4

**Figure 4.61:** Point estimator of the total mean consumer surplus divided into surplus from video streams and audio streams.

video streams by the total number of connection requests. As expected the number of dropped video calls is very high in scenario 1 (39.99%). The minimum connection blocking occurs in scenario 2, as all active connections are first reduced to their minimum quantity before a new request is rejected.

For the PSP-based simulations we observed the expected positive impact of allowing for multiple-access of the different stages on the dropping ratio. While with single access the dropping rate in cell 1 is 6.58%, it decreases to 3.36% and 1.62% in scenario 4 and 5, respectively. It is somewhat surprising that multiple-access to the highly congested network of provider 2 could further halve the dropping rate. This can be explained with the additional flexibility introduced by allowing more users multi-access.

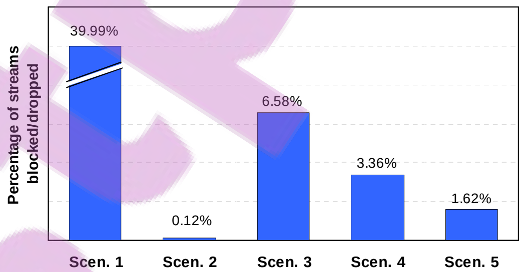
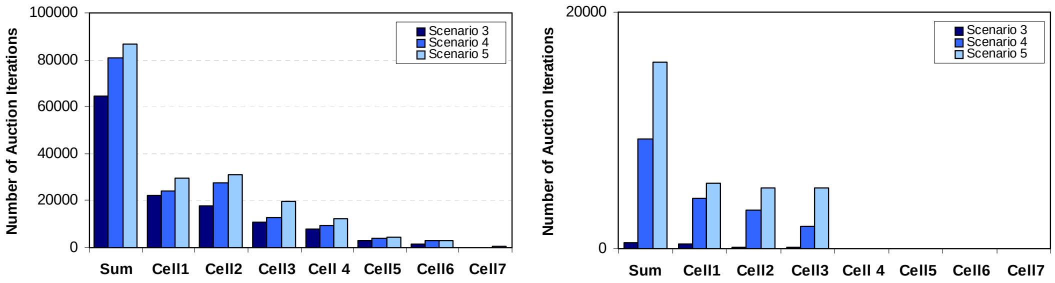



**Figure 4.62:** Mean video blocking/dropping rates in the five scenarios for cell 1 compared to overall connection requests (Provider 1).

Finally, we have analysed the total number of bidding steps over the simulation period (this can only be done for the scenarios 3 to 5, in which the PSP auction is used as allocation mechanism). As already shown in the previous experiments, the multiple-access option does increase the number of bidding iterations. This general circumstance does also apply to a setting in which users join and leave in a random fashion. Figure 4.63(a) shows the point estimator of the mean number of bidding steps for all providers for all cells. As can be observed the number of iterations increases with the multiple-access options (scenarios 4 and 5) but the increase is moderate, being about 15-25%.

In Figure 4.63(b) the number of bidding iterations are shown for provider 3. Because scenario 4 opens up this network to other non-subscribed users, a steep increase in bidding rounds can be observed.




(a) Number of iterations summed over all providers.

(b) Number of iterations for provider 1.

**Figure 4.63:** Number of iterations for each PSP scenario.

## ▶ 4.6.8 Conclusions

The development of a reasonably complex simulation scenario for resource allocation using the PSP auction mechanism and the use of the *BalancedBid* strategy in the multiple-access scenario have deepened our understanding of the practical impact of a dynamic pricing system for wireless network resources. While many aspects of the overall scenario have been kept simple and many assumptions needed to be made, we can derive some overall conclusions from the simulation results.

First, the PSP auction has shown to be feasible to be used in a multi-auctioneer setting with a large number of agents, of which some have access to multiple auctions. We could also show that the use of the *BalancedBid* strategy increases the overhead from bidding messages by only about 15-25%, which might be insignificant if the allocated flow sizes are large and long in nature.

Second, multiple-access increases consumer surplus by allowing the balancing of resources across auctions. In the specific case of this experimental series the overall increase in surplus from single-access to a two-provider multiple-access scenario was about 35% but very much depends on the specific loads in the different cells and the differences in consumer utility.

Third, the effects of multiple-access on provider revenue are small. While the specific increase in revenue depends on the definition of $\epsilon$, as it determines the maximum range of congestion-based revenues, the specific contribution of the congestion fees is arbitrary and cannot be seen as a reliable additional source of revenue for the provider. This is directly understandable as the PSP has been designed as a congestion avoidance mechanism, which minimises congestion by incentivising users to resubmit truthful bids depending on the market situation.

Subsequently, the only stable revenue source are usage-based charges from the static reserve price. As providers are usually unaware of the distribution of the users' utilities it is not possible to set a price to maximise revenue under the given objective to maximise consumer surplus from resource allocation.

A somewhat surprising outcome of this research was the good performance of the simple central allocation mechanism, which was able to proportionally reduce flow sizes of active streams if new requests arrive (Scenario 2). It was unexpected that this scenario would perform equally well compared to the distributed model by using the PSP auction.

## ▶ 4.7 Chapter Summary

We have presented a decentralised, flow-based resource allocation scheme based on pricing. The PSP auction format has been chosen as allocation mechanism because of its properties such as efficiency and incentive compatibility. The main extension of this work has been based on the idea of allowing players to simultaneously bid on multiple auctions, each independently offering network resources, to multiply the chances of increasing the obtained resource share as the sum of all sub-flows.

This design reflects the situation in a competitive wireless access network setting, in which providers allow mobile terminals to dynamically request resources without predefined contractual relationships. Users in such a market are free to bundle resources from multiple providers to increase the bandwidth.

The *BalancedBid* bidding strategy resembles truthful bidding in a multi-auction environment. By using it a bidder maximises his utility gained from the sum of resources by simultaneously truthfully expressing his demand on the market. While each auctioneer will only receive a sub-bid of this demand, the summed demand corresponds to the user's demand.

We have analytically examined that *BalancedBid* is the strategy which maximises a bidders utility independently of how other bidders behave. We have also shown that if used by all bidders a Nash equilibrium is reached. The equilibrium allocation maximises social welfare.

In addition to the *BalancedBid* strategy we have presented several alternative bidding strategies for cases in which it is impossible for bidders to coordinate their demand or bidders are unable to bundle resources from more than one auction. We have experimentally explored the properties of such bidding strategies compared to *BalancedBid*. None of the alternative strategies is able to combine all properties; either they do not converge in finite time or the resulting allocation results in inefficiencies.

Much of the intuition behind the bidding strategies has been gathered by the extensive use of the agent-based simulation model, which has been developed during the research period in parallel to the theoretical formulation of the bidding strategies. Many different variations of bidding strategies have been tested to understand the consequences on efficiency, convergence behaviour and revenue generation. For example, the *BalancedBid* strategy has been the result of extensive experimentation with other bidding strategies. Even if the bidding behaviour to a single auction did not give the impression of a converging algorithm, the cumulated results indicated convergence.

An extensive simulation study has revealed further qualities of *BalancedBid* in a systematic way. We could show the characteristics of efficiency improvements with access to multiple auctions with changing user valuation and different numbers of bidders in each auction. We showed that even if not all bidders gain access to all auctions an efficient equilibrium is reached. We also tested the convergence behaviour in several settings in which bidder access created circular relationships, which showed no significant effect on the convergence time.

A second line of interest was on the difference in revenue when bidders gain access to multiple auctions. While the congestion-based charges in equilibrium are minimal and not predictable with small $\epsilon$, we gradually increased the reserve price at one or more auctions and recorded the overall revenue. We can observe that with the possibility of access to multiple auctions revenue decreases since bidders have options to gain resources from other auctions with lower reserve prices. To set an optimal reserve price an auctioneer not only needs to know the distribution of utility but also how many other auctioneers compete for bidders.

When looking further into convergence we also conclude that, compared to the single auction case, the possibility of accessing multiple auctions increases the convergence steps to equilibrium with an increasing number of active bidders. Since multiple sub-bids per round are submitted the communication overhead increases with the number of available auctions. However, introducing multi-access does not increase the convergence time *per se* but only the communication overhead from bidding.

We also conducted experiments to understand the implications of new players joining or existing bidders leaving an already stable market. The consequences of an agent leaving the market were more severe; the average number of bidding steps to reach equilibrium was almost double compared to the case of a new agent joining a market. In both cases the more bidders were present in the platform, the less influence the disturbance of the equilibrium had on the convergence process.

One series of simulation experiments aimed at creating a multi-cell example for resource allocation under the presence of three competing providers and different user densities requesting two types of services. We have compared the allocation performance of PSP with centralised allocation models. A central allocation scheme, which proportionally reduces flow sizes if new bidders requested to join, reached nearly equally efficient average allocations compared with the single auction scenario. With the introduction of multi-auction access, efficiency of the allocation could be further increased with only minor improvements in overall provider revenue. The central allocation scheme performed best with regard to the service dropping ratio because it decreased allocations independent from user valuation.

# Chapter 5

# Admission-based pricing in a competitive wireless bandwidth-on-demand market

## ▶ 5.1 Introduction

In this chapter we look at dynamic pricing in wireless networks on the admission level. Prices are formed at the time of request and stay valid over the entire duration of the connection. Since pricing on the admission level is isolated from network control, quality guarantees can be given for the entire duration of the connection (Wang, 2006). This makes the proposed pricing scheme suitable for applications requiring static quality commitments with regard to minimum bandwidth or maximum packet loss.

As in the previous chapter, our main extension to the many existing models comes from the assumption that multiple, competing provider organisations offer wireless network resources in the market place. However, we now look at a free market setting, in which wireless resources are sold as a private good. Instead of maximising the efficiency of the allocation, in this chapter, we see the provider in the centre of the analysis. Consequently, the main objective for the design of the pricing model becomes maximising revenue in a competitive multi-provider setting. We see this work as an important continuation of the work conducted in the previous chapter. While resources in Chapter 4 where sold on a flow level, we are interested in setting a price on the admission level. While resource allocation was granular in the previous chapter, we now look at the case when resource allocation stays fixed for the duration of the connection to a wireless network.

In the assumed setting wireless resources are sold on-demand instead of requiring long-term contractual agreements. Customers requesting resources are able to compare the offers of multiple providers and to select the network with the highest net utility. The customer's decision is based on a one-dimensional description of his valuation, which he uses to determine if the resulting usage of the offered services results in a positive net utility.

Unlike with flow-based pricing, in which customers stay in continuous competition for resources, in the new setting pricing decisions need to be made at the time of the service request and cannot be changed afterwards. Since customers arrive in an asynchronous fashion at the network, providers are unable to use decentralised mechanisms, such as auctions, to let customers compete for resources in the market.[1] This also limits the possibility of a provider incentivising users to reveal their true valuation about the offered resources. Because of such limitations we design a setting in which providers set prices centrally, based on the information they have about the demand structure in the market and the competitive situation.

In this setting, the main objective becomes identifying suitable strategies for a wireless network provider facing direct competition in an on-demand market for wireless resources, given the available decision variables and the constraints imposed by the underlying network technology. To develop the main concepts we abstract from the real complexity in a mobile cell-based network and concentrate on a setting in which each provider operates a single cell. While this setting may be artificial in some sense, it allows us to understand the influence of competition in the price setting process and reveals properties which have been previously unknown.

The presented approach takes account of the increasingly competitive structure in wireless networks, where multiple providers compete for new customers more dynamically than with the traditional business models of long-term subscription. When more than one provider offers network resources in an area of overlapping wireless cells, the pricing strategy of each provider needs to consider the actions of the other providers in the market. Thus, the situation can be modelled as a non-cooperative game, in which providers become the players and base their decisions on the information they have about the market and their opponents. An important characteristic of this game is that some customers may be served by only one provider while other customers have multiple options to obtain resources. Since providers cannot distinguish between the two customer groups they need to find a strategy to "balance" between these two groups in order to maximise their revenue depending on the number of customers in each group. Sometimes it may be of advantage to serve only customers not connected to another wireless network and

[1] In principle, this would be possible if providers were able to bundle customer arrivals by waiting until the required minimum of requests has been received. However, this would lead to delays and would create problems in how to let customers compare offers from multiple providers.

to increase prices as in a monopolistic situation. In other cases such customers may be to few and prices need to be lowered to also attract customers connected with more than one network.

The chapter is organised as follows. Section 5.2 presents the general revenue maximisation problem of a single provider, given the resource constraints imposed by the implemented wireless technology. We define the problem as an optimal control model and discuss the principle problems of finding an explicit solution. We then turn our attention to a time-stationary model and show several solution approaches of how to determine the optimal decision variables, price and cell radius, under different assumptions of the active resource constraints. We also show the limitations in finding explicit solutions of the constrained maximisation problem. In Section 5.3, we develop a competition model using non-cooperative game theory. Two models are discussed: the game under complete information, and the game of incomplete information, in which providers have only partial information about the opponents cell setup. Section 5.4 presents an approximation framework for finding near-optimal solutions of optimal price/cell-radius combination for the two-provider case. This section also introduces the simulation approach and explains the main assumptions taken for the experiments. Section 5.5 provides the simulation results of the experiments. Section 5.6 concludes the chapter and summarises the research findings.

## ▶ 5.2 Revenue Maximisation in a Monopolistic Market for Wireless Resources

In this section we discuss the principle optimisation problem of a single provider which sells bandwidth with certain QoS guarantees in an on-demand market at which prices are defined at admission time of a new request. The developed model design has been inspired by the work described in Wang et al. (1997) and Wang (2006), in which an optimal control model for providing Quality-of-Service in a fixed network is developed and which has been briefly described in the literature review.

After describing our main modelling assumptions we present the general optimal control problem a provider in a monopolistic situation faces. We then turn our attention to a time-stationary model in which the service activation rate is constant over time and which allows us to further examine the model.

### ▶ 5.2.1 Main model assumptions and system parameters

We develop an analytical model for a single service class. We assume that customers can request only one constant-bit-rate (CBR) service with bit rate $c$ and a maximum bit-error-ratio (BER). We model the Quality-of-Service for a service class only by the maximum loss given by a maximum BER.[2] Depending on the coding model the BER directly defines the required Signal-to-Interference-to-Noise Ratio (SINR) and the power needed to supply the customer at a given distance from the base station.

Wireless services are provided by a single cell using WCDMA-based technology with the available code slots $C_{max}$, aggregated forward link transmission power $P_{max}$ and a maximum cell radius $Z_{max}$. We assume that the capacity in the cell is limited only by the forward link since multimedia applications usually require more resources in the downlink than in the uplink. We assume that the provider can decide on the maximum active coverage area within which it serves users' requests. The selected cell size[3] is described by a circular area with radius $z(t)$.

Providers do not know the exact position of users in their cell but only the distance between the base station and the mobile terminal. We assume that customers are spatially distributed over the service area and the distribution is a uniform random variable in the two-dimensional space. Service activation is modelled as a Poisson process and the service activation rate, $\lambda(x(t), z(t), t)$ is time-dependent and changes with the price $x(t)$ and the cell radius $z(t)$.

Without making additional assumptions about the functional form of $\lambda$ we can state that

$$\frac{\partial \lambda(x(t), \cdot, t)}{\partial x} \leq 0,$$

and

$$\frac{\partial \lambda(\cdot, z(t), t)}{\partial z} \geq 0.$$

This is because, with an increasing price, we can expect equal or less customers activating services with a valuation higher than $x(t)$ while the number of users increases with an increasing active cell radius $z(t)$.

The service duration for each customer is assumed to be exponentially distributed with mean $1/r$ and to be independent from the price $x(t)$. We further assume that we can express the two resource constraints, rate and power, as functions $C(x(t), z(t))$ and $P(x(t), z(t))$, respectively. Table 5.1 describes the main model variables used in the following sections.

### ▸ 5.2.2 The optimal control model for revenue maximisation

With the above assumptions, we can formulate the optimal control model for the interval $[0, T]$ as `MODEL1`

$$\max_{x(t), z(t)} \int_0^T \frac{\lambda\left(x(t), z(t), t\right)}{r} x(t) dt \tag{5.1}$$

---

[2]The taken approach has been extensively used in the existing literature. For example, see Siris (2002); Zhou et al. (2002, 2003).

[3]The maximum cell size is usually defined at the design time of the network but can be changed during operations in order to maximise overall network capacities in different overlapping cells. The method is called *cell breathing*, a mechanism that attempts to keep the forward and reverse link handoff boundaries balanced by changing the forward link coverage according to the changes in the reverse link interference level. We use the possibility of artificially shrinking the coverage area as a second decision variable for a provider to maximise revenue.

| Variable Name | Description |
|---|---|
| $c$ | Service rate defined for the service class |
| $p$ | Power needed to supply a user with a service rate $c$ and a maximum BER as defined by the service class |
| $z(t)$ | Maximum cell radius within which customers are served at time $t$ |
| $x(t)$ | Price set by a provider at time $t$ |
| $\lambda(x(t), z(t), t)$ | Service activation rate at time $t$ of users with a willingness-to-pay of at least $x(t)$ and a maximum radius $z(t)$ from the base station |
| $1/r$ | Service departure rate |
| $C_{max}$ | Maximum rate of the cell |
| $P_{max}$ | Maximum transmission power of the cell |
| $Z_{max}$ | Maximum radius of the cell |

**Table 5.1:** Overview of system variables

subject to

$$C\left(x(t), z(t)\right) \leq C_{max}, \tag{5.2}$$

$$P\left(x(t), z(t)\right) \leq P_{max}, \tag{5.3}$$

$$z \leq Z_{max}. \tag{5.4}$$

In general, the solution to this model consists of a set

$$A(t) = \{(x(t), z(t)) : t \in [0, T], \operatorname{argmax}[\texttt{MODEL1}]\},$$

which defines the revenue maximising pairs consisting of a price and maximum radius that solve `MODEL1`. While it is not possible to provide an explicit solution without knowing the functional form of $\lambda\left(x(t), z(t), t\right)$ and without making additional assumptions on the modelling of the resource constraint functions $C\left(x(t), z(t)\right)$ and $P\left(x(t), z(t)\right)$, we can provide some general observations about the solution space if none of the resource constraints apply or only one constraint applies.

### Unconstrained system

If `MODEL1` is not limited by any of the two constraints (5.2) and (5.3), i.e., the number of requests in the service area is insufficient to fully utilise the capacities of the wireless cell at any price and cell radius, the model can be converted to an unconstrained maximisation problem, which can then be easily solved. To maximise its active user base, and, with it, the number of service activations, a provider sets its coverage to the maximum cell size and searches for the service price maximising its revenue with

$$z^*(t) = Z_{max}; \, x^*(t) = \frac{\lambda}{\frac{\partial \lambda}{\partial x}},$$

where $x^*(t)$ has been derived by solving the equation (5.1) within the integral. The maximum revenue can be obtained at the point where the price elasticity of demand $\epsilon$

equals 1, where $\epsilon = -\frac{\partial \lambda}{\partial x}\frac{x}{\lambda}$. As before, the demand is expressed by the service activation rate $\lambda_z(x)$.

### Rate-constrained system

If the system is constrained only in the available code slots for data transmission, i.e., if users request high-bandwidth services with a large maximum bit-error-ratio, the available transmission power is sufficient to serve all users at the maximum cellsize $Z_{max}$. Therefore, the provider will announce a price so that $C(x^*(t), z^*(t)) = C_{max}$ and $z^*(t) = Z_{max}$. Under the realistic assumption of a linear relationship of the rate constraint in the form $C\left(x(t), Z_{max}\right) = \frac{\lambda(x(t), Z_{max})}{r} c$, we can solve

$$z^*(t) = Z_{max};\, x^*(t) = \frac{\epsilon}{1-\epsilon}\gamma c,$$

where $\gamma > 0$ is the Lagrangian multiplier of the constraint given in Equation (5.2).

### Power-constrained system

If the system is limited by the available transmission power, i.e., many users request low-bandwidth services at a low maximum BER, the question about the optimal cell size becomes relevant to the solution, since the power requirements grow exponentially with the distance of the terminal from the base station. If a terminal far from the base station is admitted, it may consume as much power as it would require to serve several terminals nearby the base station at the same rate and service class. Loosely speaking, a base station must balance its active customer base, given by the cell radius $z^*(t)$, with the available transmission power. If it chooses too a large cell radius, it risks binding the available transmission power to terminals far away from the base station, and is unable to serve the projected number of services to reach the projected revenue with the given $(x^*(t), z^*(t))$ combination. If it chooses $z^*(t)$ as too small, the available customer base willing to pay a price of $x^*$ is too small and network resources will be underutilised.

For the case of a power-constrained network cell, `MODEL1` can be partially solved explicitly by formulating a suitable $P(x(t), Z_{max})$ function. We provide an example of the resulting system in Section 5.2.3. However, in most cases, the solution of the resulting non-linear constrained maximisation problem can only be derived numerically.

### Rate and power-constrained system

If both power- and rate-constraints are active we cannot provide any general solution. The existence of an explicit solution depends on the functional form of $\lambda$ and both resource constraint functions $C(x(t), z(t))$ and $P(x(t), z(t))$.

### ► 5.2.3 Revenue maximisation under resource constraints for the time-stationary case

Though it would be ideal to solve `MODEL1` directly, it has been shown to be mathematically intractable (Wang et al., 1997). The difficulty comes from the fact that it is difficult to describe network states, such as the average number of active services, as functions of prices and time. We have therefore decided to look at a model in which the service activation rate $\lambda(x,z)$ is time-stationary. With this assumption we can transform the optimal control model into a constrained maximisation problem. Additionally, since customers are assumed to be uniformly distributed over the service area, we can rewrite $\lambda(x,z) = \frac{\pi z^2}{\pi Z_{max}^2}\lambda(x, Z_{max}) = \frac{z^2}{Z_{max}^2}\lambda_z(x)$. Thus, $\lambda$ becomes only dependent on the price, with $\lambda_z(x)$ defining the service activation rate in a circular area with radius $z$. We can state the constrained maximisation problem as `MODEL2`:

$$\max_{x,z} \frac{\lambda(x,z)}{r}x = \frac{z^2\lambda_z(x)}{rZ_{max}^2}x,$$

subject to

$$C(x,z) \leq C_{max},$$

$$P(x,z) \leq P_{max},$$

$$z \leq Z_{max}.$$

We can use the Lagrange Multiplier Method to solve the above system. The Lagrangian is given by

$$L(x,z,\gamma_1,\gamma_2,\gamma_3) = \frac{z^2\lambda_z(x)}{rZ_{max}^2}x - \gamma_1(C(x,z)) - \gamma_2(P(x,z)) - \gamma_3 z,$$

the first-order conditions

$$\frac{\partial L}{\partial x} = \frac{z^2\lambda_z'(x)}{rZ_{max}^2}x + \frac{z^2\lambda_z(x)}{rZ_{max}^2} - \gamma_1\left(\frac{\partial C(x,z)}{\partial x}\right) - \gamma_2\left(\frac{\partial P(x,z)}{\partial x}\right) = 0, \tag{5.5}$$

$$\frac{\partial L}{\partial z} = \frac{2z\lambda_z(x)}{rZ_{max}^2}x - \gamma_1\left(\frac{\partial C(x,z)}{\partial z}\right) - \gamma_2\left(\frac{\partial P(x,z)}{\partial z}\right) - \gamma_3 = 0, \tag{5.6}$$

and the slackness conditions

$$\gamma_1(C_{max} - C(x,z)) = 0, \gamma_1 \geq 0, \tag{5.7}$$

$$\gamma_2(P_{max} - P(x,z)) = 0, \gamma_2 \geq 0, \tag{5.8}$$

$$\gamma_3(Z_{max} - z) = 0, \gamma_3 \geq 0. \tag{5.9}$$

The objective is to identify the pair ($x^*$, $z^*$), which maximises revenue for the time-stationary $\lambda_z(x)$ function. In this system, the Lagrange multipliers serve as shadow prices for the amount of resources consumed by the average number of active services, under the assumption that users are uniformly distributed on the service area.

The solvability of the equation system given by (5.5) - (5.9) depends on the definition of the service activation function $\lambda_z(x)$ as well as on the complexity of the resource constraints. While the rate constraint is usually a linear equation, the transmission power consumed by a service can only be adequately described by a non-linear function which depends on the distance between base station and mobile terminal as well as other environmental factors. In the following we discuss some specific cases of the constrained maximisation problem.

### Revenue maximisation for a single base station with rate constraint

In this step we define the general revenue maximisation problem for a provider which is constrained in the number of code slots available for data transmission. We can express the constrained maximisation model as `MODEL3`:

$$R = \frac{\lambda_z(x)}{r} x$$

subject to

$$\frac{\lambda_z(x)}{r} c \leq C_{max}.$$

The constraint expresses that for this average number of users the summed rate of all service requests needs to be smaller than or equal to the available capacity of the network link. The rate constraint says that the maximum rate $c$ consumed by each admitted connection multiplied by the average number of active services must not be larger than the available code slots $C_{max}$. We can solve this system by using the Lagrange method. The Lagrangian function is defined as

$$L(x, \gamma) = \frac{\lambda_z(x)}{r} x - \gamma \frac{\lambda_z(x)}{r} c.$$

The first derivative $\partial L / \partial x$ of the Lagrangian is given by

$$x\lambda_z'(x) + \lambda_z(x) - \gamma c \lambda_z'(x) = 0,$$

and the slackness condition is defined as

$$\gamma(C_{max} - \frac{\lambda_z(x)}{r} c) = 0, \gamma \geq 0.$$

Solving the constrained maximisation model gives

$$x^* = \frac{\epsilon}{\epsilon + 1} \gamma c$$

as the optimal price, where $\gamma$ is the Lagrange multiplier to the capacity constraint and

$$\epsilon = \frac{x}{\lambda_z(x)} \frac{\partial \lambda_z(x)}{\partial x}$$

is the price elasticity of the demand expressed by the service activation rate $\lambda_z(x)$ depending on the price $x$.

**Proposition 5.1** (The optimal pricing policy). *Suppose $x^*$ is the optimal solution for the pricing model defined with `MODEL3`. Then*

$$\begin{array}{lll} \textit{Case 1:} & x^* = \frac{\epsilon}{\epsilon+1}\gamma c & \textit{if } \gamma > 0, \\ \textit{Case 2:} & x^* = x^0 & \textit{if } \gamma = 0, \end{array}$$

*where $x^0$ maximises $x\lambda_z(x)/r$ without considering the rate constraint.*

In this general solution two cases have been distinguished. In the first case the network is tightly constrained and demand for network resources exceeds the supply measured in the available code slots per time unit. In Case 1, the Lagrange multiplier $\gamma$ can be interpreted as a shadow price for using resources in the form of code slots.

If the network is under-utilised, i.e. the given demand is smaller than the available cells per time-unit, the provider sets the price without considering the capacity constraint and solves the maximisation problem according to Case 2. The shadow price becomes zero, and the provider maximises its revenue without considering the resource constraint.

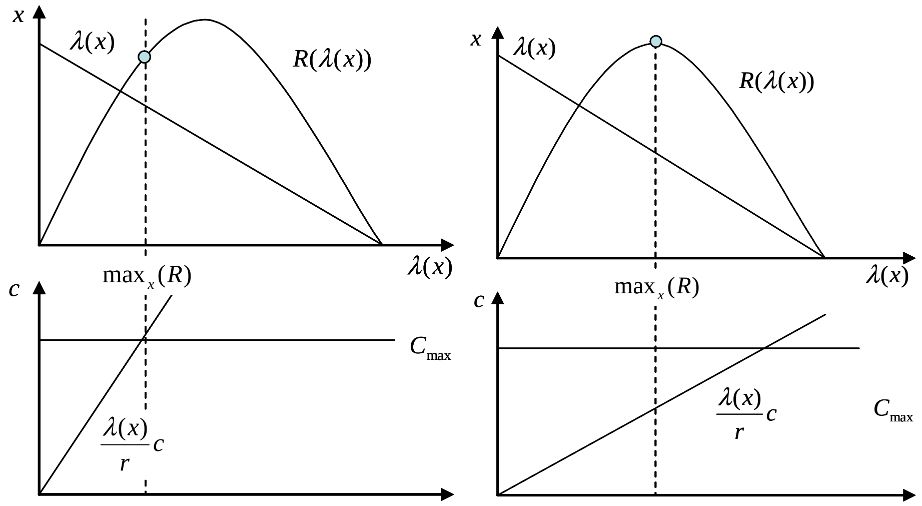
**Example 5.1** (Solution for a linear service activation rate $\lambda_z(x)$). *Proposition 5.1 provides the general solution to the constrained maximisation problem without making any assumptions on the functional form of $\lambda_z(x)$. In the following we assume that $\lambda_z(x)$ is a linear function of the form $\lambda_z(x) = e - fx$. Then, $\epsilon = \frac{fx}{fx-e}\gamma c$ and the constrained maximisation problem becomes a linear programming problem with two equations and two unknown variables $x$ and $\gamma$. Under the condition that all parameters are equal or larger than zero we can solve*

$$\begin{array}{lllll} \textit{Case 1:} & x = \frac{e}{f} + \frac{rC_{max}}{fc} & \textit{and} & \gamma = \frac{2}{c} - e & \textit{for } e < \frac{2}{c} \\ \textit{Case 2:} & x = \frac{e}{2f} & \textit{and} & \gamma = 0 & \textit{for } e \geq \frac{2}{c} \end{array}$$

*The condition ensures that in Case 1, $\gamma$ is always $> 0$. Otherwise, Case 2 applies and the system has more supply of resources than the demand expressed by $\lambda$. Figure 5.1 illustrates the two cases with the linear demand function. Figure 5.1(a) shows the first case, when the system is constrained in code slots. Figure 5.1(b) illustrates the case of a system in which supply exceeds demand.*

### Revenue maximisation for a base station with a rate constraint and two service classes

We now turn our attention to a scenario in which customers request resources from two different service classes. We express the demand in both service classes by $\lambda_{z1}(x_1)$ and




(a) A rate-constrained setup.

(b) A setup in which supply exceeds demand.

**Figure 5.1:** Illustration of the two solution cases in the constrained maximisation problem.

$\lambda_{z2}(x_2)$, respectively. The constrained maximisation problem can be written as `MODEL4`:

$$\max \frac{\lambda_{z1}(x_1)}{r_1}x_1 + \frac{\lambda_{z2}(x_2)}{r_2}x_2,$$

subject to

$$\frac{\lambda_{z1}(x_1)}{r_1}c_1 + \frac{\lambda_{z2}(x_2)}{r_2}c_2 \leq C_{max}.$$

We derive the Lagrangian with

$$L(x_1, x_2, \gamma) = \frac{\lambda_{z1}(x_1)}{r_1}x_1 + \frac{\lambda_{z2}(x_2)}{r_2}x_2 - \gamma(\frac{\lambda_{z1}(x_1)}{r_1}c_1 + \frac{\lambda_{z2}(x_2)}{r_2}c_2),$$

the two first-order conditions

$$\frac{\partial L}{\partial x_1} = \lambda_{z1}(x_1) + \lambda'_{z1}(x_1)x_1 - \gamma c_1 \lambda'_{z1}(x_1) = 0,$$

$$\frac{\partial L}{\partial x_2} = \lambda_{z2}(x_2) + \lambda'_{z2}(x_2)x_2 - \gamma c_2 \lambda'_{z2}(x_2) = 0,$$

and the slackness condition

$$\gamma(C_{max} - \frac{\lambda_{z1}(x_1)}{r_1}c_1 + \frac{\lambda_{z2}(x_2)}{r_2}c_2) = 0, \gamma \geq 0.$$

The general solution can again be written by using the elasticity of the service arrival rate for each service class.

**Proposition 5.2** (The optimal pricing policy)**.** *Suppose $x_1^*$ and $x_2^*$ are the optimal solutions for* *the pricing model defined with* `MODEL4`*. Then*

$$x_1^* = \frac{\epsilon_1}{\epsilon_1 + 1}\gamma c_1,$$

$$x_2^* = \frac{\epsilon_2}{\epsilon_2 + 1}\gamma c_2,$$

*for $\gamma > 0$, and*

$$x_1^* = x_1^0 \text{ and } x_2^* = x_2^0,$$

*for $\gamma = 0$, and $x_1^0$, $x_2^0$ being the solutions of the unconstrained maximisation problem. As in the one service-class case, $\epsilon_i$ expresses the price elasticity of demand in the service class i and is defined as:*

$$\epsilon_i = \frac{x_i}{\lambda_{zi}(x_i)}\frac{\partial\lambda_{zi}(x_i)}{\partial x_i}$$

The model for more than two service classes can be developed accordingly.

**Example 5.2** (Solution for linear service activation function $\lambda_i(x_i)$)**.** *We use the linear functions $\lambda_1(x_1) = e_1 - f_1x_1$ and $\lambda_2(x_2) = e_2 - f_2x_2$. We can solve*

$$x_1^* = \frac{e_1}{2f_1} + \frac{c_1\left(e_2c_2r_1 - 2C_{max}r_2r_1 + e_1c_1r_2\right)}{2\left(f_1r_2c_1^2 + f_2c_2^2r_1\right)},$$

*and*

$$x_2^* = \frac{e_2}{2f_2} + \frac{c_2\left(e_2c_2r_1 - 2C_{max}r_2r_1 + e_1c_1r_2\right)}{2\left(f_1r_2c_1^2 + f_2c_2^2r_1\right)},$$

*for $e_2c_2r_1 + e_1c_1r_2 > 2C_{max}r_2r_1$ (Case 1) or*

$$x_1^* = \frac{e_1}{2f_1} \text{ and } x_2^* = \frac{e_2}{2f_2},$$

*for $e_2c_2r_1 + e_1c_1r_2 \leq 2C_{max}r_2r_1$ (Case 2).*

### Revenue maximisation for a single base station with rate and power constraints

To model the power constraint we use the model presented in Siris (2002), which presents a framework for resource control in WCDMA networks, providing Quality-of-Service in terms of the maximum average BER. In our work we only consider the forward link. Table 5.2 defines the variables and parameters used in the subsequent models.

In the downlink, the interference $I_i$ experienced by a mobile terminal $i$ is given by the orthogonality $\theta_i$ of the spreading codes used by the mobile terminals and can be expressed as $I_i = \theta_i g_i \sum_{j\neq i} p_j$. Typical values for $\theta_i$ range between 0.1 and 0.6 (Holma and Toskala, 2000). With a $\theta_i = \theta$ we can then express the bit-energy-to-noise-density ratio $E_b/N_0$ as

$$\left(\frac{E_b}{N_0}\right)_i = \frac{W}{c_i}\frac{g_ip_i}{\theta g_i \sum_{j\neq i} p_j + \eta_i}. \tag{5.10}$$

| Parameter Name | Description |
|---|---|
| $W$ | Chip rate of the wireless system |
| $a_i$ | Average throughput of user $i$ with a transmission rate of $c_i$ and a target BER as defined in the service profile |
| $g_i$ | Channel gain between sender and receiver. We use a simple propagation model, which defines $g_i = k\hat{d}_i^{-u}$, where $\hat{d}_i$ is the distance between the base station and the user, $k = 1.8E - 14$, and $u = 4$ |
| $I_i$ | Power of the total interference experienced by user $i$ |
| $\gamma_i$ | Target signal-to-interference-to-noise (SINR) ratio for user $i$ |
| $\eta$ | Power of the background noise, identical for all users in the system |
| $\theta$ | Orthogonality factor of the codes assigned to the active users |

**Table 5.2:** Overview of system parameters

We assume that the background noise $\eta_i$ is the same $\eta$ for all users within the cell. The value of $\frac{E_b}{N_0}$ corresponds with the signal quality and, thus, directly influences the BER (Siris, 2002). In particular, the BER is a non-decreasing function of $\frac{E_b}{N_0}$, which depends on many factors such as the multipath characteristics, the used modulation techniques and the implemented forward-error-correction algorithms. Under perfect power control we can set $\gamma_i = \frac{E_b}{N_0}$, with $\gamma_i$ defining the target signal-to-interference-to-noise density ratio (SINR). For this study we assume a QPSK modulation. For a non-fading channel, the relationship between BER and SINR can then be expressed as (Rulnick and Bambos, 1997):

$$BER(\gamma_i) = 0.5 \text{ erfc } \sqrt{\gamma_i} \tag{5.11}$$

where *erfc* is the complementary error function. Additionally, we need to consider the effect of forward error correction for a given target BER. If we assume a continuous flow[4], and by using Shannon's Second Theorem, the maximum average throughput with a given sending rate $r_i$ equals $a_i = r_i \times F(BER(\gamma_i))$, with $1/F(BER)$ representing the minimum redundancy factor to recover the message without any error. $F(BER(\gamma_i))$ is given by (Maillé, 2004):

$$\begin{aligned} F(BER) &= 1 - H(BER) \\ &= 1 + BERlog_2(BER) + (1 - BER)log_2(1 - BER). \end{aligned} \tag{5.12}$$

With the above equations we can now relate a target BER level to the corresponding target SINR level to achieve the desired channel quality.

The downlink direction of the CDMA network is power-constrained and the corresponding resource constraint is:

$$\sum_i p_i \leq P_{max}.$$

If we assume that a base station always allocates all its power among active users we can rewrite Equation (5.10) as:

$$\gamma_i = \frac{W}{c_i} \frac{g_i p_i}{\theta_i g_i (P_{max} - p_i) + \eta_i}. \tag{5.13}$$

[4]The relationship becomes significantly more complex for packetised flows. For details see Siris (2002).

To derive the channel gain from the distance between the base station and the mobile station we use a simple attenuation model $g(d) = kd^{-u}$, where $k$ is a constant set to $k = 1.82E - 14$ and $u = 4$ (Parsons, 2000). In practice, the channel attenuation is not necessarily an increasing function of distance. This is due to shadowing and the multipath effects a signal experiences on the way from the sender to the receiver. However, we can interpret the distance $d$ as a "radio distance" rather than a physical distance (Zhou et al., 2002).

We now look for an expression, which gives the transmission power needed to supply a single mobile terminal in a given distance $d$. We therefore solve Equation (5.13) to $p_i$ and obtain:

$$p_i = \frac{c\gamma(\eta d_i^4 + k\theta P_{max})}{k(W + c\gamma\theta)}.$$

In the next step we want to determine the transmission power required to supply a circular service area with radius $z$ with a given user density $\lambda(x)$, if all users request services only in on service class (with identical target SINR $\gamma$ and sending rate $c$). Since we have previously assumed that customers are homogeneously distributed throughout the service area and with $\lambda(x, z)$ developed in the previous section we can derive the user density at a cell radius $z$ by differentiating

$$\frac{\partial\lambda(x, z)}{\partial z} = \frac{2\pi z\lambda_z(x)}{Z_{max} r}.$$

The total power needed to supply users up to the radius $z$ can be derived by multiplying the power with the user density and integrating over the radius. We obtain:

$$P(x, z) = \int_0^z \frac{2\pi\delta\lambda_z(x)}{Z_{max}^2 r} \frac{c\gamma(\delta^4\eta + k\theta P_{max})}{k(W + c\gamma\theta)} d\delta. \tag{5.14}$$

$$= \frac{\pi c\gamma z^2\lambda_z(x)(\eta z^4 + 3k\theta P_{max})}{3Z_{max} kr(W + c\gamma\theta)} \tag{5.15}$$

(5.16)

With this definition of the power constraint we can solve `MODEL2`. The derived system of equations is a non-linear programming problem with polynomial functions. Because of the high order of polynomials given by the power constraint, no explicit solution can be provided even if we assume a linear service activation function $\lambda_z(x)$. In Section 5.5.2 we compare the numerical results of `MODEL2` with results which have been derived by simulation and by using identical parameters.

## ► 5.3 A Game-Theoretic Discussion to the Two-Provider Case

In this section we discuss the situation when several competitive providers serve a market in which customers can select networks based on the price information they receive at the time of request. Unlike in the previous section, in which a provider needed to solve the constrained maximisation problem given by the demand structure and the resource constraints for the network cell, in the new setting a provider needs to take the actions of the other providers into account when making its pricing decision. Thus, non-cooperative game theory can be used to model the situation as a game played between providers.

In principle, many different settings can be thought of for describing a possible game played between the market participants. One could see a game in each interaction between the providers and the customer. In this type of game the customer can be interpreted as an auctioneer, which requests bids from providers. The provider submitting the lowest bid wins the auction and closes the contract with the customer (Mitchell and Vogelsang, 1991). Depending on the auction format, providers may be incentivised to reveal their true cost to the customer. Thus, the problem becomes identifying the underlying cost structure of the network, which may be difficult to determine (Courcoubetis and Weber, 2003).

Another alternative is to assume that players in the market have different market power and players choose their actions subsequently. In such a game one player becomes the leader, while the other players are the followers. The leader moves first by choosing its price, being aware that the followers can infer its action in their strategy. In a second step the followers announce their prices in the market.

We have decided to model the situation as a game, in which providers have symmetrical information about the opponent and choose their actions in the form of price announcements simultaneously. Since we have assumed the customer demand function (given by $\lambda_z(x)$) to be time-stationary, the game has to be played only once to set prices in the market and to let players maximise their average revenue obtained from selling network resources. However, it may be played repeatedly when the demand situation changes. In this case we assume that providers behave myopically by not using the information collected in the past rounds to infer the actions of the opponent players.

In the described setting the position and size of the players' wireless network cells become a critical element of the game. Since the cell overlap between the cells determines how many customers are "shared" between the players, and how many customers can be served monopolistically, the position of the opponents' base stations is important information in the game. In the case of a full overlap of the cells of two providers the described game becomes similar to the Bertrand game, in which duopoly firms compete over prices. In this game, in which two identical price-setting firms produce homogeneous products at constant marginal costs, marginal cost pricing is the unique Nash equilibrium (Tirole, 1988). In effect, each firm makes zero profit. An important difference to this game is the introduction of resource constraints, which limits players in lowering their prices below a certain level.

The situation looks different when cells do only partly overlap. In such a setting the customer base is divided into two groups: customers, which can choose between several offers from different providers and customers which have only one access option. The price is expected to reflect this situation in which the provider needs to balance between these two groups if it is allowed to set only one price for all customers.[5]

To simplify the description of the game, we limit our attention to a two-provider case, which both operate a single cell. We further assume that both providers use the same wireless technology and that they face identical technological constraints. In all such models the price is the only optimisation variable available to the providers. We first model the game of complete information in which players have full information about the competitive situation and then turn our attention to a situation in which the provider's location is only known by the provider itself but not by its opponents.

### ▶ 5.3.1 The situation as a game of complete information

In a game of complete information the payoff functions of the two players are common knowledge in the market (Gibbons, 1992). The payoff functions of the players are determined by the resource constraints of the wireless cell, and the "level of competition", which, in our model, is solely given by the percentage of cell overlap. Since we have assumed that the players use identical wireless technologies, and the maximum cell radii are identical, the distance between the base stations solely determines the competition level. We define $t_i$ as player $i$'s absolute geographical position. To simplify the analysis we assume that the position of both players only differs in one dimension. The revenue function $R_i$ of player $i$ is then given by:

$$R_i = x_i \frac{\lambda_z(x_i)}{r} \Big( \alpha(t_i, t_j) + (1 - \alpha(t_i, t_j))\beta(x_i, x_j) \Big), \tag{5.17}$$

where $\lambda_z(x_i)$, as in the previous section, is defined as the service activation rate in a circular area with radius $z$; $\alpha(t_i, t_j)$ is the percentage of the cell which is not covered by the opponent player given the positions $t_i$ and $t_j$. It therefore denotes the share of customer requests over which player $i$ may exert monopoly power and $1-\alpha$ is the share of customer requests who may decide for alternative network access. $\beta(x_i, x_j)$ describes if a player is able to gain customer requests in the overlapping area, which depends on its price and the opponent's price. We define $\beta$ as follows:

$$\beta_i(x_i, x_j) = \begin{cases} 1 & \text{if } x_i < x_j \\ \frac{1}{2} & \text{if } x_i = x_j \\ 0 & \text{if } x_i > x_j \end{cases} .$$

To derive a function for $\alpha$ we need to find an expression for the percentage of cell area owned monopolistically compared to the overall cell area $\pi z^2$. Figure 5.2 shows the known parameters in such a setting. The area of the "symmetric lens" is given by the circle segment $A_1 = \frac{2\arccos(t_2 - t_1)}{2z}$ minus the area of the triangle given by $A_2 = z(t_2 - t_1)\sin(\arccos(\frac{t_2 - t_1}{2z}))$. We can then divide $A_1 + A_2$ by $\pi z^2$ and, after some transformations,

[5] This also reflects the situation in which a provider cannot directly distinguish between the two customer groups and thus needs to set a price for all customers.

obtain:

$$\alpha(t_2, t_1) = 1 - \left(\frac{4z \arccos\left(\frac{t_2 - t_1}{2z}\right) - (t_2 - t_1)\sqrt{1 - \frac{(t_2 - t_1)^2}{4z^2}}}{2\pi z}\right). \tag{5.18}$$

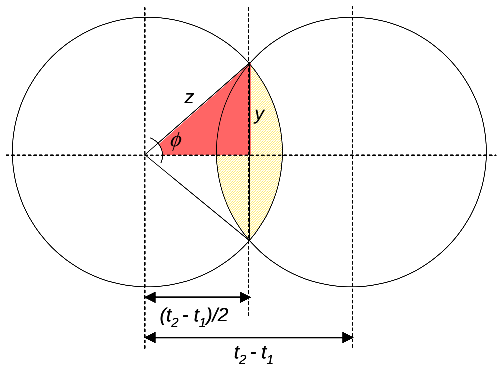



**Figure 5.2:** Calculation of the area of the symmetric lens.

With the revenue function given in Equation (5.17) we can proceed in defining the constrained maximisation problem to be solved by each player to derive the optimal price:

$$\max_{x_i} R_i = x_i \frac{\lambda_z(x_i)}{r} \left(\alpha + (1 - \alpha)\beta\right), \tag{5.19}$$

subject to

$$c \frac{\lambda_z(x_i)}{r} \left(\alpha + (1 - \alpha)\beta\right) \leq C_{max}. \tag{5.20}$$

As seen in Example 5.2.3, the above problem can be solved explicitly only for certain types of service activation functions $\lambda(x)$. To be able to continue our discussion, in the following, we assume a linear service activation function $\lambda(x) = e - fx$. We can now solve Equations (5.19) and (5.20) for the three possible values of $\beta$ and obtain two solutions each for $x_i$, one for the unconstrained case (given by $x_A$ in all three cases) and one for the constrained case ($x_B$ to $x_D$). The solutions are:

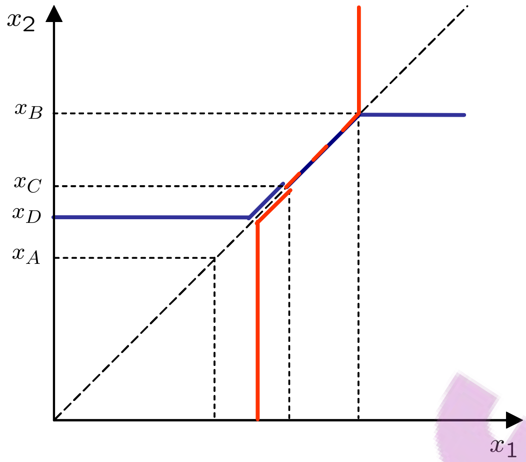
$$x_i = \begin{cases} x_A = \frac{e}{2f},\ x_B = \frac{e}{f} - \frac{rC_{max}}{cf} & \text{for } \beta = 0 \\ x_A = \frac{e}{2f},\ x_C = \frac{e}{f} - \frac{2rC_{max}}{cf(1+\alpha)} & \text{for } \beta = \frac{1}{2} \\ x_A = \frac{e}{2f},\ x_D = \frac{e}{f} - \frac{rC_{max}}{cf\alpha} & \text{for } \beta = 1 \end{cases}.$$

The above solutions give us the values for $x_i$ at which revenue is maximised, as well as the boundary prices at which the rate constraint starts to apply. Given the formal solution we now need to identify the reaction function of each player. The reaction function shows what strategy one player chooses, given the strategy of the other player (Gibbons, 1992). Given the opponent's price $x_j$, a player has the following options to "react":

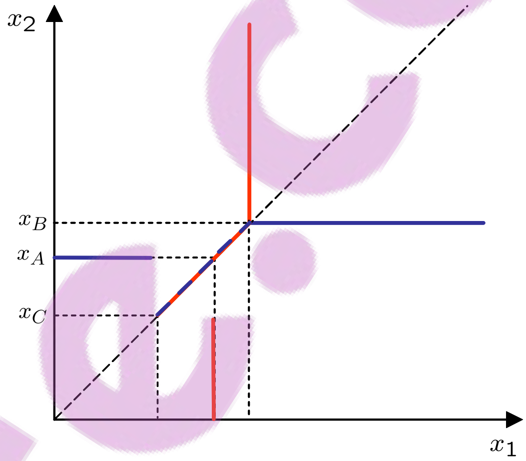
- Set his price $x_i$ equal with $x_j$,
- Set his price $x_i - \delta$ slightly below $x_j$,

- Set his price $x_i + \delta$ slightly above $x_j$,
- Set his price equal to the monopolistic price for the entire cell,
- Set his price equal to the monopolistic price for serving only customers belonging to the non-overlapping area $\alpha$.

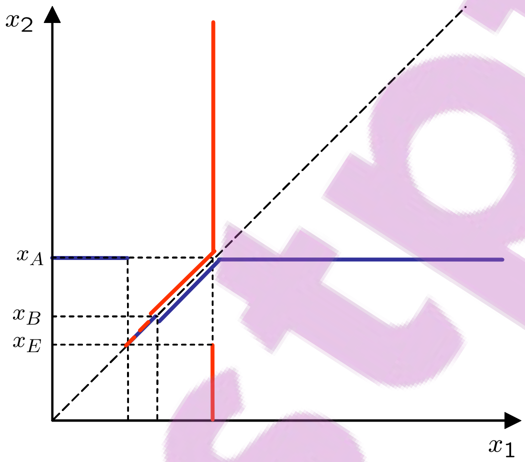
Depending on both, the given demand in form of the service activation rate $\lambda_z(x)$, and the rate constraint, we can distinguish four cases, which are shown in Figure 5.3.1. The cases can be described as follows:



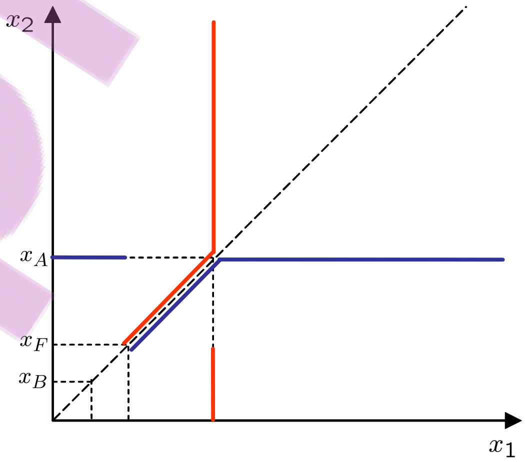

(a) Case 1: $x_B \geq x_C \geq x_A$.




(b) Case 2: $x_B \geq x_A \geq x_C$.




(c) Case 3: $x_A \geq x_B \geq x_E$.




(d) Case 4: $x_B \leq x_E \vee x_B < 0$.

**Figure 5.3:** Price reaction functions of the game of complete information.

**Case 1: $x_B \geq x_C \geq x_A$** The player always sets his price equal to $x_B$ if the opponent chooses a price higher than $x_B$. As long as the opponent price stays above $x_C$ the player sets its price equal to the opponent's price to equally share the customers in the overlapping area. With any price lower than $x_j$ the player violates the rate constraint, since all customer requests would be won by him and thus the available capacity would be insufficient.

With opponent prices between $x_C$ and $x_D$ the player cannot match this price since the resource constraint is violated. Instead, he chooses a price slightly above the opponent's price and only serves the customers in the non-overlapping part of the cell. With an opponent's price below $x_D$ the player chooses $x_D$, since this is the revenue maximising price for the customers located in the non-overlapping part of the cell. Case 1 is shown in Figure 5.3(a). The reaction function becomes:

$$x_i(x_j) = \begin{cases} x_B & \text{for } x_j > x_B \\ x_j & \text{for } x_C < x_j \leq x_B \\ x_j + \delta & \text{for } x_D < x_j \leq x_C \\ x_D & \text{for } x_j \leq x_D \end{cases} .$$

**Case 2:** $x_B \geq x_A \geq x_C$ The player always sets his price equal to $x_B$ if the opponent chooses a price higher than $x_B$. He matches the opponent's price between $x_B$ and $x_C$. With an opponent's price below $x_C$ he sets his price back to $x_A$ since it is more beneficial to charge to monopolistic price for customer requests outside the overlapping area. Case 2 is shown in Figure 5.3(b). The reaction function becomes:

$$x_i(x_j) = \begin{cases} x_B & \text{for } x_j > x_B \\ x_j & \text{for } x_C < x_j \leq x_B \\ x_A & \text{for } x_j \leq x_C \end{cases} .$$

**Case 3:** $x_A \geq x_B \geq x_E$ Since $x_B < x_A$ the player sets his price to $x_A$ whenever the opponent's price is above $x_A$. For opponent prices between $x_A$ and $x_B$ the player sets his price slightly below the opponent's price to gain all customer requests. For prices below $x_B$ this is not possible since the resource constraint applies and the chosen price is infeasible. Instead, the player now matches the opponent's price to equally share customers. With an opponent's price below $x_C$ it becomes better for the player to set his price back to $x_A$ and to serve only the customers in the non-overlapping area. To find $x_E$ we set

$$x_E \frac{e - x_E f}{r}(\alpha + (1 - \alpha)) = x_D \frac{e - x_D f}{r} \alpha,$$

and set $x_D = \frac{e}{2f}$, which is the revenue-maximising price of the unconstrained system. We solve the equation to $x_E$ and obtain $x_E = \frac{e}{2f} - \frac{\sqrt{e^2 f^2 (1-\alpha^2)}}{2f^2(1+\alpha)}$. Case 3 is shown in Figure 5.3(c). The reaction function becomes:

$$x_i(x_j) = \begin{cases} x_A & \text{for } x_j > x_A \\ x_j - \delta & \text{for } x_B < x_j \leq x_A \\ x_j & \text{for } x_E < x_j \leq x_B \\ x_A & \text{for } x_j \leq x_B \end{cases} .$$

**Case 4:** $x_B \leq x_E \vee x_B < 0$ The player chooses $x_A$ for an opponent's price above $x_A$. With opponent prices below $x_A$ but above $x_F$ he chooses a price slightly below to gain all customers. With prices below $x_F$ it becomes more beneficial to set the price to $x_A$ and to only serve customers outside the overlapping area. $x_F$ has been derived by setting

$$x_F \frac{e - x_F f}{r}(\alpha + \frac{1}{2}(1-\alpha)) = x_D \frac{e - x_D f}{r}\alpha,$$

setting $x_D = \frac{e}{2f}$ and solving to $x_F$. We obtain $x_F = \frac{e}{2f} - \frac{\sqrt{e^2 f^2 (1-\alpha)}}{2f^2}$. Case 4 is shown in Figure 5.3(d). The reaction function becomes:

$$x_i(x_j) = \begin{cases} x_A & \text{for } x_j > x_A \\ x_j - \delta & \text{for } x_F < x_j \leq x_A \\ x_A & \text{for } x_j \leq x_F \end{cases}.$$

Now that we know the price reaction functions of the players we can proceed in identifying candidate points for a Nash equilibrium, from which both players do not want to unilaterally deviate. From the reaction functions shown in Figure 5.3.1 (a)-(d) we can identify the candidate points for possible Nash equilibria, which need to be located at points where the price reaction functions overlap. We discuss each case separately.

**Case 1:** We can directly see that all prices above $x_B$ are dominated by $x_B$. This is because when setting a price above $x_B$, all customers in the overlapping area are served by the opponent. Thus, the revenue with a price above the monopolistic price must become smaller. We can also see that all prices below $x_C$ are dominated by $x_C$. Since below $x_C$ a player is not able to serve half of the customer requests in the overlapping area it needs to set a price so that it only serves customers outside the overlapping area. Therefore, the obtained revenue must be smaller than at point $x_C$. This leaves us with the area between $x_B$ and $x_C$. With any opponent price $x_B \geq x_j \geq x_C$ it is optimal for the player to set $x_i = x_j$. With any price higher he loses the customer requests in the overlapping area and obtains lower revenue. Furthermore, the rate constraint does not allow him to set a price lower than $x_j$ since he would win all customers in the overlapping area. Thus, all prices between $x_B$ and $x_C$ are possible Nash equilibria of the game.

The next question is which price a player would chose when he is aware of the opponent's payoff structure. Since we have already solved the constrained-maximisation problem to $x_C$, we know that this point maximises the player's revenue. Thus, a player would set his price $x_i = x_C$ at which all resources in both wireless cells get fully utilised.

Figure 5.4 plots the $x_C$ against $t_1$ and $t_2$ for a numerical example, which resembles the situation of case 1. We can see how $x_C$ decreases with a decreasing distance between both base stations. For all situations in which the distance of the cells is larger than $|t_i - t_j| > 2z$, the price is set to the monopolistic price $x_B = 70$.

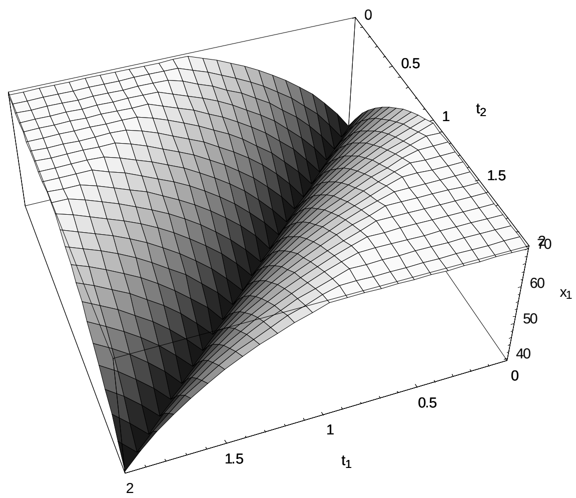



**Figure 5.4:** Plot of the revenue-maximising price $x_C$ against the position $t_i$ of player 1 and 2 with cell radius $z = 1/2$.

**Case 2:** As in case 1 all prices above $x_B$ are dominated by $x_B$. For all prices below $x_C$ the player loses all customers in the overlapping area, which results in a smaller average revenue. As in Case 1 all prices between $x_B$ and $x_C$ are Nash equilibria, since with a given opponent's price $x_j$, a player cannot do better than setting $x_i = x_j$. Any price above $x_j$ lets him lose half of the customer requests in the overlapping area. Any price below is infeasible because of the rate constraint. Again, we can identify a price $x_i$, which maximises a player's revenue under the assumption of full knowledge of the opponent's strategy, yielding $x_A$ as the optimal solution. All prices above and below give a player less revenue.

**Case 3:** All prices above $x_A$ are dominated by $x_A$. We can also see that prices between $x_A$ and $x_B$ are dominated by $x_A$ since a player would gain more revenue by charging the revenue-maximising price by charging any price below this value. Prices below $x_E$ are dominated by $x_E$. This leaves us with prices between $x_B$ and $x_E$ as Nash equilibria, from which no player wants to unilaterally deviate. By solving the constrained-maximisation problem we obtain $x_B$ as the revenue-maximising price of all possible Nash equilibria.

**Case 4:** The reaction functions of the players do not cross for any price $x_i$. Thus, no Nash equilibrium in pure strategies exists.

This definition of the prices forming a Nash equilibrium and the identification of the price, which denotes the revenue-maximising Nash equilibrium, concludes our analysis of the game of complete information.

### ▶ 5.3.2 The situation as a game of incomplete information

We now turn our attention to a game, in which part of the payoff function of each player is private information. As in the case of complete information, we assume that both players have identical resource constraints and that they operate a cell with identical cell radius $z$. All this is common knowledge to both players. However, we now assume that the position $t_i$ of the base station of player $i$ is private information to this player. Since a player's position $t_i$ is the only information a player holds privately, we denote $t_i$ as the *player's type*. A player's type also determines his payoff function when different positions of a wireless cells hold different probabilities of cell overlap with the opponent. For example, if some type allows a player to know for sure that no overlap can exist, his payoff will be different than in the case that he knows that overlap occurs with a certain probability.

With the uncertainty about the opponent's type introduced in this game, a player needs to form beliefs $p_i(t_j|t_i)$ about the opponent's type, given his own type $t_i$. This belief allows him to define a pricing strategy $x_i(t_i)$, which determines his action $a_i \in A_i$. Since in our case the action is defined as the price announced on the market, the action space becomes $A_i = [0, \infty)$. The beliefs of a player describe the uncertainty of that player about the types of the other player. Since the positions of the base station are assumed to be independent, the probability function simply becomes $p_i(t_j)$.

To form this belief function we additionally assume that the type spaces $T_1$ and $T_2$ are common knowledge and that types are drawn from the type spaces by a uniformly distributed random variable, which is also known to both players. With these additional assumptions $p_i(t_j)$ becomes a uniformly distributed random variable on the space $T_2$.

As with the game of complete information we can formulate the constrained maximisation problem with a rate constraint as:

$$\max_{x_i} \mathbb{E}_i[R_i(x_i)] = x_i \frac{\lambda(x_i)}{r} \int_{t_2} \Big[\alpha + (1-\alpha)\,\mathrm{Prob}\{x_i < x_j\}\Big]\, dt_2 \tag{5.21}$$

subject to

$$c\frac{\lambda(x_i}{r} \int_{t_2} [(\alpha) + (1-\alpha)\,\mathrm{Prob}\{x_i < x_j\}] dt_2 \leq C_{max}. \tag{5.22}$$

Equation (5.21) defines the expected revenue of player $i$, which distinguishes between two types of customers: customers located outside the overlapping area, denoted by $\alpha$, always accept the offer. Customers within the overlapping area only accept the offer if the price set by $i$ is lower than that of $j$, which occurs with $\mathrm{Prob}\{x_i < x_j\}$.

Equation (5.22) defines the rate constraint of the wireless cell. Note that unlike the one-provider case, the resource constraint contains the factor distinguishing between the two customer groups to only include the share of customers in the constraint, which are expected to accept the offer of player $i$.

After having defined the Bayesian game, the goal becomes to understand if an equilibrium pricing strategy $x_i(t_i)$ for each player exists, which forms a Bayesian Nash equilibrium of the game. A strategy in a Bayesian game is a function $x_i(t_i)$, where, for each type $t_i \in T_i$, $x_i(t_i)$ specifies an action from a feasible set $A_i$ (Gibbons, 1992). Strategies form a Bayesian Nash equilibrium if for each player $i$ and for each type $t_i \in T_i$, $x_i^*(t_i)$ solves:

$$\max_{a_i \in A_i} \sum_{t_j \in T_j} R_i(a_i, s_j^*(t_j); t) p_i(t_i).$$

As with the Nash equilibrium, in the Bayesian Nash equilibrium no player wants to deviate from his strategy, given the strategy chosen by the other player.

To simplify the following analysis we define the type spaces of the players to $T_1 = [0,1]$ and $T_2 = [1,2]$ and set the cell radius of both players to $z = 1/2$. The defined bounds allow for $0 \leq \alpha \leq 1$ and $t_1 \leq t_2$.[6] Since it is common knowledge that the distribution of the opponent's type is uniform, we now need to define the limits of the integral of the opponent's type in Equation (5.21). When player 1 is of type $t_1 = 0$, the type of player 2 is irrelevant for his pricing decision since no overlap can occur. Thus, the probability becomes 1. If he is of type $t_1 = 1$, all possible types $t_j$ become relevant. We can therefore form the integral as follows:

$$\int_1^{t_1+1} \left[\alpha + (1-\alpha)\,\mathrm{Prob}\{x_i < x_j\}\right] dt_2 + \int_{t_1+1}^{2} 1 dt_2. \tag{5.23}$$

Figure 5.5 explains the limits of the integral graphically. The upper triangle defines all type combinations at which the cells do not overlap and, thus, the probability of overlap is zero. In the lower triangle $t_2 - t_1 < 1$ and the player needs to integrate over the opponent's type. The integral to be formed by player 2 can be found accordingly and is given by:

$$\int_0^{t_2-1} 1 dt_2 + \int_{t_2-2}^{1} \left[\alpha + (1-\alpha)\,\mathrm{Prob}\{x_j < x_i\}\right] dt_2. \tag{5.24}$$

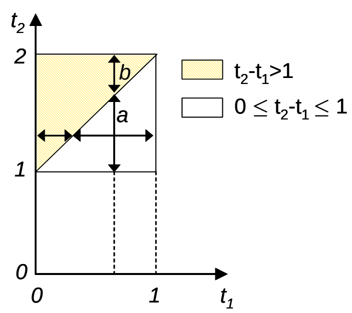



**Figure 5.5:** Graphical representation of the players types in the game of incomplete information, showing the two parts of the integral $a$ and $b$.

The second simplification for the following analysis concerns the definition of $\alpha$. Instead

[6]The chosen setup may seem artificial. However, the setup allows for any cell overlap scenario and avoids potential discontinuities for $t_2 - t_1 = 0$ since it can only occur at the space boundary when $t_1 = t_2 = 1$.

of using the exact Equation (5.18) for describing the percentage of cell overlap, we use a linear approximation $\alpha = \frac{t_2 - t_1}{2z} = t_2 - t_1$. The approximation is based on the observation that the exact function is reasonably close to $t_2 - t_1$.

Figure 5.6 provides an example for a specific setup of the game with $t_1 = 1/2$ and $t_2 = 5/4$. In the following we describe three solution approaches for identifying possible equilibrium pricing strategies that form a Bayesian Nash equilibrium. In all examples we use the setup given in Table 5.3.

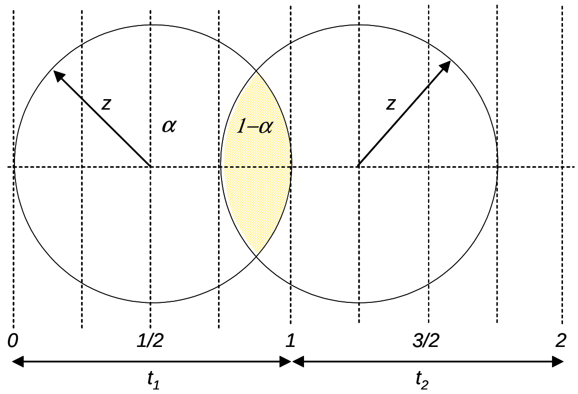



**Figure 5.6:** Graphical representation of the game with $t_1 = 1/2$ and $t_2 = 5/4$ and the intervals on which both players' types are uniformly distributed.

| Parameter | Value |
|---|---|
| $e$ | 100 |
| $f$ | 1 |
| $r$ | 3/10 |
| $C_{max}$ | 100 |
| $c$ | 1 |

**Table 5.3:** The setup used for the examples

## ▸ 5.3.3 Bayesian Nash equilibrium in linear pricing strategies

As a first step we simplify the exposition by looking for a linear equilibrium. By assuming a linear equilibrium we do not restrict the player's strategy spaces to include only linear strategies but allow them to choose arbitrary strategies and see if there is an equilibrium which is linear (Gibbons, 1992). We choose the following linear equilibrium strategies:

$$x_1(t_1) = b_1 - c_1 t_1, x_2(t_2) = b_2 - c_2(2 - t_2).$$

With $b_i \geq 0$ and $c_i \geq 0$ we can transform the probability in Equations (5.21) and (5.22) to

$$\text{Prob}\{x_1 < x_2\} = \text{Prob}\{x_1 < b_2 - c_2(2 - t_2)\} = \text{Prob}\{t_2 < \frac{b_2 - x_1}{c_2}\} = \frac{b_2 - x_1}{c_2}, \tag{5.25}$$

and

$$\text{Prob}\{x_2 < x_1\} = \text{Prob}\{x_2 < b_1 - c_1 t_1)\} = \text{Prob}\{t_1 < \frac{b_1 - x_2}{c_1}\} = \frac{b_1 - x_2}{c_1}. \tag{5.26}$$

By integrating over the opponent's type according to Equation (5.23) we obtain the constrained maximisation problem for player 1, which now includes the parameters of player 2's pricing strategy:

$$\max_{x_1} \frac{x_1 \left(e - f x_1\right)\left(-\frac{x_1 t_1^2}{2c_2} + \frac{b_2 t_1^2}{2c_2} - \frac{t_1^2}{2} + 1\right)}{r} - \frac{c\gamma \left(e - f x_1\right)\left(-\frac{x_1 t_1^2}{2c_2} + \frac{b_2 t_1^2}{2c_2} - \frac{t_1^2}{2} + 1\right)}{r},$$

subject to

$$\frac{c \left(e - f x_1\right)\left(-\frac{x_1 t_1^2}{2c_2} + \frac{b_2 t_1^2}{2c_2} - \frac{t_1^2}{2} + 1\right)}{r} \leq C_{max}.$$

As in the one-provider scenario we can use the Lagrangian multiplier method to find a solution for $x_1$. As a valid solution in the defined range for $x_1$ we obtain:

$$x_1(t_1, b_2, c_2) = \frac{e}{2f} + \frac{b_2 t_1^2 + c_2\left(2 - t_1^2\right)}{2t_1^2} \tag{5.27}$$

$$- \frac{\sqrt{c^2 t_1^4 \left(e - f b_2 + f c_2\right)^2 - 4cfc_2\left(ce - 2rC_{max} + cft_1^2\left(c_2 - b_2\right)\right) + 4c^2 f^2 c_2^2}}{2cft_1^2} \tag{5.28}$$

The solution still contains the opponent's parameters of the linear pricing strategy. Since with $t_2 = 2$ the resulting price is simply $b_2$, we can interpret this parameter as the monopolistic price; $c_2$ can be interpreted as a proxy for the player to anticipate how strongly competition influences prices. In the numerical example $b_2 = 70$ and we can plot the function $x_1(t_1, c_2)$ as a three-dimensional graph, which is shown in Figure 5.7[7].

Following the same procedure for player 2 provides us with the pricing function $x_2(t_2, b_1, c_1)$, which we omit here. If an equilibrium in linear pricing strategies exists, it must be possible to find a $c_1$ for which Equation (5.25) becomes equal to Equation (5.27) for all possible types $t_1$. However, as can be seen in Figure 5.7, the derived solution is non-linear in $t_1$ and therefore no equilibrium solution in linear pricing strategies can exist.

## ▶ 5.3.4 Bayesian Nash equilibrium in hyperbolic equilibrium pricing strategies

The results in the game of complete information have motivated us to analyse if the game of incomplete information has a Bayesian Nash equilibrium in hyperbolic pricing

[7]The monopolistic price is given by $x_B = \frac{e}{f} - \frac{rC_{max}}{cf}$. (See Section 5.3.1)

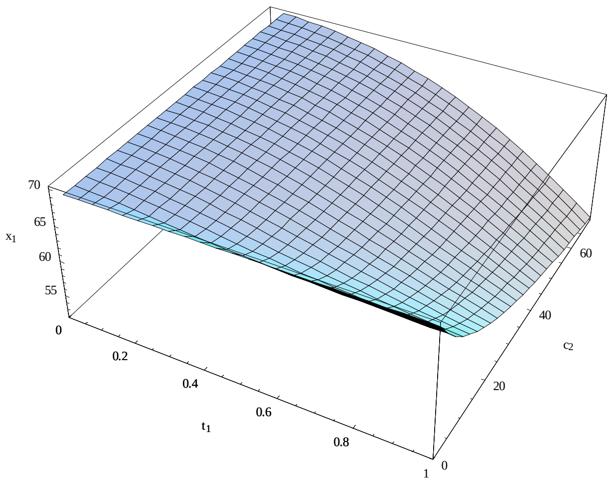



**Figure 5.7:** The resulting pricing strategy $x_1(t_1, c_2)$ of player 1 for $b_2 = 70$ when assuming an equilibrium in linear pricing strategies.

strategies. We therefore define the pricing functions as

$$x_1(t_1) = b_1 - \frac{c_1}{1 + (t_1 + 1)},\; x_2(t_2) = b_2 - \frac{c_2}{1 + (t_2 - 1)}.$$

As with the linear strategies and $b_i \geq 0$ and $c_i \geq 0$ we can use the pricing strategies to find

$$\text{Prob}\{x_1 < x_2\} = \text{Prob}\{x_1 < b_2 - \frac{c_2}{1 + (t_2 - 1)}\} = \text{Prob}\{t_2 < 2 - \frac{c_2}{b_2 - x_1}\} = 2 - \frac{c_2}{b_2 - x_1}, \quad (5.29)$$

and

$$\text{Prob}\{x_2 < x_1\} = \text{Prob}\{x_2 < b_1 - \frac{c_1}{1 + (1 - t_1)}\} = \text{Prob}\{t_1 < 2 - \frac{c_1}{x_2 - b_1}\} = 2 - \frac{c_1}{x_2 - b_1}. \quad (5.30)$$

We follow the identical solution path as described in the previous section by using the Lagrange Multiplier method and solving the constrained maximisation problem consisting of $\partial L / \partial x_1$ and the slackness condition defining the resource constraint of the wireless network. We obtain the following solution for player 1:

$$\begin{aligned} x_1(t_1, b_2, c_2) \quad = \quad & \frac{ct_1^2 (e - fc_2) + 2ce - 2rC_{max} + cfb_2 \left(t_1^2 + 2\right)}{2cf \left(t_1^2 + 2\right)} \\ - \quad & \frac{\sqrt{\left(ct_1^2 (e + fb_2 - fc_2) + 2ce - 2rC_{max} + 2cfb_2\right)^2 - 4cf \left(t_1^2 + 2\right) \left(ce (b_2 - c_2) t_1^2 + 2ceb_2 - 2b_2 rC_{max}\right)}}{2cef \left(t_1^2 + 2\right)} \end{aligned} \quad (5.31)$$

The interpretation of the factors $b_i$ and $c_i$ in the case of a hyperbolic pricing strategy is less straightforward. Since $\frac{c_2}{1+(t_2-1)}$ becomes $\frac{c_2}{2}$ for $t_2 = 2$, we cannot directly interpret $b_2$ as the monopolistic price. However, when we set $t_1 = 0$, Equation (5.31) becomes independent from $c_2$ and we need to find $b_2 = \frac{e}{f} - \frac{rC_{max}}{fc}$ (for the constrained case). Then, we can again plot $x_1(t_1, c_2)$, which is shown in Figure 5.8. This time we cannot directly

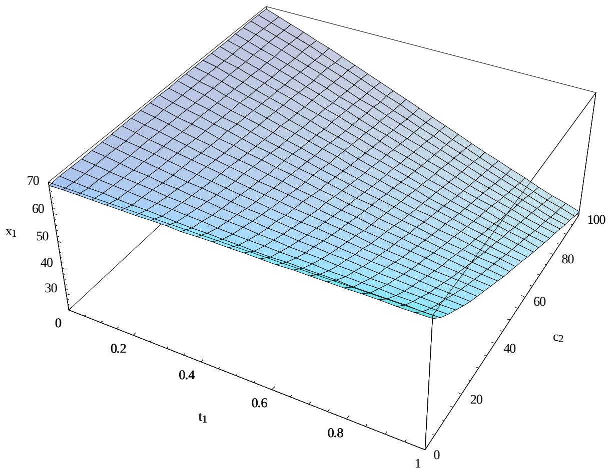



**Figure 5.8:** The resulting pricing strategy $x_1(t_1, c_2)$ of player 1 for $b_2 = 70$ when assuming an equilibrium in hyperbolic pricing strategies.

infer the conclusions from the graph as in the case of linear pricing strategies. But the resulting system of equations, consisting of the two resulting pricing functions $x_1(t_1, c_2)$ and $x_2(t_2, c_1)$, together with Equations (5.29) and (5.30) cannot be solved so that $c_1$ and $c_2$ take fixed values for all possible types $t_1$ and $t_2$. Thus, no equilibrium can be found in hyperbolic pricing strategies.

### ▶ 5.3.5 Bayesian Nash equilibrium in symmetric pricing strategies

In the last approach to find an explicit expression for the pricing strategy we examine if we can identify a symmetric equilibrium pricing strategy, in which both players' pricing function can be expressed by a single strategy function $s(\dot{})$. For this we define the probability function for each player as:

$$\text{Prob}\{x_1 < x_2\} = \text{Prob}\{x_1 < s(2 - t_2)\} = \text{Prob}\{s^{-1}(x_1) < 2 - t_2\} = 1 - s^{-1}(x_1),$$

and

$$\text{Prob}\{x_2 < x_1\} = \text{Prob}\{x_2 < s(t_1)\} = \text{Prob}\{s^{-1}(x_2) < t_1\} = 1 - s^{-1}(x_2).$$

With this definition we can now derive the solution to the integral as defined in Equation (5.23) and use the Lagrangian multiplier method for solving the constrained maximisation problem. In this process we can replace $x_i = s(t_i)$, which also gives $s_i^{-1}(s(t_i)) = t_i$ and

$s'^{-1}(t_i) = 1/s'(t_i)$. $\partial L/\partial x_1$ becomes a differential equation:

$$\begin{aligned}
\frac{\partial L}{\partial x_1} &= \frac{c\gamma t_1^2 s'^{-1}(x_1)(e - fx_1)}{2r} - \frac{x_1 t_1^2 s'^{-1}(x_1)(e - fx_1)}{2r} + \frac{cf\gamma\left(2 - t_1^2 s^{-1}(x_1)\right)}{2r} \\
&\quad - \frac{x_1 f\left(2 - t_1^2 s^{-1}(x_1)\right)}{2r} + \frac{(e - fx_1)\left(2 - t_1^2 s^{-1}(x_1)\right)}{2r} \\
&= \frac{c\gamma t_1^2 (e - fs(t_1))}{2rs'(t_1)} - \frac{t_1^2 s(t_1)(e - fs(t_1))}{2rs'(t_1)} + \frac{cf\gamma\left(2 - t_1^3\right)}{2r} - \frac{fs(t_1)\left(2 - t_1^3\right)}{2r} \\
&\quad + \frac{(e - fs(t_1))\left(2 - t_1^3\right)}{2r}.
\end{aligned} \tag{5.32}$$

Note that through the substitutions the above equation becomes independent of $x_1$ and only depends on the player's type $t_1$ and the function $s(t_1)$. We can now set Equation (5.32) to zero and write the slackness condition as:

$$\gamma\left(C_{max} - \frac{c(e - fs(t_1))\left(2 - t_1^3\right)}{2r}\right) = 0. \tag{5.33}$$

If an equilibrium in symmetrical pricing strategies exists, there must exist a function $s(\dot{})$ which satisfies Equations (5.32) and (5.33) for all $t_1 \in T_1$. Figure 5.9 visualises the function $s(t_1)$ when solving Equation (5.33). We find that $s(t_1)$ is not a solution for Equation (5.32). Consequently, with the given approach, no equilibrium in symmetrical strategies can be found for the game.

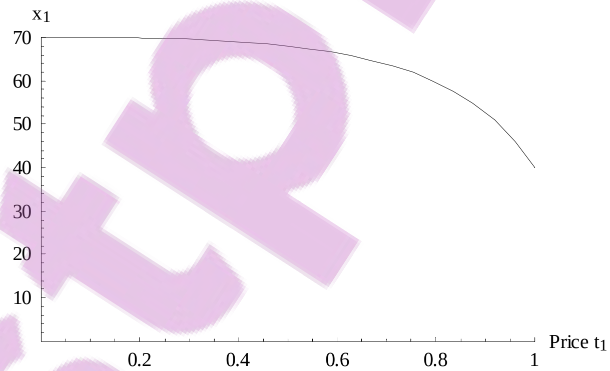



**Figure 5.9:** The resulting function when solving the slackness condition to $s(t_1)$.

## ▸ 5.4 A Heuristic Approximation Framework for the Two-Provider Case

In the previous two chapters we have seen that the proposed pricing strategy for a wireless provider facing direct competition over prices is difficult to fully comprehend analytically. This is mainly for two reasons: first, the constrained maximisation problem can usually not be solved in explicit form when we model resource constraints with non-linear behaviour such as transmission power. Second, when introducing a second cell, which is operated by an alternative provider and either fully or partially overlaps with the cell of the first provider, price setting becomes a game, for which we could not find an equilibrium pricing strategy in the case of incomplete information. We have therefore decided to explore the described pricing model by simulation. To use simulation we need to implement our model in a way that does not require the full solution of the mathematical model, but can derive an approximate solution to the constraint maximisation problem.

In the following we describe how demand on the wireless market is modelled and how providers derive estimates from the various variables. We then present the approximation procedure to find a near-optimal solution to the constrained maximisation problem and briefly explain the technical admission control function implemented in the simulation platform. Finally, we elaborate on how the statistical analysis of the simulation output has been performed to retrieve the results presented in the next chapter.

### ► 5.4.1 User demand and utility

The developed simulation platform allows us to represent each individual customer as an agent. The service request behaviour of each agent is modelled as an alternating on-off process, which is described by two, exponentially distributed random variables with the means $\mu_{inactive}$ and $\mu_{active}$, respectively (Figure 5.10). Therefore, the resulting demand for a provider is an overlaid process consisting of multiple alternating on-off processes depending on the user density and the cell size.

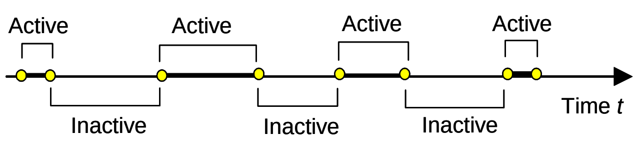



**Figure 5.10:** The activity and inactivity process for an agent representing a single user.

To model user utility for services we use a uniform random variable. Every time a customer changes from the inactive to the active state he generates a new random utility value. Once a customer receives one or more price offers he maximises his net utility by selecting the offer with the lowest price if this price is below or equal to his utility. Otherwise, he rejects all offers.

### ► 5.4.2 Modelling of the estimators

To let a provider learn about the demand situation and the level of competition within its cell range, we need to implement a procedure to form estimators. While some events are directly observable for the provider, we need to take further assumptions about other estimators such as the estimator of the user demand structure and the estimator of the prices set by the competitor.

All estimators are formed with a sliding window technique, which collects a predefined number of events (or objects). The sliding window $\Delta t$ moves with the progressing simulation time as shown in Figure 5.11. The estimator is an approximation of the real variable and the accuracy of the measurement depends on the number of data points taken into consideration.

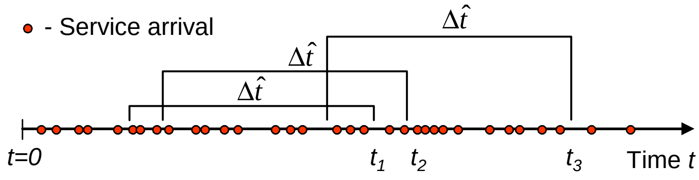



**Figure 5.11:** Sliding window of the historical data used for calculating the estimators.

The following estimators are created by each provider to be used with the approximation procedure:

**Service activation rate ($\hat{\lambda}(x_i, z_i)$)** An essential element of the model is the service activation rate $\hat{\lambda}(x_i, z_i)$, which changes with the price $x_i$ and the cell radius $z_i$. While providers can form a general estimator for the service activation rate, they are unaware of the underlying customer utility structure. To provide a base station agent with information about the users' willingness-to-pay we include such information in every service request. However, the base station agent cannot read this value directly but can only use it to form the aggregate estimation function. Otherwise, and in absence of competition, a base station could perfectly price-discriminate between users. It would be hard to justify this approach since users usually have no incentives to reveal their true utility to the provider directly.[8] By using customer utility in aggregated form we enable a provider to learn from the ongoing market activity. Note that by using the described estimation method we are not restricted by using simple demand functions, such as linear functions in the examples in Section 5.1.

**Service duration ($1/\hat{r}$)** To measure service duration a provider records the start and stop time of each active service. It can then form an estimator $1/\hat{r}$ from the historical data using the sliding window technique. We assume that the service duration is independent from the price $x_i$.

**Percentage of single access customers ($\hat{\alpha}(t_i, t_j)$)** In addition to the estimation of the service activation rate, a provider needs to form an estimator about the share of customers

[8]However, a customer may be willing to reveal this information indirectly. Data could be collected by conducting customer surveys or by observing customer behaviour over longer periods.

it can serve monopolistically. To make this information available to the provider in the simulation environment, we assume that each customer transmits his access options with each service request message. As previously explained, $\alpha$ is a function of both providers types, which may be determined by the position of the base station or the maximum cell radius given by the technological constraints of the network cell.

**Competitor price ($\hat{\beta}(x_i, x_j)$)** As with the estimator for $\alpha$ we need to let the provider form an estimate of the current competitive price. With this estimator we can form a probability function, which gives the probability of having a price set below the competitor or formally $\hat{\beta}(x_i, x_j) = \text{Prob}\{x_i < x_j\}$. To let providers have access to competitive prices we let customers transmit the alternative price offers with each service accept or service reject message. As with the estimation of the demand structure, a provider cannot use this information directly (for example, by reading out the latest price update of the competitor) but can only form the competitive price estimator in aggregate.

While with the model developed above we require the provider to have access to a lot of information about the customer demand structure and the competitive environment, we can argue that providers usually have precise information about the their customer base. Information about competitive prices may also be obtained indirectly by, for example, measuring offer acceptance ratios in different rings of the cell and comparing them with the utility structure of the customer demand. While this may also be possible to model in the simulation, it would require extensive learning iterations to let providers form stable and reliable estimators before running the actual experiments.

Figure 5.12 provides an overview of the different estimators and the points at which data is collected.

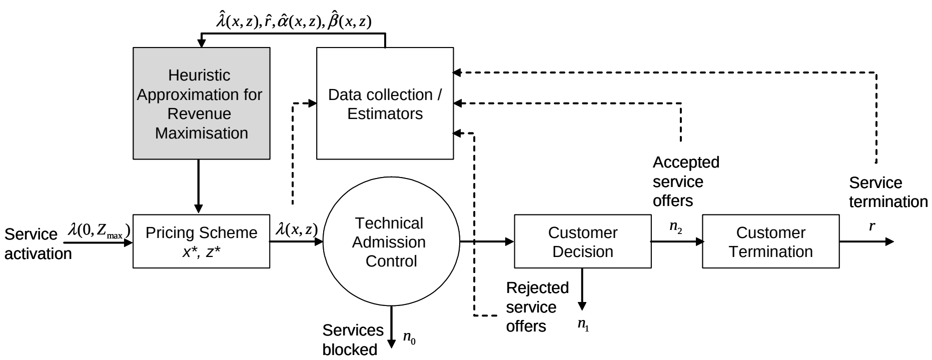



**Figure 5.12:** Overview of the estimators derived by a provider using the sliding window technique.

### ▶ 5.4.3 The approximation procedure

To find a near-optimal solution to the constrained maximisation problem described in Section 5.2.3 we have developed a straightforward approximation model, which uses a finite grid to calculate feasible revenue for a finite number of price/cell-radius combinations. First, we let each provider calculate its expected revenue with the following equation:

$$\mathbb{E}\left[R_i(x_i, z_i)\right] = x_i \frac{\hat{\lambda}(x_i, z_i)}{\hat{r}}(\hat{\alpha} + (1 - \hat{\alpha})\hat{\beta}), \tag{5.34}$$

which matches the model described in Section 5.3.2.

Second, a provider needs to check the feasibility of all price/cell-radius combinations. As described in Section 5.2.3 we consider two constraints the constraint in code slots and the constraint in transmission power in the forward link:

$$\frac{c}{2}\frac{\hat{\lambda}(x_i, z_i)}{\hat{r}}(\hat{\alpha} + (1 - \hat{\alpha})\hat{\beta}) \leq C_{max} \tag{5.35}$$

and

$$\frac{c\pi\gamma d^2 \hat{\lambda}(x_i, z_i)(\hat{\alpha} + (1 - \hat{\alpha})\hat{\beta})(\eta d^4 + 3k\theta P_{max})}{6 Z_{max} k \hat{r}(W + c\gamma\theta)} \leq P_{max} \tag{5.36}$$

To determine the approximate revenue-maximising values for possible $x, z$ combinations, the model uses a matrix $R_{xz}$, which, for a given grid size $\Delta x$ and $\Delta z$, contains the corresponding revenue for each value combination using Equation (5.34). The grid size used in this matrix determines the precision of the solution. An example of the matrix is shown in Table 5.13(a).

In a second and third step the provider calculates the feasibility of every price/cell-radius combination used in the revenue matrix in terms of the code and power constraints given by the Equations (5.35) and (5.36). He creates two additional matrices $C_{xz}$ and $P_{xd}$ with identical dimensions, which contain the results of the feasibility check. If, for the given combination of $\hat{\lambda}$ and $1/\hat{r}$, sufficient resources are available, the corresponding cell in the matrix is marked with a 1. If resources are insufficient, the cell is marked with a 0. An example is shown in Figures 5.13(b) and 5.13(c).

Finally, the provider multiplies the three matrices, $R_{xz} \times C_{xz} \times P_{xz}$, and derives a new matrix $F_{xz}$ containing the feasible revenue for all combinations of price/cell-radius combinations considered in the matrix (Figure 5.13(d)). He determines the $(\hat{x}^*, \hat{z}^*)$ combination that maximises revenue under the given capacity and power constraints. The resulting matrix can also be visualised in a three-dimensional graph as shown in Figure 5.14. It is easy to see the revenue-maximising $(\hat{x}^*, \hat{z}^*)$-pair in the graphic. The solution is an approximation of the exact values.

Figure 5.15 gives an overview of the implementation from a process view, which has been described above. The factor $y$ determines the number of events until the approximation is rerun.

| | $0 | $30 | $50 | $70 | $90 |
|---|---|---|---|---|---|
| 100m | $0 | $200 | $140 | $98 | $69 |
| 300m | $0 | $600 | $420 | $294 | $206 |
| 500m | $0 | $1,230 | $861 | $603 | $422 |
| 700m | $0 | $2,190 | $1,533 | $1,073 | $751 |
| 900m | $0 | $4,530 | $3,171 | $2,220 | $1,554 |

(a) Example revenue matrix.

| | $0 | $30 | $50 | $70 | $90 |
|---|---|---|---|---|---|
| 100m | 1 | 1 | 1 | 1 | 1 |
| 300m | 1 | 1 | 1 | 1 | 1 |
| 500m | 1 | 1 | 1 | 1 | 1 |
| 700m | 0 | 0 | 1 | 1 | 1 |
| 900m | 0 | 0 | 0 | 1 | 1 |

(b) Example rate constraint matrix

| | $0 | $30 | $50 | $70 | $90 |
|---|---|---|---|---|---|
| 100m | 1 | 1 | 1 | 1 | 1 |
| 300m | 0 | 1 | 1 | 1 | 1 |
| 500m | 0 | 1 | 1 | 1 | 1 |
| 700m | 0 | 0 | 1 | 1 | 1 |
| 900m | 0 | 0 | 0 | 0 | 0 |

(c) Example power constraint matrix

| | $0 | $30 | $50 | $70 | $90 |
|---|---|---|---|---|---|
| 100m | $0 | $200 | $140 | $98 | $69 |
| 300m | $0 | $600 | $420 | $294 | $206 |
| 500m | $0 | $1,230 | $861 | $603 | $422 |
| 700m | $0 | $0 | $1,533 | $1,073 | $751 |
| 900m | $0 | $0 | $0 | $0 | $0 |

(d) Resulting feasible revenue matrix

**Figure 5.13:** Example matrices in the numerical solution to calculate revenue and capacity constraints for different value pairs $(x, z)$.

### ▸ 5.4.4 Modelling of the technical admission control function

Besides modelling the code and power constraints of a WCDMA-based network for use in the approximation module, we need to implement a technical admission control function to decide if a new service request can be accepted or needs to be rejected from a technical point of view (see Figure 5.12). Since the heuristic approximation procedure only considers the average number of service requests in the cell, we also need to model resource allocation in the actual system.

The admission control function of the *AdSim* platform is based on a simple physical model of the bit-energy-to-noise-density ratio, which has already been given in Equation (5.10). While in the downlink, the interference $I_i$ is simply the sum of the transmission power assigned to all other active mobile terminals ($I_i = \sum_{j \neq i} p_j$), for the uplink, the interference for a user is given by the sum of the sending power $p_j$ of all other mobiles times the individual channel gain $g_j$ ($I_i = \sum_{i \neq j} g_j p_j$). A feasible setting in the uplink therefore requires that all mobile terminals can maintain the target SINR, given their limitations in transmission power.

In both transmission directions we have implemented a simple algorithm that iteratively approximates the power levels for each active link.[9] With this algorithm the base station, for a given moment in time, can check the feasibility of the current setup of active links and associated quality guarantees. If a new service request arrives at the base station

[9]For each service class we have defined the parameters of a target BER and a transmission rate in both directions.

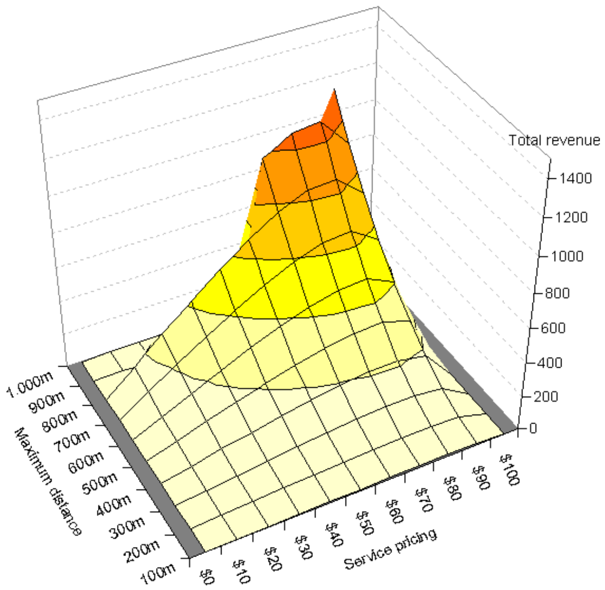



**Figure 5.14:** Example for the graphical visualisation of the feasible revenue matrix $S_{xz}$.

it can be admitted if, by including the new request, the quality guarantees of all existing links can still be met (Bambos et al., 1995). If, in both directions, a feasible power allocation can be found, the service request is admitted, otherwise the request is rejected.[10]

While this model can guarantee an average quality of the link in terms of the BER on the physical level, no guarantees can be made in regard to other quality parameters such as delay and jitter on the link layer.

### ► 5.4.5 Statistical analysis of simulation output

After having explained the implementation of the approximation framework and the properties of the estimator functions used with the approximation technique, we explain the output analysis of the simulation data. This is important since the quality of the obtained results and corresponding conclusions strongly depends on the rigor used with the analysis.

Besides more complex methods to analyse simulation output, two principle approaches can be distinguished. The first, called *Batch Means (BM)*, clusters output data of a single long simulation run into smaller batches (Alexopoulos and Goldsman, 2004). The batches are then used to calculate the estimators for the steady-state mean and variance. For large batch sizes the experimenter assumes that the batch means are approximately *independent and identically distributed* (IID). The advantage of this method is that it is less influenced by the initiation period.

[10]Note that by implementing this algorithm at the base station we do not aim at providing a continuous power control function. The algorithm is solely used by technical admission control to check the feasibility of admitting the new request. The underlying radio resource management functions, which are not modelled in the simulation, are assumed to perform continuous power control and other functions to maintain the link quality.

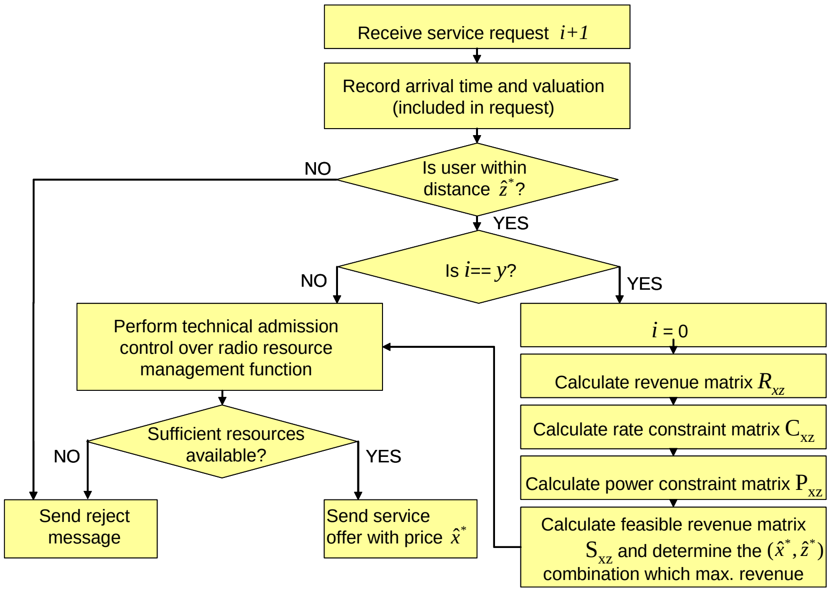



**Figure 5.15:** Process flow-chart of the model to approximate the optimal values for $x$ and $z$ in the simulation environment.

The second method, called *Independent Replications (IR)*, requires conducting multiple, independent replications of the simulation process of shorter length (Alexopoulos and Goldsman, 2004). Then, for each replication, an estimator for the steady-state is formed. The results can be assumed to be IID and are used in the next step to generate the output estimators (the mean and variance) of the experiment. While IR has the advantage of creating simulation data, which can be assumed to be IID since different initial conditions can be used, it suffers from the influence of the initiation period, which needs to be eliminated first.

A recently presented method combines both approaches by running a small number of replications and grouping the data of each replication into batches. This method, called *Replicated Batch Means (RBM)*, provides a good balance between the BM and IR methods and has found a wider acceptance (Argon et al., 2006). The method requires $k$ independent and identically distributed replications to estimate the underlying steady-state mean of the underlying stochastic process. $k$ is assumed to be small so that the replication means will not be sufficient to form the estimator. Therefore, the data in each replication is further grouped into $b$ batches, each containing $m$ observations. Let $X_j^{(r)}(m)$ be the $j$th batch mean of the $r$th replication for $r = 1, .., k$ and $j = 1, .., b$. $\overline{Y}_r(b, m) = \sum_{j=1}^{b} X_j^{(r)}(m)/b$ is the $r$th replication mean for $r = 1, .., k$. Then, the point estimator for the steady-state mean $\mu$ is given by:

$$\overline{X}_{RBM}(k, b, m) = \sum_{r=1}^{k} \overline{Y}_r(b, m)/k,$$

and

$$\hat{V}_{RBM}(k,b,m) = \frac{m}{kb-1}\sum_{r=1}^{k}\sum_{j=1}^{b}(\overline{X}_j^{(r)}(m) - \overline{X}_{RBM}(k,b,m))^2$$

is the RBM estimator of $\sigma^2$.

We use the RBM method for the purpose of this simulation study. We run 5 replications of each experiment, each of 10 hours length. To eliminate the influence of the initial transient period, we follow the procedure described by Heidelberger and Welch (1983)[11] using a plot of the smoothed simulation data from multiple replications. We have determined that for all tested configurations the initiation period is below 45 minutes. We therefore let each replication run for 11 hours and discard the first 60 min of data in the analysis.

To create a single batch we collect the data of each base station over a period of $5min$. We take a time-weighted mean over the period, which is recorded in the database. For the calculation of the estimators we set $m = 5$, according to the length of the interval.[12] Since our analysis involves two decision variables, the price $x$ and the cell-radius $z$, the definition of the batches becomes more complex. Regularly, providers jump between different cell radii to optimise their revenue depending on the current state of the estimators. When multiple cell-radii have been selected by a base station within a single batch, we create several separate batches, which can be distinguished by the cell radius. In this way we collect data for each cell radius and statistical analysis can be conducted separately for each cell radius. If the number of batches for one cell radius is too small, i.e. below 10 in any of the replications, we eliminate the particular cell radius from our result.

## ► 5.5 Experimental Results

This section presents the results of an extensive simulation study to experimentally explore the developed pricing strategy and the approximation procedure as described in Section 5.2.3. We use the *AdSim* simulation platform, which is described in detail in Chapter 6. The experiments were run in the computer labs of the University of Auckland Business School. We concurrently used 32 machines to run the required replications for each data point, which took about 4 weeks to complete.

We first present the results from a one-provider, one-cell scenario to develop a basic understanding about the sensitivity of the main output variables when gradually changing a single input variable such as the user density or the service class. We then turn our attention to a two-provider scenario, in which each provider operates one cell and present the results from different simulation scenarios.

In Table 5.4 the input parameters of the simulation platform are given. Unless other-

[11] The method is described in detail in Law and Kelton (2007).

[12] We have measured that in average, a provider updates its price about once a minute. However, this value depends on the arrival rate and can differ between different simulation setups. To not further complicate the analysis we have decided to fix m to 5.

| Description | Value |
|---|---|
| Simulation Length | 10 hours |
| Number of Replications | 5 |
| Maximum Cell Size $Z_{max}$ | 1,000m |
| Code Slots / Chip Rate $C_{max}$ | 3.84 Mcps |
| Transmission Power $P_{max}$ | 30W |
| Channel Gain $g_i$ | $g_i = kd_i^{-u}$,<br>$k = 1.80E - 14$ and $u = 4$ |
| Background Noise Power $\eta$ | 1.00E-13 |
| Code Orthogonality Factor $\theta$ | 0.1 |
| Size of Sliding Window | 500 Objects |
| Size of Approximation Grid - Price | 50 Steps |
| Size of Approximation Grid - Radius | 5 Steps |
| Price Update Sequence | After 50 New Requests |
| Number of Users per $km^2$ | 50 - 250 |
| Customer Distribution | Random Spatial Distribution |
| Mean Service Inactivity Time $\mu_{inactive}$ | 20 minutes |
| Mean Service Activity Time $\mu_{active}$ | 5 minutes |
| Demand Structure | $U(0, 100)$ |

**Table 5.4:** Simulation Parameters used in all simulations unless otherwise stated.

wise stated, all experiments are run with this setup. This allows us to directly compare the results from different experiments and understand the consequences of changing a single input parameter. The input parameters can be further clustered into four categories; general parameters, the parameters for the cell setup, the parameters of the approximation module, and the configuration of customers in the simulation area. In Table 5.5 the parameters of the service classes which have been used with all experiments are defined.

### ► 5.5.1 Parameterisation of the simulation platform

Before we start with the description of the actual simulation results, we provide a brief description of the results of the pre-experiments. Such experiments have been conducted to parameterise the simulation platform so that output precision and computational demand are reasonably balanced. Two areas can be distinguished: the general environment parameters and the model-specific parameters. All experimental results shown in the following were conducted with a fixed user density of 200 users/$km^2$ and all users requesting service class 2. We have run identical experiments with different user densities and service classes and have derived similar results.

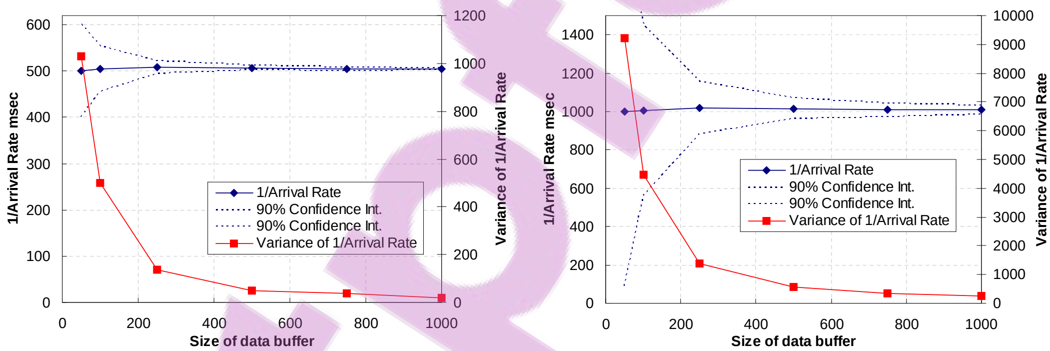
| Service Class | Guaranteed Bandwidth | Maximum Bit Error Ratio (BER) | Example Application |
|---|---|---|---|
| 1 | 32 | $1.00E - 8$ | Medium Quality Audio Streaming |
| 2 | 64 | $1.00E - 8$ | High Quality Audio Streaming |
| 3 | 144 | $1.00E - 6$ | Medium Quality Video Streaming |
| 4 | 256 | $1.00E - 5$ | High Quality Video Streaming |

**Table 5.5:** Service classes used in the experiments.

### General environment parameters

The most fundamental environment parameter is the size of the sliding window to derive the estimators for the service activation rate $\lambda(x)$, the service activity duration $r$, and the maximum customer valuation $v_{max}$. To understand the quality of the estimators relative to the size of the sliding window we conducted an experiment in which we varied the size of the window between 50 objects and 1000 objects. Figure 5.16(a) depicts the estimator $\hat{\lambda}_z(0)$, the 90% confidence interval, and the variance against the buffer size.[13] We can observe that the variance quickly decreases up to a buffer size of 250, after which it decreases at a slower rate. Figure 5.16(b) shows the same information for $\hat{\lambda}_z(50)$. With a window size larger than 250 objects, the variance becomes reasonable small.

From the data of the same experiment we analysed how prices develop when varying the size of the sliding window (Figure 5.5.1). For values smaller than 250 objects the approximation procedure selects smaller cell radii at lower prices. This is because the estimators have a high variability, which lets the approximation procedure run into resource constraints more often and forces it to select a price/cell-radius combination with smaller cell radius. For values above 250 objects prices stabilise and the approximation procedure reduces the price/cell-radius selection to the largest and second-largest cell radius. In the following experiments we therefore use a buffer size of 500.




(a) Variance of the arrival rate estimator $\hat{\lambda}(0)$ for the entire cell for different buffer sizes.

(b) Variance of the arrival rate estimator $\hat{\lambda}(50)$ for the entire cell for different buffer sizes.

**Figure 5.16:** Variance of estimator functions for different buffer sizes.

### Model-specific parameters

The main model-specific parameters are the price update frequency and the grid size of the approximation function. Ideally, we want to update prices with each new request arriving at the network cell. However, the computational procedures to update the estimators and to approximate the revenue-maximising prices are computationally demanding. We therefore define a certain value which determines after how many new service requests the revenue-maximisation procedure is rerun. We vary this value from 5 to 100 events

[13] As in the previous sections $\lambda_z$ denotes the service activation rate in a circular area with radius $z$.

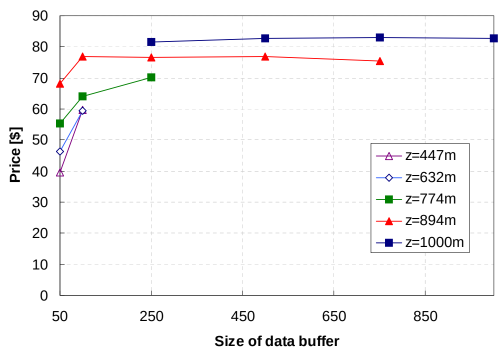
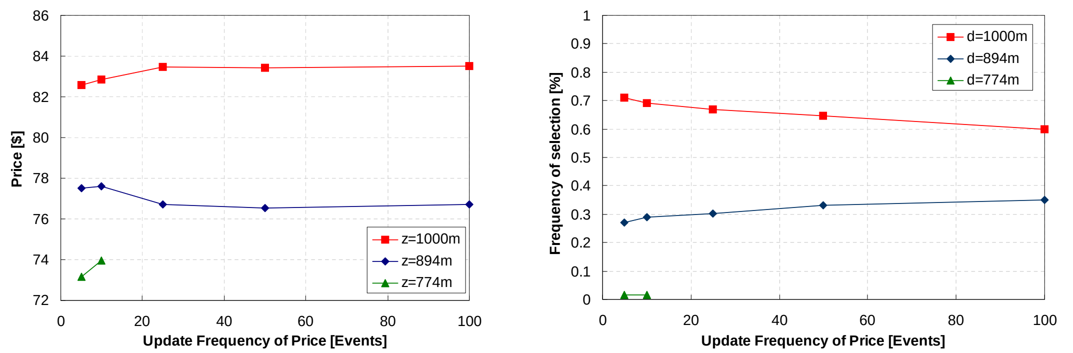



**Figure 5.17:** Prices for different buffer sizes.

and observe the change in output variables. Figure 5.18(a) shows that with shorter update cycles the approximation procedure identifies a solution with a cell radius of $z = 774m$, which is not selected when lowering the frequency of the price updates. However, Figure 5.18(a) shows that the percentage of the selection is very low. We have therefore decided to run all following experiments with a value of 50 events before the price update procedure is rerun.




(a) Prices against update frequency.

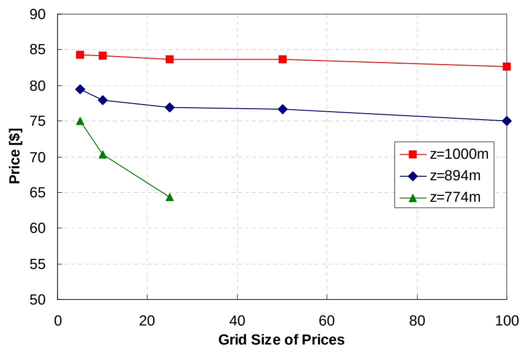
(b) Selection frequency of different radii against update frequency.

**Figure 5.18:** Prices and selection frequency for different update frequencies.

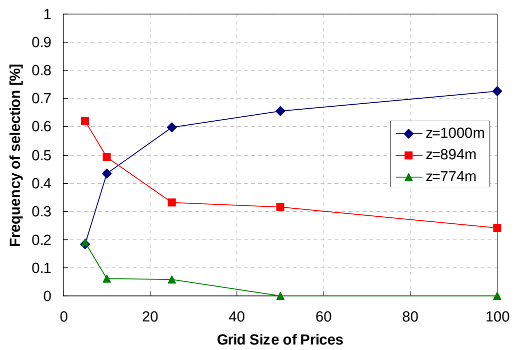
To understand the influence of the revenue grid dimensions on the quality of the output variables we conducted an experiment in which we used an identical setup for users and cells and varied the grid size of the price dimension between 5 and 100.[14] Figure 5.19(a) shows that with values below 25 the approximation procedure selects solutions on three different cell radii, while for larger values, only two different solutions are picked. Since with larger values the heuristic is able to make a more granular decision, the revenue-maximising solution is always found on the largest or second largest cell radius. As in the previous case the selection percentage of the solution with radius $z = 774m$ is very low (Figure 5.19(b)). It can also be observed that the selection percentages stabilise

[14]With a value of 5 the heuristic can only distinguish between five different price levels

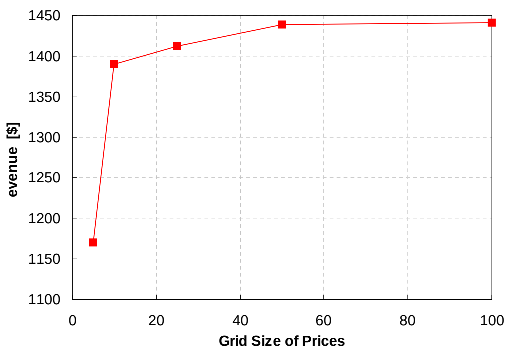
with a grid size larger than 50. Figure 5.19(c) shows the revenue against the grid size. This reveals that smaller grid sizes have an adverse effect on the ability of a provider to maximise revenue. With grid sizes above 25 revenue stabilises. We have therefore decided to use a price grid size of 50 for all following experiments.




(a) Price against price grid size.




(b) Cell radius selection frequency against price grid size.



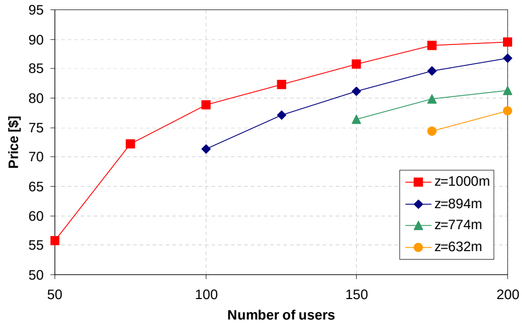

(c) Revenue against price grid size.

**Figure 5.19:** Prices, selection frequency and revenue for different grid sizes of price.

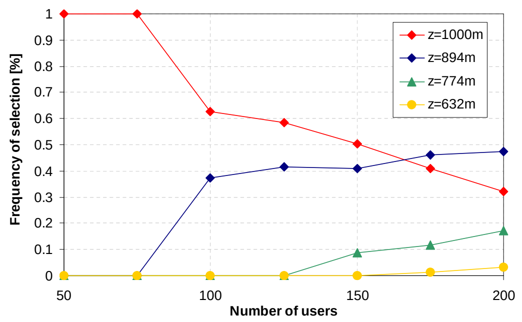
In a second experiment we varied the cell radius grid size between 3 and 20 while keeping the simulation setup fixed. As already explained, the main limitation of increasing the radius grid size arises from the quality of the estimators in each cell ring. Since with an increasing grid size the area of the rings get smaller, less data can be collected for each ring. Based on the results we have set the cell radius grid size to 5 to ensure the quality of the estimators. While this clearly constrains the cell in finely adjusting its cell radius to the optimal value, a small grid size allows us to identify statistically relevant price/cell-radius pairs. With a larger grid the results between the different cell radii cannot be clearly distinguished, which leads to problems in the overall analysis.

### ▸ 5.5.2 Simulation results for the one-provider, one-cell scenario

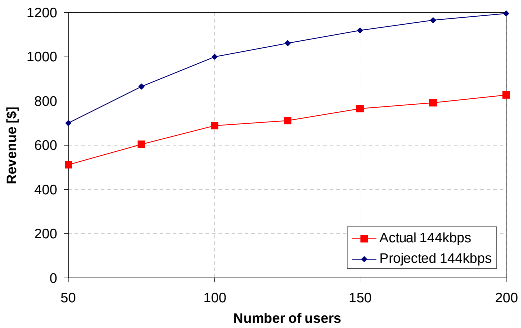
In this section we present the results from the one-cell, one provider case. The results are mainly intended to complement the results derived analytically and to verify the implemented approximation routines in the simulation environment. Additionally, the results serve as a baseline to be compared with the results in the following experiments.



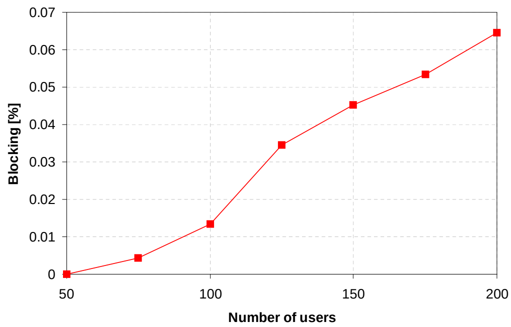

(a) Prices per cell-radius against user density.




(b) Selection frequency of different cell sizes against user density.




(c) Actual and projected average revenue against user density.




(d) Average blocking ratio against user density.

**Figure 5.20:** Analysis of the one-provider, one-cell scenario for service class 3.

### Basic experiment with varying user densities

In the first experiment we have analysed how prices and cell radii change with an increasing user density, which we vary from 50 users/$km^2$ to 200 users/$km^2$. Figure 5.5.2 provides an overview of the typical analysis we performed for each configuration for service class 3 and a maximum cell radius of $Z_{max} = 1{,}000m$. Figure 5.20(a) shows the price/cell-radius combinations against the user density and Figure 5.20(b) shows the corresponding percentage with which the price/cell-radius combinations were selected.

To gain a better understanding about the statistical significance of the results we have analysed the frequency distribution prices according to the different cell radii. Figure 5.21 shows the result of the analysis for a user density of 150 users/$km^2$ and for service class 3.[15] While all distributions overlap, we can clearly observe the shifted peak as the cell radius becomes smaller. The variance and 90% confidence interval are given in Table 5.6. This demonstrates that the different pairs can be distinguished and that the boundaries of the 90% confidence intervals do not overlap.

[15]Shown here is the frequency distribution of the batch means. The picture looks similar when drawn for reach replication separately, however with a higher diffusion.

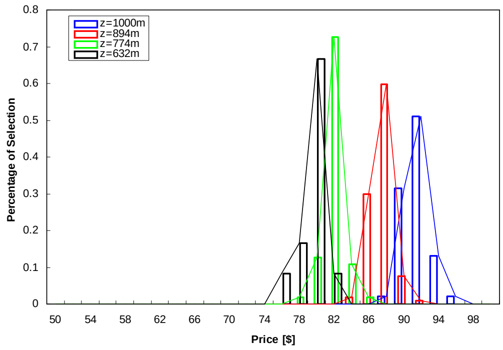
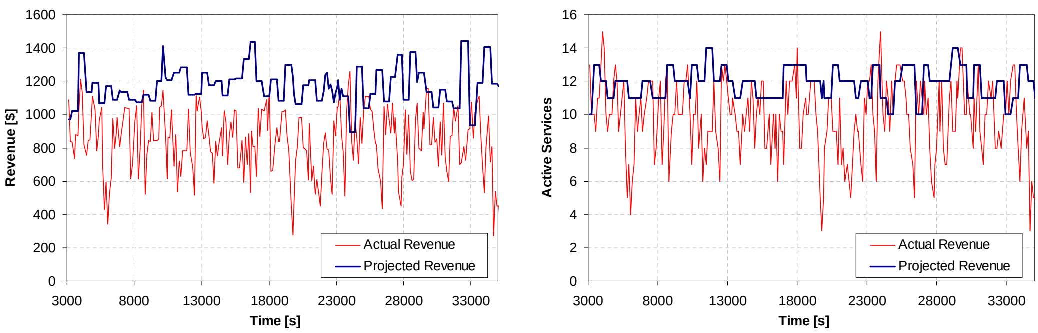



**Figure 5.21:** Frequency distribution for a user density of 150 and service class 3.

In Figure 5.20(c) the actual revenue is compared with the projected revenue from the approximation module. We can see that with an increasing load, the actual and projected values diverge. This is mainly due to three reasons. First, the approximation module optimises revenue for a steady stream of service activation. In contrast, the activation process is a stochastic process from the combined activations of all customers in range. As explained, the activation and deactivation is modelled with an exponential distribution by each customer. In times of arrival bursts, the network cannot accommodate all requests, while during other periods free resources are available.




(a) Revenue and projected revenue. (b) Active services and projected active services.

**Figure 5.22:** Comparison of revenue and service projection with actual values for one simulation run over time.

Figure 5.22(a) demonstrates this effect by showing the actual revenue versus the

| | | | | 90% Confidence Interval | |
|---|---|---|---|---|---|
| **Cell Radius** | **Average Price** | **Variance** | **Batches** | **Lower Boundary** | **Upper Boundary** |
| 1,000m | 90.09 | 4.61 | 88 | 89.67 | 90.51 |
| 894m | 86.37 | 6.30 | 112 | 85.92 | 86.82 |
| 774m | 81.10 | 9.93 | 53 | 79.96 | 82.24 |
| 632m | 77.21 | 8.33 | 13 | 77.91 | 78.90 |

**Table 5.6:** Batch variance and confidence interval for a user density of 150 and service class 3.

projected revenue for one single simulation run. We can observe that the projected revenue is sometimes matched by the actual revenue. Most of the time, however, the actual revenue is significantly lower. A second effect that plays a role is that, after a price increase, existing services are still charged with the price committed at setup time. This causes the actual revenue to be lower than the projected revenue. This explanation is confirmed by Figure 5.22(b), which shows the actual active services versus the projected active services. In difference to the large difference between projected and actual revenue, the number of projected services is more often reached or surpassed by the actual number of active services.

A third reason for a significantly lower actual revenue is that the approximation procedure assumes a uniform distribution of active customers over the coverage area of the cell. However, this may be not the case and more customer requests may be received from the outer areas of the cell while, during other periods, more customers from an area around the base station may request services.

Figure 5.20(d) shows the actual blocking ratio, which was measured during the simulation process. We measure blocking as the percentage of customers with a higher valuation than the price within the cell-radius currently selected by the approximation module. With the highest user density of 200 users/$km^2$, the blocking ratio rises to about 7%. The blocking ratio is higher with service classes requiring more bandwidth since less active services can be served and customers farther away from the base station consume significantly more resources, which leads to higher overall blocking in the cell.

The results of the entire experiment for all service classes and user densities between 50 and 200 / $km^2$ are given in the chapter appendix. Figure 5.7 shows the simulation results for a maximum cell radius of $Z_{max} = 1,000m$ and in Figure 5.7 for a maximum cell radius of $Z_{max} = 1,500m$. We show only the prices selected for the different cell radii and the percentage of selection, which tells how often a certain price/cell-radius combination has occurred at a given user density.

### Comparison of the simulation results with the analytical solution

In the next experiment we compare the results derived by the experiments with the analytical results. As already explained in Section 5.2.3, we are able to derive a numerical solution for simple service activation functions $\lambda(x)$. Since we use a uniform distribution for the users' valuation in the simulation, $\lambda(x)$ is a linear function that can be easily obtained from the input parameters. The numerical solution can then derived by solving the constrained maximisation problem, as given in Equations (5.1) to (5.4), with Mathematica 5.1. The output consists of multiple solutions, of which some are saddle points with some having imaginary elements. We select the solutions which are within the valid range of price and cell radius and calculate the corresponding revenue. Table 5.7 compares the results for service class 3 with the simulation results from the last experiment.

The results of the analytical solution with the simulation results for $z = 1000m$ are most

| | Analytical Solution | | | Experimental Solution | | | |
|---|---|---|---|---|---|---|---|
| User density | Price $x$ | Radius $z$ | Revenue $R$ | Price $x$ | Radius $z$ | Revenue $R$ | Prob. Selection |
| 50 | 51.38 | 1000 | 749.43 | 56.52 | 1000 | 737.25 | 1 |
| 100 | 78.18 | 1000 | 1023.58 | 78.47 | 1000 | 1013.67 | 0.62 |
| | 73.11 | 905 | 966.18 | | | | |
| | | | | 69.88 | 894 | 1009.43 | 0.37 |
| 150 | 85.78 | 1000 | 1097.66 | 86.14 | 1000 | 1074.49 | 0.5 |
| | | | | 80.72 | 894 | 1119.48 | 0.41 |
| | 77.45 | 807 | 1023.77 | | | | |
| | | | | 77.23 | 774 | 948.02 | 0.09 |
| 200 | 89.34 | 1000 | 1143.14 | 90.09 | 1000 | 1071.56 | 0.32 |
| | | | | 86.37 | 894 | 1129.23 | 0.48 |
| | | | | 82.34 | 774 | 1045.19 | 0.17 |
| | 80.60 | 745 | 1041.22 | | | | |
| | | | | 78.34 | 632 | 813.25 | 0.03 |

**Table 5.7:** Comparison of the solutions obtained analytically and by simulation for service class 3 and a maximum cell radius of $Z_{max} = 1,000m$.

straightforward to compare with the simulation results. We can observe a close matching of results with the prices chosen in the simulation being slightly higher. The largest difference can be found at the lowest user density (50 users/$km^2$), where the solution derived by simulation is significantly higher.

For the rest of the results we can compare only the approximate closeness of the simulation results. Since the approximation procedure uses a finite grid, the precision of the results depends of the size of the grid. We can observe that the solutions found by simulation on the given grid yield similar revenue but may not be identified as a maximum value in the analytical solution.

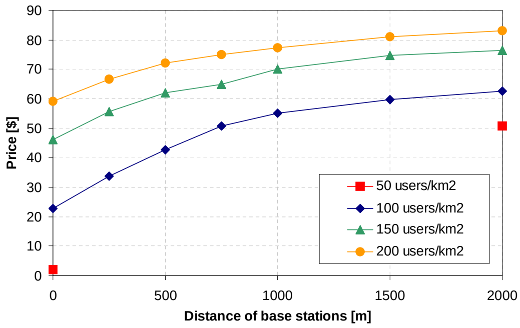
It can be seen that the analytical solution always provides two solutions, one at maximum radius, and one at a lower radius and lower price. In contrast, the simulation results provide price/cell-radius combinations for each finite radius. It can also be observed that the global maximum of the analytical solution is always reached with the price/cell-radius combination that sets the cell radius to the maximum value. This is not the case in the data obtained from simulation.

### ► 5.5.3 Simulation results for the two-provider scenario

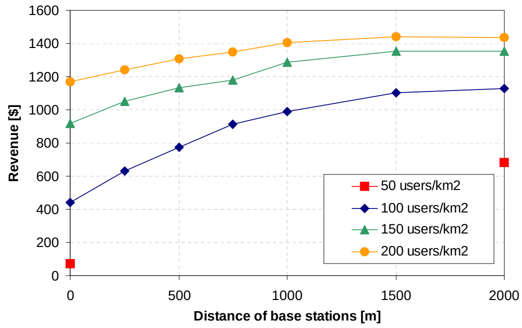
We now introduce a second provider, which, as the first provider, operates a single cell. All customers are given the ability to access both base stations to inquire for price information. We conduct multiple experiments in which vary the user density, the cell overlap, the cell sizes, and in which we introduce additional users with single access. We are mainly interested in the changes in prices but also analyse other output parameters such as revenue and blocking ratio.

#### Variation of the cell overlap with price as the only decision variable

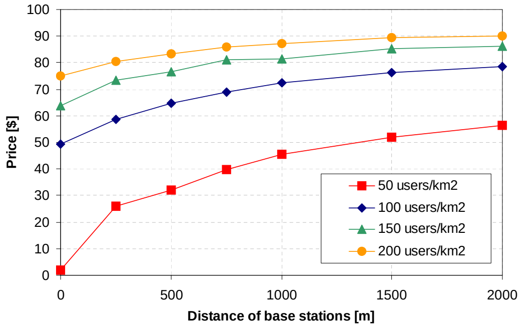
In the first experiment we let providers only choose prices and fix the cell radius to $Z_{max} = 1,000m$. We then vary the cell overlap by changing the position of the base station



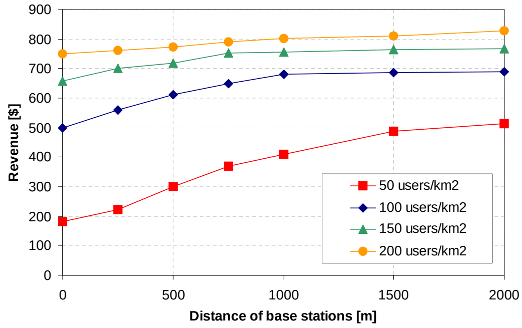

(a) Prices against the distance of the two base stations for service class 2.




(b) Revenue per provider against the distance of the two base stations for service class 2.




(c) Prices against the distance of the two base stations for service class 3.




(d) Revenue per provider against the distance of the two base stations for service class 3.

**Figure 5.23:** Prices and revenue per provider against the distance of the two base stations for service class 2 and 3.

of the second provider from $0m$ to $2,000m$ relative to the position of the base station of the first provider.

Figures 5.23(a) and 5.23(c) show the selected prices against the distance of the two cells for user densities ranging from 50-200 users/$km^2$ for service class 2 and service class 3, respectively. With an increasing cell overlap, the selected prices decrease for all user densities. Figure 5.23(b) shows the obtained revenue against the cell distance.

An important observation of this first experiment is the instability of prices when the user density is reasonably low and cells only partly overlap. This behaviour is consistent with the analysis of the game of complete information, in which no equilibrium exists in the case of an unconstrained system. Figure 5.24 shows the selection of prices for both providers against time. Both providers continuously lower their prices up to the point at which it becomes more beneficial to switch to the monopolistic price and to only serve customers outside the overlapping cell area. Therefore, in Figure 5.23(a) we cannot show prices for the case of partly overlapping cells and for a user density smaller than 100 users/$km^2$. For service class 3 the demand is large enough to produce stable prices for the lowest user density of 50 customers/$km^2$.

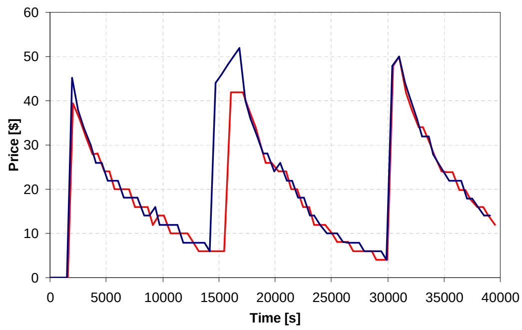



**Figure 5.24:** Prices of both providers against time in the case of a low user density (50 users/$km^2$) and partly overlapping cells (distance of base stations set to 500m). All results shown for service class 2.

### Variation of the cell overlap with price and cell radius as decision variables

We now allow both providers to maximise their revenue by varying their price $x$ and the active cell radius $z$. The visualisation of the results becomes more complex because we cannot show the outcome for different user densities in one graphic. To be able to visualise the results we show the results for selected user densities and service classes.

Figure 5.25 provides the results for service class 3 and a user density of 100 and 200 users/$km^2$, respectively. As with the previous experiments, prices increase with a decreasing cell overlap.

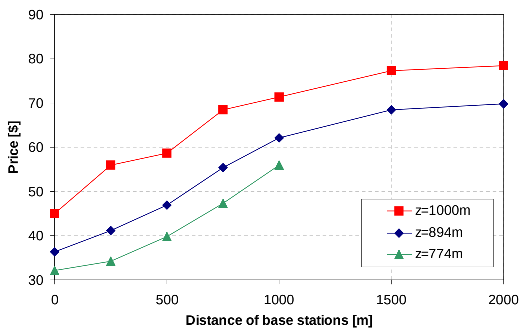
Figure 5.25(c) gives the average revenue for all four simulated user densities ranging from 50 to 200 users/$km^2$. As expected, revenue increases at the same rate as prices. The average revenue for a user density of 50 users/$km^2$ can only be generated for small cell overlap as prices become unstable for larger overlaps of the cells.

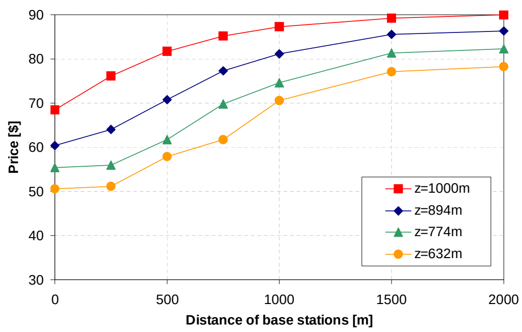
In Figure 5.25(d) the average service blocking ratio is shown against the distance of the base stations. With full cell overlap, the blocking ratio from a provider perspective is above 40%. This is because providers equally share customer requests and customers always request prices from both networks. Since providers have insufficient capacity to supply all customers alone, customers are rejected and count towards the blocking ratio. The blocking ratio decreases as the cell overlap becomes smaller and reaches about 8% in the monopolistic case. This demonstrates the ”conflict” of a provider in balancing between setting prices for the monopolistically served customers versus setting prices for the customers served under competition.

Figure 5.26(a) and 5.26(b) show the batch variance for a user density of 100 and 200 users/$km^2$, respectively.[16] While the variance with full overlap is relatively low it increases when cells only partially overlap. It again decreases with only small overlaps and reaches the minimum value in the monopolistic case.

The chapter appendix provides the same analysis for service class 2 (Figure 5.32).

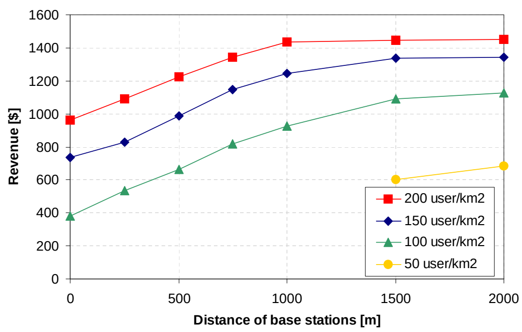
[16] As explained, the variance has been determined from the batch means of the five replications. Each batch summarises the simulation results for a one minute time window.

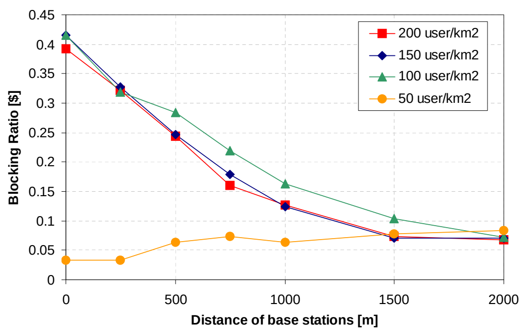



(a) Price/cell-radius selection against the distance of the base stations for a user density of 100 users/$km^2$.

(b) Price/cell-radius selection against the distance of the base stations for a user density of 200 users/$km^2$.




(c) Average revenue against the distance of the base stations for four user densities.

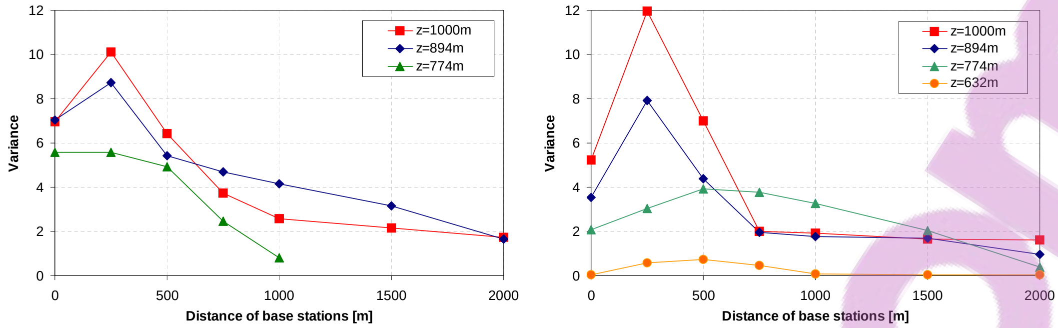
(d) Average blocking ratio against the distance of the base stations

**Figure 5.25:** Prices, average revenue and average blocking ratio for service class 3.

### Different cell sizes with full and partial overlap

In this experiment we vary the maximum cell size operated by the providers and vary the cell overlap. We set the maximum cell size of the first provider to $Z_{max1} = 500m$, while the second provider uses a maximum cell size of $Z_{max2} = 1,000m$. We use a constant user density of 200 users/$km^2$ and let customers request services of service class 3.

For the provider with the smaller maximum cell radius, the price remains constant at around 51 up to the point when the cell overlap is about 70% (corresponding to a base station distance of 800m). After this point the selected price increases to 67.41 for the monopolistic case. The picture looks different for provider 2 (Figure 5.27(b)). Since provider 2 can access more customers outside the coverage area of provider 1, it needs to set higher prices to not violate the resource constrained. While the selected price level for the largest cell radius remains relatively constant at around 86 and only increases to about 88, prices selected for smaller cell radii increase. Figure 5.27(c) shows the average revenue for both providers, which reveals that the cell movement does not have any influence on revenue. Since the demand is large enough, providers are able to balance the increasing "price pressure" by adapting the cell radius. Also, we can note that the revenue difference between the providers is small since the provider with the smaller maximum cell size can "make up" for this by decreasing prices. Note that this is only possible as long as




(a) Batch variance against the distance of the base stations for a user density of 100 users/$km^2$.

(b) Batch variance against the distance of the base stations for a user density of 200 users/$km^2$.

**Figure 5.26:** Batch variance against the distance of the base stations.

the system is constrained in resources. If user demand is lower, providers with smaller maximum cell radius will have significantly lower revenue.

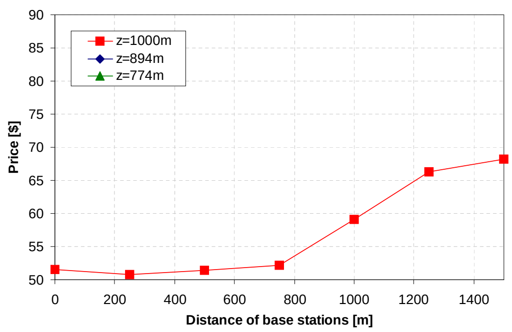
The result of the same experiment in which users select service class 2 is shown in Figure 5.33 in the appendix. The results are similar to the results of the presented scenario and we omit a detailed discussion.

### Introduction of single-access customers

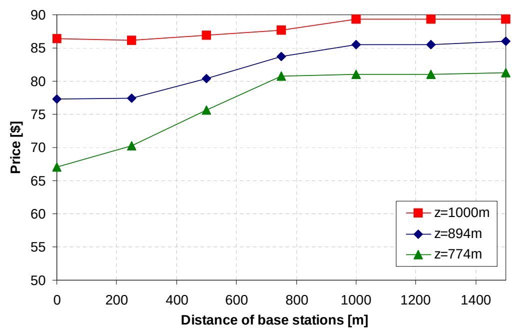
The objective of the following experiment is to understand how prices diverge when one provider measures significantly higher demand than the other provider. We can simulate this situation by introducing a second customer group, which has access to only one provider. We start with a scenario with a "base demand" created by users with a user density of 100 users/$km2$. Then, we gradually increase the density of single-access customers from 50 to 250 users/$km^2$.

We can observe in Figure 5.28(a) that prices diverge once the number of single-access customers increases. While the prices of provider 1 only increase slightly from 42.35 to about 54 for the largest cell radius, prices of provider 2 steadily increase with an increasing number of single-access customers. Since provider 1 cannot "see" the additional customer requests, it optimises its price so that revenue is maximised. In contrast, provider 2 needs to increase prices to not violate the resource constraints. In Figure 5.28(c), the average revenue per provider is depicted. With increasing user density of single-access users revenue increases for both providers, since both are able to gain more services at higher prices. After a user density of 100 users/$km^2$ the revenue increase stops for both providers.

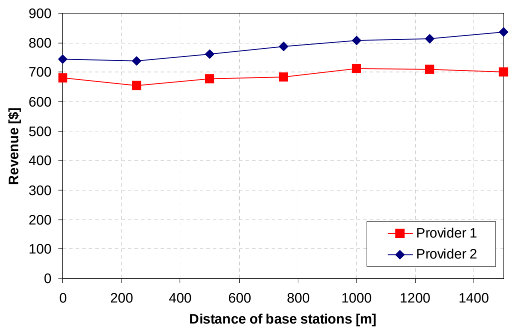
The identical experiment with users requesting service class 2 is given in Figure 5.34 in the chapter appendix. We can observe that in this setting the increasing user density of single-access customers has a positive influence on provider revenue of provider 2. While revenue also increases for provider 1 due to more multi-access customers deciding for the offer, the increase is smaller than for provider 2. We can also observe that once the density of single-access customers is large enough and prices in network 2 increase,




(a) Price/cell-radius combinations selected by the provider 1 against the distance of the base stations.




(b) Price/cell-radius combinations selected by the provider 2 against the distance of the base stations.




(c) Average Revenue for both providers.

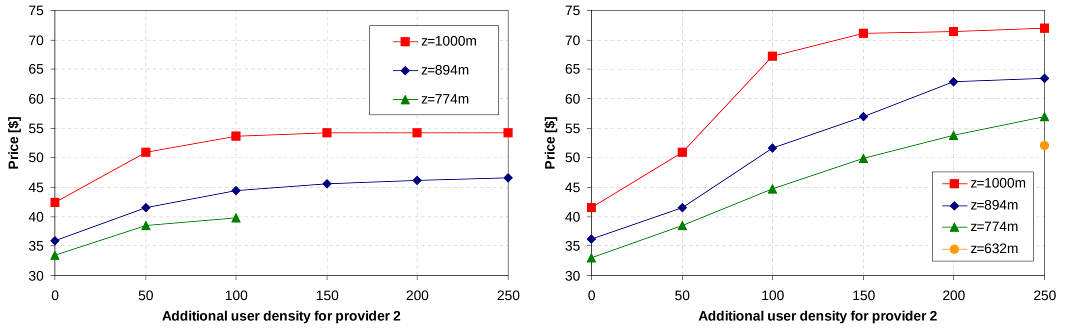
**Figure 5.27:** Prices and average revenue with different maximum cell radii and when shifting one base stations away from the other (service class 3).

provider 1 charges the monopolistic price of 50.

## Changing the cell size of one provider

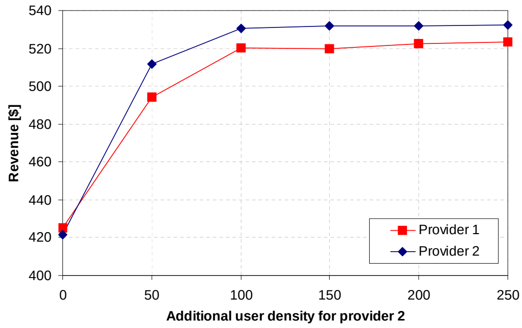
Instead of varying the distance between cells we now vary the cell size of one provider and observe the changes in prices and revenue. The setup is as follows. Provider 1 operates a cell with a fixed maximum cell radius of $Z_{max} = 750m$ while provider 2 changes its maximum cell radius from $Z_{max} = 500m$ to $Z_{max} = 2,000m$. The distance between the base stations is fixed at $d = 1,000m$. We have chosen a user density of 150 users/$km^2$ with users requesting service class 3.

Figure 5.29(a) shows the price/cell-radius combinations selected by both providers. Provider 1 selects prices only with the largest radius. Note that since we gradually change the cell size of provider 2 we cannot compare the prices directly. Instead, we show the price/cell-radius combinations with index 1-5, indicating the index of the radius, counting from the smallest to the largest radius. We can observe that while the first provider needs to lower its price because of the increasing competition, provider 2 has to increase prices and reduce its cell radius to cope with the increasing number of users. When both providers have the same maximum cell radius the prices match. With larger cell sizes prices diverge since provider 2 can increasingly source customers outside the overlapping




(a) Price/cell-radius combinations selected by provider 1 against additional single-access user density with connection to provider 2.

(b) Price/cell-radius combinations selected by provider 2, which experiences the additional demand from single-access customers.




(c) Average revenue per provider.

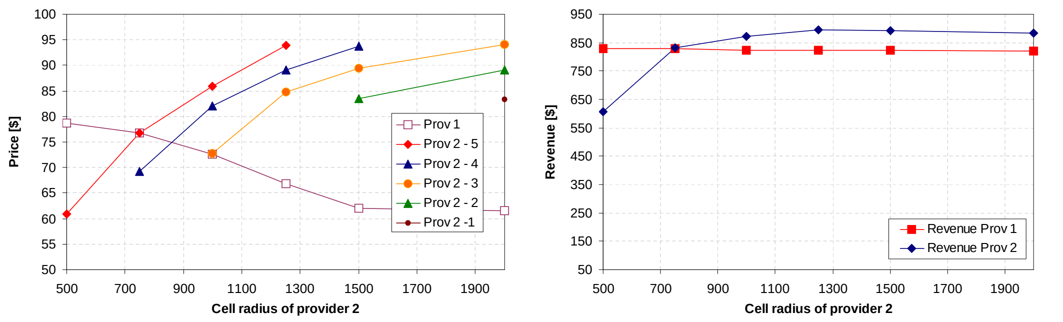
**Figure 5.28:** Prices and average revenue for two providers when introducing single-access customers (service class 3).

cell area.

In Figure 5.29(a) we show the average revenue obtained by both providers. While with a maximum cell radius below $750m$ revenue is significantly lower, cell radii larger than $1,000m$ seem to not benefit a provider to gain additional revenue. Even with more overall customers within the cell area, providers are unable to increase the number of active services as customers farther away from the base station require higher transmission power. We can also say that the limited grid size of the radius selection seems to make it impossible with larger maximum cell radii to set the optimal price/cell-radius combination.

## ▸ 5.6 Chapter Summary

We have presented an admission-based pricing approach for a wireless multi-provider setting, which allows providers to set prices at the time of the customer request. In this scenario providers set prices centrally by considering the customer demand and the competitive situation, which is given by the degree of cell overlap as well as the technological constraints present in both cells. Our focus was on understanding the provider's problem for revenue maximisation when it is faced with competition from




(a) Price/cell-radius combinations for provider 1 and 2 against the cell size of provider 2.

(b) Average revenue for provider 1 and 2 against the cell size of provider 2.

**Figure 5.29:** Prices and average revenue for two providers when varying the cell size of one provider (provider 2). User density of 150 users/$km^2$ and service type 3.

other providers partially or fully covering the same area.

Besides describing the general constrained maximisation problem a provider faces in order to optimise its revenue, we have modeled the situation of two providers as a game of complete and incomplete information. In the game of complete information the entire cell setup is common knowledge. In contrast, in the game of incomplete information, players are uncertain about the opponents position and are therefore unaware about the degree of cell overlap. By forming beliefs about the probability of the opponents position, a player can deduce the cell overlap and can calculate the expected influence of the opponent on his pricing decision.

For the game of complete information we could solve the constrained maximisation problem for a linear service activation function and when assuming a single rate constraint. We have illustrated the player's reaction function for different cases and have shown which prices form a Nash equilibrium of the game from which no player wants to unilaterally deviate. We could also identify the price at which provider revenue is maximised, and have shown that, under certain circumstances, the price is part of a Nash equilibrium.

For the game of incomplete information, in which the players are characterised by their absolute position and which, together with the providers own position, determines the degree of cell overlap, we could not identify an equilibrium in an explicit pricing function analytically. To gain an understanding about the price formation process and to derive an indication, if price/cell-radius combinations exist from which providers do not deviate in the described setting, we made use of the developed simulation environment and the approximation framework to identify near-optimal solutions to the constrained maximisation problem.

By conducting multiple simulation experiments with different setups we could gain an understanding about the formation of steady-state price/cell-radius combinations selected by the providers. By running multiple replications of the same experiment we could prove such combinations to be statistically significant in most experimental setups. In certain setups with low user demand we could also observe a periodic behaviour without reaching a steady-state. This behaviour could be matched with the findings in the game of complete information, in which, in the absence of any resource constraints, no non-zero revenue Nash equilibrium exists.

The results show that in the described setting, non-zero prices exist which allow providers to maximise revenue under the given competition. This also implies that in future wireless networks, in which resources are allocated on demand, multiple providers may coexist, even if access is fully transparent and customers can freely select the provider to join. The resulting price levels are "stable" and allow the provider to collect positive revenue from operating its network cell.

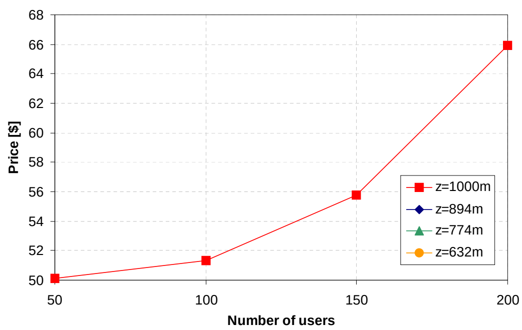
While many technical barriers would need to be overcome, we see a direct applicability of the developed concept for today's WLAN environments. With a small portable applet running an intelligent agent on the mobile client, it would be possible to allow customers access to residential or privately-owned WLAN networks and to use the developed pricing concept to charge customers for QoS services. The access points running the pricing engine can adapt pricing over time and can set up time-of-day profiles to react to the changing user demand. Additionally, access points can dynamically react to new network cells from competitive networks by adapting their price and range accordingly.

Several avenues for future research can be identified. First, the described model can be extended to support multiple service classes. While we have provided a preliminary discussion of two service classes for the monopolistic case in Section 5.2.3 and have implemented the required functionality in the simulation environment, we did not further pursue this research direction. In a multi-service class environment providers would be able to set preferences for certain service classes which yield the highest revenue per consumed resource unit. The preference would be expressed by the price set in each service class since wireless resources would first be used to fulfill the demand in the higher contributing classes.

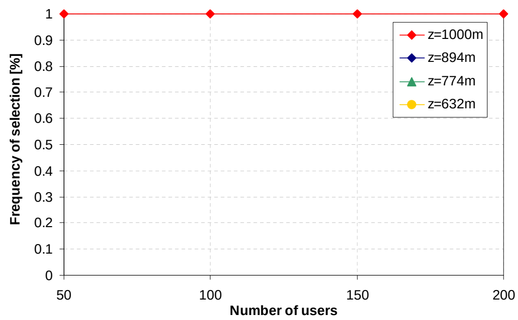
Second, the model needs to be extended to also capture the uplink direction and to include more than one wireless technology. By including different technological constraints and running similar experiments as we have done in Section 5.5, valuable information can be gained about the relative competitiveness of a technology when deployed in an existing, competitive environment.

Third, user mobility and handover functionality need to be integrated with the model to develop a general framework in the current context beyond 3G mobile access networks. This would require including mobility aspects in the admission pricing decision by, for example, projecting the path of movement when a mobile terminal moves away from the base station.

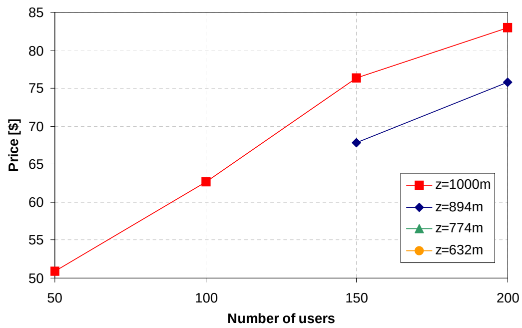
## ▶ 5.7 Chapter Appendix: Additional Simulation Results



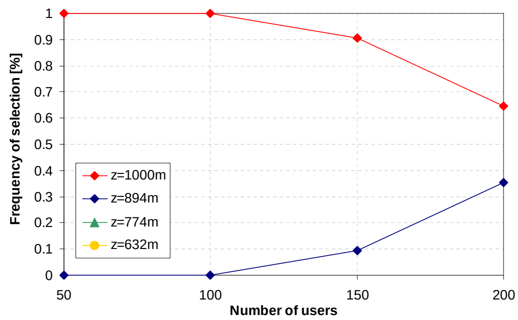

(a) Prices per cell-radius for service type 1 (32kbps).



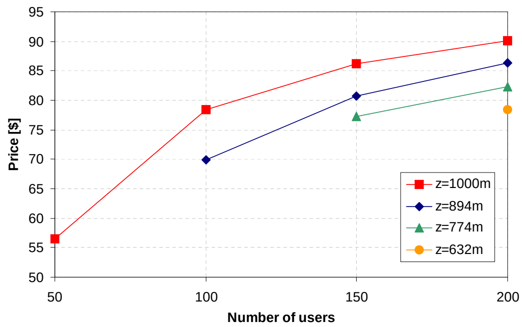

(b) Frequency of cell-radius selection for service type 1 (32kbps).



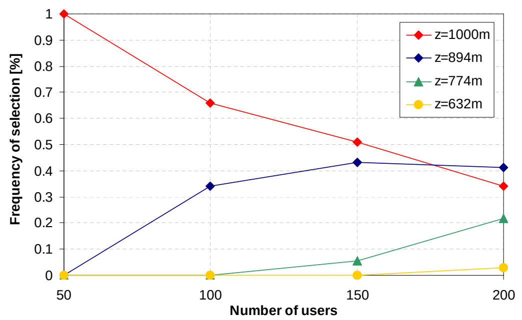

(c) Prices per cell-radius for service type 2 (64kbps).



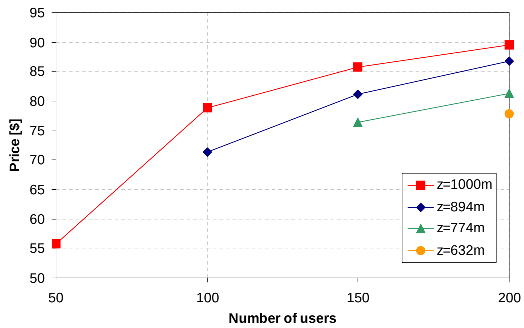

(d) Frequency of cell-radius selection for service type 2 (64kbps).



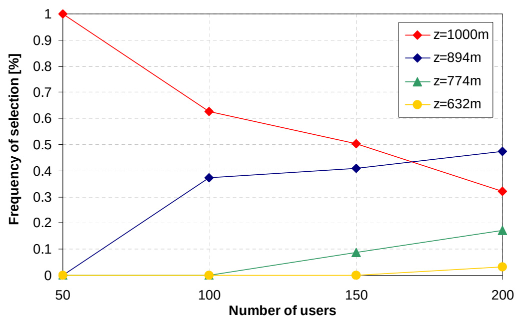

(e) Prices per cell-radius for service type 3 (144kbps).




(f) Frequency of cell-radius selection for service type 3 (144kbps).

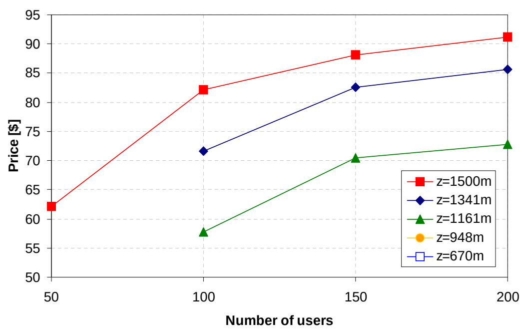



(g) Prices per cell-radius for service type 4 (256kbps).

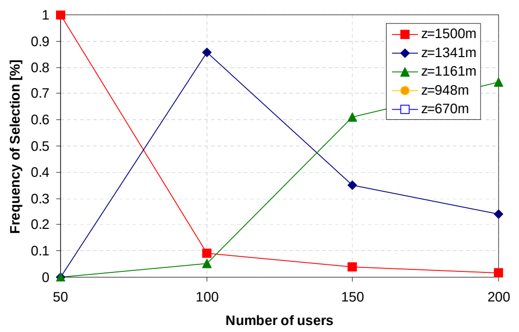



(h) Frequency of cell-radius selection for service type 4 (256kbps).

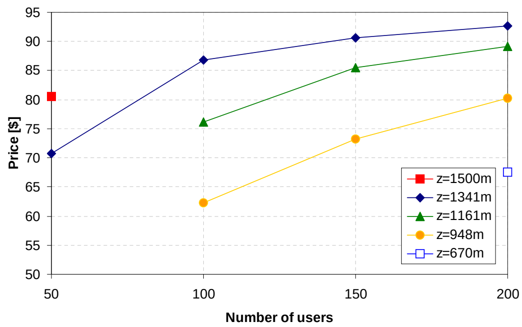
**Figure 5.30:** Prices and cell radii in a one-provider one-cell scenario for a maximum cell size of $Z_{max} = 1,000m$.



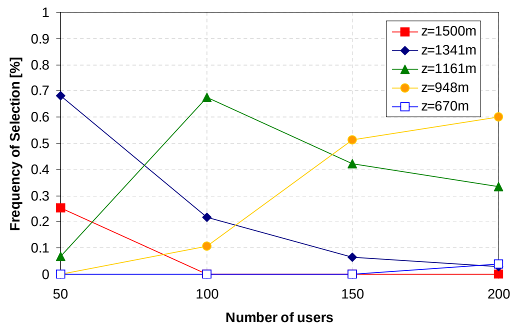

(a) Prices per cell-radius for service type 1 (32kbps).



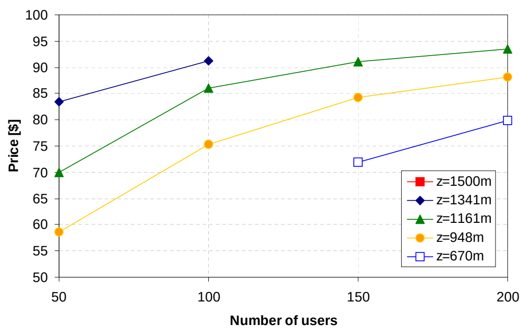

(b) Frequency of cell-radius selection for service type 1 (32kbps).



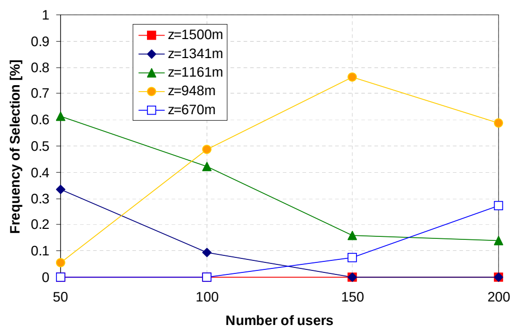

(c) Prices per cell-radius for service type 2 (64kbps).



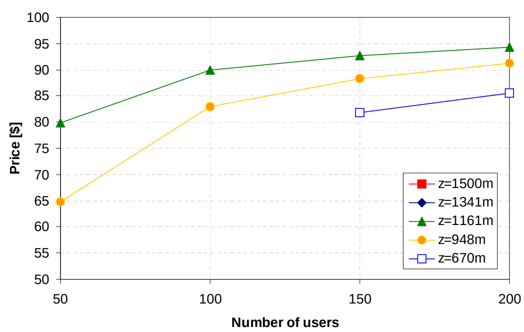

(d) Frequency of cell-radius selection for service type 2 (64kbps).



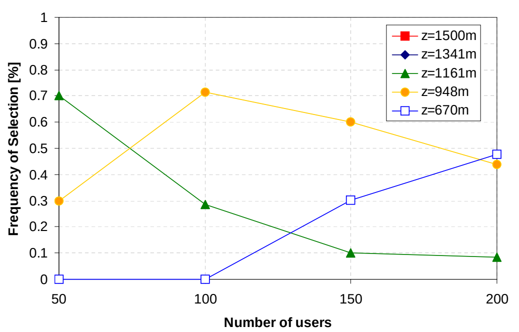

(e) Prices per cell-radius for service type 3 (144kbps).




(f) Frequency of cell-radius selection for service type 3 (144kbps).

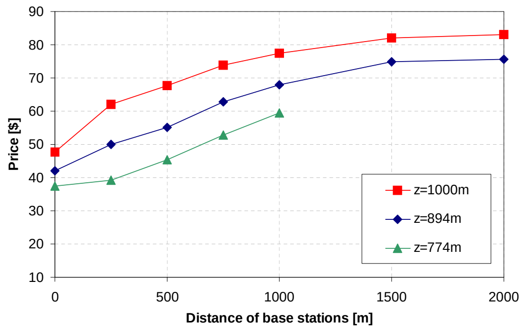



(g) Prices per cell-radius for service type 4 (256kbps).

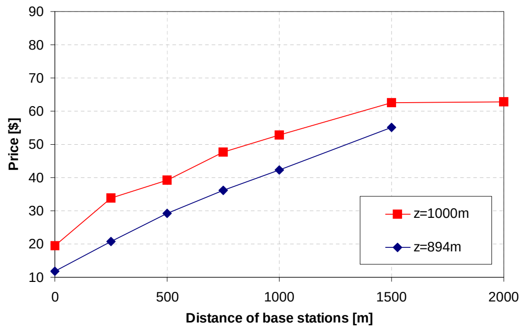



(h) Frequency of cell-radius selection for service type 4 (256kbps).

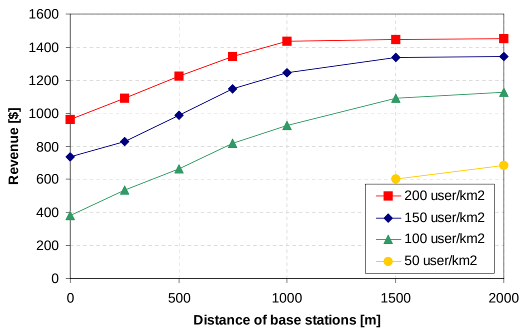
**Figure 5.31:** Prices and cell radii in a one-provider one-cell scenario for a maximum cell size of $Z_{max} = 1,500m$.



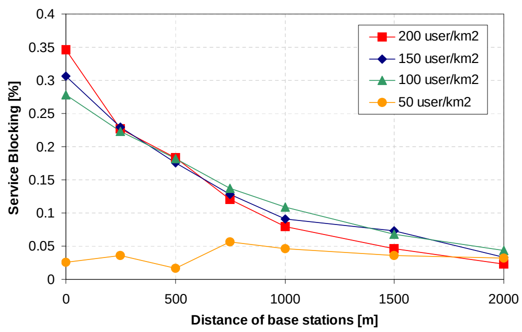

(a) Prices against cell distance for service type 2.




(b) Revenue per provider against cell distance for service type 2.

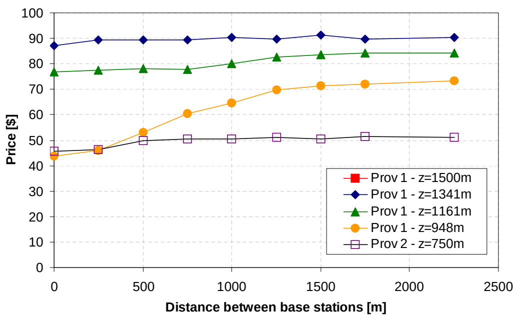



(c) Prices against cell distance for service type 3.

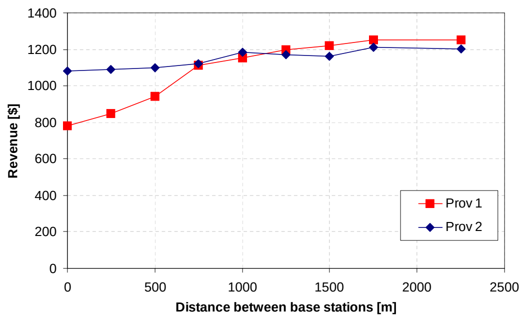



(d) Revenue per provider against cell distance for service type 3.

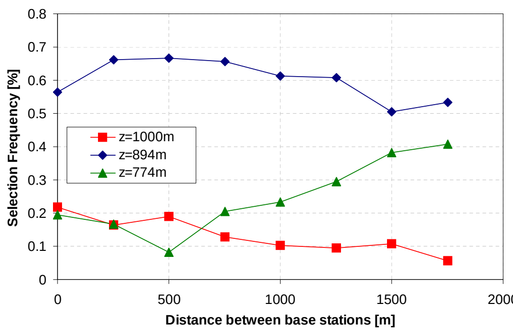
**Figure 5.32:** Prices and revenue per provider against cell distance.




(a) Price/cell-radius combinations selected by provider 1 and 2 against the distance between base stations.

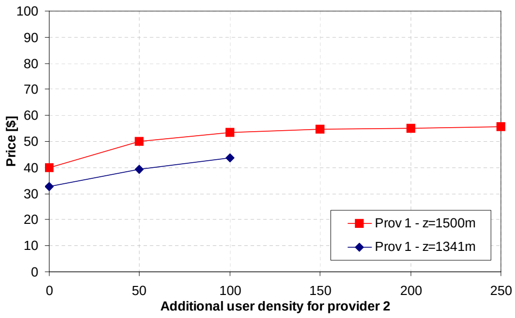



(b) Average revenue per provider against the distance between base stations.

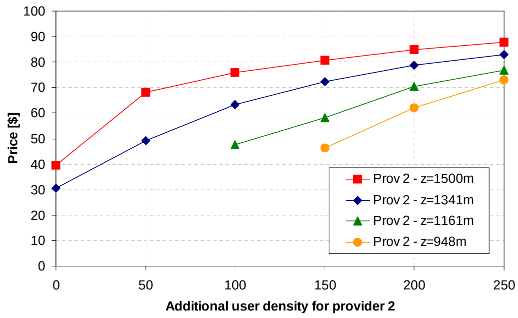



(c) Percentage of radius selection for provider 1 against the distance between base stations.

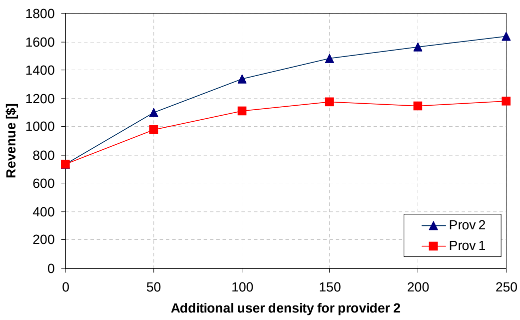
**Figure 5.33:** Prices and average revenue with different maximum cell radii (provider 1 z=1,500m, provider 2 z=750m) and when shifting base stations away from each other (service class 2).



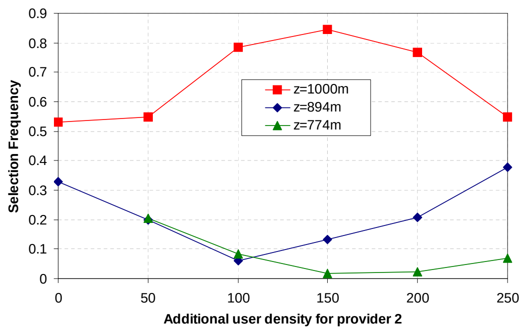

(a) Price/cell-radius combinations selected by provider 1 against additional single-access user density with connection to provider 2.




(b) Price/cell-radius combinations selected by provider 2, which experiences the additional demand from single-access customers.




(c) Average revenue per provider against the increasing number of single access customers for provider 2.




(d) Percentage of radius selection for provider 2 against the increasing number of single access customers for provider 2.

**Figure 5.34:** Prices, average revenue, and percentage of radius selection for two providers when introducing single-access customers (service class 2).

# Chapter 6

## The Simulation Architecture

### ▸ 6.1 Introduction

This chapter provides an overview of the simulation environment, which has been developed and extensively used in both research streams, PSPSim and AdSim, since the completion of the first software prototype in December 2004. Much of the intuition behind the bidding strategies and the optimisation algorithms was derived from experimenting with the simulation architecture. Part of the material presented in this chapter has been published in Roggendorf et al. (2006), a paper describing the general agent-based simulation approach for the dynamic pricing of wireless resources in a competitive setting.

Since the development of the experiment environment is a major accomplishment, we provide a detailed description of the simulation platform and the structures developed for simulating dynamic pricing. The material presented here is not explicitly required for the understanding of the models and simulation results from the previous two chapters.

#### ▸ 6.1.1 General objectives of the simulation platform

The simulation environment has been developed with the main objective to create a general framework for dynamic pricing in wireless networks. Instead of designing a specific tool only for the purpose of the particular simulation experiments, we aimed at developing a basic architecture of micro-entities together with a generic ontology, a language between such micro-entities. In particular, the following objectives were central for the platform selection and the architecture design:

- Entities are required to act independently according to predefined behaviour and decision rules. Therefore, each entity needs to be able to implement individual algorithms to model its internal structures, as well as to model its communication behaviour with other entities or the outside environment.
- We also aim at developing an environment which is mostly independent from the specific market institution to be implemented. This means that the ontology needs to allow for a wide range of communication work flows without already defining the specifics of the content being exchanged.
- An entity model is required which allows for the easy extension of behaviour and functional blocks within the entities. A modular approach, in which entities can inherit functionality from generic entities, was an important requirement throughout the specification and implementation of the generic architecture.
- To not be restricted in the size of simulation experiments, an objective for the platform implementation is to create an environment which can be distributed on multiple machines. For example, it should be possible to move certain entities with intense computational workload to separate machines to allow for load distribution.
- To be able to use the platform as a base for subsequent research projects which can build on the general functionality, it is important that the underlying platform is continuously maintained and updated.

### ▶ 6.1.2 Delimitation from technical network simulations

The simulation environment to be created differs in some substantial points from the engineering perspective of using simulation in communication networks. In the following we contrast the main differences of both approaches.

Network simulation on the technical level is usually concerned with measuring some sort of technical efficiency measure such as maximum throughput, packet loss, or signal delay. Engineers are interested in how a modified protocol stack performs in a complex setting when implemented in all entities of the network. All network entities are expected to act and react in the same way as given by the protocol implementation. Entities, such as mobile devices or base stations, are described purely by technical parameters such as the maximum transmission power, the frequency bands used for the transmission, or the chip rate of a wireless network cell. To be able to collect meaningful data, a technical network simulator needs to imitate the network operation as realistic as possible.

In contrast, the environment for simulating dynamic pricing in a wireless network has its focus on the economic aspects of the network operations. In this sense, factors such as user satisfaction and revenue contribution of the resources allocated become important output variables of the simulation. This also means that additional parameters need to be considered as input parameters. For example, to model user satisfaction, each entity representing a user needs to implement a utility description, which can be different for each entity in the system.

Another difference between the purely technical view and the economic view is the level of abstraction from the many factors underlying wireless network traffic. While from a technical view it is desirable to implement as much detail as possible, to model resource allocation from an economical viewpoint, the simulation needs to be able to abstract only a few aspects. For example, for testing new MAC protocols it is essential for the experimenter to have access to data on the IP level to reconstruct the observed behaviour. In contrast, for experiments from the economic viewpoint, it may be sufficient to collect data on higher levels and disregard the specifics of the underlying transport protocol.

The above description does not automatically mean that available simulation tools, which have been created with the objective of technical network simulation, are not suitable for our task. However, they need to allow for a modular extension beyond the technical description of entities and for building additional logic above the protocol stack.

### ▸ 6.1.3 Special requirements of simulating an environment of wireless networks

Developing a simulation environment for wireless access networks created additional requirements for the general architecture. In contrast to a wired network, in which access options of each entity are given by the available physical connections, in a wireless network several factors influence the possibility of entities to access a wireless link. These factors, such as the location of the user, the user's end terminal, and the contractual relationships with wireless network providers needed to be modelled in the simulation environment. Additionally, entities may not be stationary, but may be mobile and change their position during the simulation experiment.

Another complex field to be modelled in the simulation environment was the management of wireless resources at these entities acting as the base stations in the simulation. Since we wanted to keep the general implementation as open as possible, resource constraints modelled by the particular implementation needed to be modular and extendable. In the simplest case, it is possible to model resource allocation with a simple rate constraint and to abstract from the particularities of the wireless channel. However, the simulation environment should also be able to handle more complex resource management such as the modelling of user interference and power constraints in a WCDMA network.

### ▸ 6.1.4 Chapter outline

This chapter is structured into five sections. The following section briefly explains the selection process for the agent platform used for the implementation of the simulation environment for both research streams. The second section provides an overview of the main concepts and characteristics of agent-based platforms, and agent-based simulation in particular. Then, we present the JADE agent platform and describe its main features. Section 5 is dedicated to the architecture and implementation of the simulation environment developed within this research. We first describe the generic architecture, which has been the base for both simulation platforms. Then, we provide an overview of the extensions developed for the flow-based PSPSim environment and the admission-based AdSim environment. Section 6 summarises the main points of the chapter.

## ▶ 6.2 Selection of the simulation platform

For the final selection of the simulation approach an analysis of the existing simulation tools and platforms was carried out to minimise the development effort needed to conduct the experiments but at the same time develop a general simulation platform for future research efforts. In this context three main groups of simulation platforms can be distinguished: mathematical software packages such as *Mathematica*, *MatLab* or *Maple* (Chonacky and Winch, 2005), network simulation environments such as *ns2* or *SSFNet* (Tyan and Hou, 2002), and agent-based simulation tools such as *JACK*, *ZEUS*, or *JADE* (Luck et al., 2004). In principle, all platforms are able to support the basic requirement of modelling individual preferences and strategies. However, the tools in the three categories have distinct advantages and disadvantages which needed to be evaluated.

The platforms of the first category, the mathematical packages, were seen as too inflexible in order to build a generic simulation tool for dynamic resource allocation. While many extensions to such software packages exist and they are used for simulations in a wide range of applications, they have not been originally developed for this purpose. Also, the focus of such packages is usually on solving complex mathematical models, which does not apply directly to our research domain. Additionally, the investment for buying such packages and the applicable add-ons is significant and only limited testing was possible to support the decision-making process.

The second category of software tools was evaluated as not fully suitable because its focus lies on the performance of the lower network layers such as the medium-access-control layer (MAC) or the Internet Protocol (IP) layer. The common simulation approach with network simulation environments is to generate realistic user data and to test alternative protocol implementations under various load scenarios and with different traffic types. Our research, which in contrast focuses on the negotiation process between entities, does not require the generation of real traffic data. Furthermore, it needs to abstract from the complexity of network traffic characteristics. Even if such tools allow for a flexible implementation of higher level protocols and the use of basic agent-based programming concepts, the functionality provided by the platform itself was seen as not sufficiently contributing to the task to be accomplished.

The third group, agent-based simulation environments, gave the best match with the requirements of an open and flexible architecture for simulating individual user behaviour in a decentralised setting. The platforms evaluated in this category all provide a significant proportion of basic building blocks required to support an operational agent-based system. Also, many toolkits come with graphical support tools for the design and the testing of the application at run-time.

After creating a short list of all software platforms with the best match to our research problem, JADE was selected because of its active user community, the implemented standards for the agent-based communication process and the flexibility it provides in terms of agent mobility and code portability.

## ▶ 6.3 Multi agent systems

In this section we give a brief introduction into what a rational agent is, the definition of multi agent systems and the key concepts centered around the MAS approach.

### ▶ 6.3.1 The rational agent paradigm

No overall accepted definition for the term *agent* exists in the literature. The term has been used by many disciplines in very different ways. In a broad sense, an agent can be defined as *"anything that can be viewed as perceiving its environment through sensors and acting upon that environment through actuators"* (Russell and Norvig, 2003). More precisely, an agent is *"an autonomous entity with an ontological commitment and agenda of its own"* (Harris, 1990, p. 3). The term, originally used in philosophy, has been adapted in distributed artificial intelligence (DAI) to describe *"computational entities that are capable of exhibiting flexible behavior in dynamic and unpredictable environments"* (Luck et al., 2004, p. 4). In this respect an agent can be seen as some kind of high-level software abstraction. The agent concept provides a convenient and powerful way to describe a complex software entity which is capable of acting exactingly in order to accomplish tasks on behalf of its user (Woolridge, 2002). From a conceptual view, the agents' paradigm applies concepts from artificial intelligence and speech-act theory to the distributed object technology (Bellifemine et al., 2003).

Four basic properties of agents which are regarded as necessary and sufficient for "agenthood" are defined as (Wooldridge and Jennings, 1995):

- *Autonomy:* functions without central control or direction.
- *Reactiveness:* have the ability to monitor their environment and to react to changes in the environment.
- *Proactiveness:* have an overarching goal that directs short-term and long-term behaviour.
- *Social ability:* have the ability to communicate and interact with other agents and entities inside and/or outside the system.

These characteristics are broadly accepted by many researchers as they capture all characteristics of most agent-based systems. However, additional aspects such as cognitive abilities or adaptive behaviour may add to the dimensions of agents depending on the environment and the specific implementation.

An agent is called *rational* if it always selects an action that optimises a certain performance measure given the current knowledge of the agent (Vlassis, 2003). The performance measure is typically predefined by the designer of the system. A rational agent is often also called *intelligent* because of its ability to autonomously reach a goal.

Assuming that an agent acts within a certain time grid $t = 1, 2, 3...$ it has to select an action $a_t$ from a set of possible actions $A$ for each discrete time step.[1] To let an agent act in its environment it needs to observe past events and use the history of such events to choose its action. If an agent's observation of each time stamp $t$ is $o_t$ and its past actions is described with $a_t$, the function mapping the past observations to the agents action can be written as

$$\pi(o_1, a_1, o_2, a_2, ...o_t) = a_t$$

If the function maps the entire history of the observation-action pairs to decide on the action $a_t$ at the time $t$, it is also called the *policy* of an agent (Vlassis, 2003).

Implementing this function can often be very difficult for various reasons. First, the number of historic observation-action pairs can be very large and may require a huge memory. Second, the computational complexity of the policy function may be very high, especially with a large number of past events. Consequently, no solution may be found in acceptable time.

An agent may therefore reduce the complexity of its policy to a smaller number of past observations. For example, an agent can only take the observation from the last round into account. Such an agent, which ignores the past and only bases its decision on the last perception is often called a *reflex agent* (Russell and Norvig, 2003). The policy implementing this behaviour, $\pi(o_t) = a_t$ is called memoryless, reactive, or myopic. Even with such a simple policy, reactive agents can be very successful in reaching their goal. In contrast, agents that maintain an internal state, and use the historic data to predict the effects of actions are called *deliberate agents*.

### ► 6.3.2 The agent's world

As described, agents perceive the environment in which they live and act with their *observations*. This environment of an agent is often called the *agent's world*. The collective information that is contained in the world at any time step $t$ is called a state $s_t$ and is part of all possible states of the world $S$. A world can either be discrete or continuous depending wether the number of states are finite or infinite. If an agent is able to observe all information of the current state the world is *fully observable* for this agent ($o_t = s_t$).

[1] We assume a discrete model in contrast to a continuous model in which actions have to be taken at no predefined time steps.

Depending on the complexity of a particular world this assumption may be more or less realistic. In many problems the state at time $t$ may contain a complete description of the entire world's history before time $t$. In such cases, a reflex agent implementing a memoryless policy can easily find the optimal action. A state that succeeds in retaining all relevant information is said to be *Markov*, or to have *the Markov property* (Sutton and Barto, 1998).

In many situations, however, agents can only oversee a certain part of the state of the world while other parts of the state is unobservable for them. The situation of a *partially observable* state of the world can stem from two sources. First, an agent's capabilities or perception framework may not allow it to observe certain details. For example, if a robot agent does not have a temperature sensor it cannot measure the current room temperature. Second, the full state may just not be observable. For example, if we think of an agent playing blackjack, even observing every detail of the last round of the game will not tell the agent the value of the next card in the deck.

In cases of partial observability an agent has to make additional assumptions about the true state of the world. A conditional probability function can be used by an agent to assign a probability to each possible state based on the partial information available ($P(s_t|o_t)$). $s_t$ is treated as a random variable that can take all states in $S$ with $0 < P(s_t|o_t) < 1$ and $\sum_{s_t \in S} P(s_t|o_t) = 1$.

When an agent chooses an action, the world changes as a result of its action. If the world is deterministic the agent knows already the new state of the world $s_{t+1}$ after the execution of the action. In contrast, in a stochastic world, the next state is unknown and can only be described with a probability function $P(s_{t+1}|s_t, a_t)$. Again, $s_{t+1}$ can take all states in $S$. The stochastic model clearly introduces more complexity in optimal decision-making but usually improves the model to better reflect a situation in the real world (Vlassis, 2003).

### ▶ 6.3.3 Characteristics of a multi agent system

In many situations the value-add of employing an agent-based methodology comes from the coexistence and interaction of multiple agents. Such a system that consists of a group of agents that can potentially interact with each other is called a *Multi Agent System* (MAS). The study of MAS focuses on systems in which many intelligent agents interact with each other. The type of software agents used in MAS are usually using rich cognitive models and sophisticated communication languages and interaction mechanisms (Davidsson, 2002). Figure 6.1 shows a general multi agent scenario and the internal properties of the single agents in the system.

MAS have emerged as a sub-discipline of *Distributed Artificial Intelligence* (DAI), which is concerned with systems that consist of multiple independent entities that interact in a defined enivironment (Stone and Veloso, 2000).[2]

[2]Another sub-discipline of DAI is *Distributed Problem Solving* (DPS), which is concerned with information

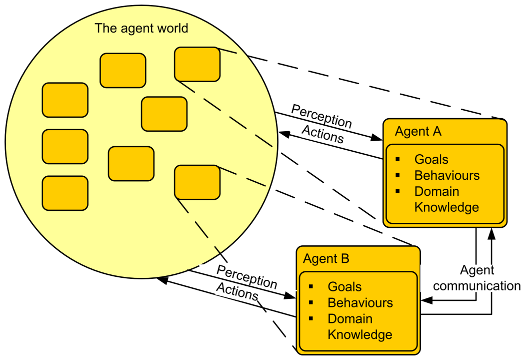



**Figure 6.1:** The fully general multi agent scenario. Adapted from Stone and Veloso (2000)

MAS have some distinct characteristics in contrast to a single agent approach (Vlassis, 2003). The design of agents in a MAS is usually heterogeneous. Agents may be implemented based on different programming languages or run on different hardware platforms. Also, agents may differ in their internal preferences for resources or goal settings. While in the single-agent approach control is centralised, in MAS it is distributed (decentralised) over all agents in the system. While this feature has many advantages, such as asynchronous computation and speedups in certain situation, it also introduces additional challenges. For example, a coordination mechanism is needed to ensure a good macro result of the joint decision.

In a MAS the knowledge about the state of the world can differ from agent to agent. For example, an agent observing the actions of another agent is usually unaware of its action levels and its current perceptions and may also not be able to infer the other agent's future plans. One particular concept used in MAS is that of common knowledge (Vlassis, 2003). Common knowledge is knowledge every agent has and every agent knows that the other agent knows, and so on.

The environment in which agents operate can either be static or dynamic. A static environment is time invariant and stable over the lifetime of the agents. Single agent systems often assume a static environment to simplify the design and the computational complexity. In MAS, the environment is often dynamic because it is influenced by the actions of each individual agent in the system.

Agents in a MAS can either act in a cooperative or competitive way. In a cooperative system agents share information and coordinate their actions to come to the best solution for their group as a whole. In contrast, competitive agents compete against each other and have the ability to negotiate once a conflict occurs. Conflicts can stem from limited resources in the system but can also be more complex in nature such as discrepancies between agents in terms of the level of expertise. In a competitive environment, the implemented communication mechanisms can drive individual agents to certain actions or decision patterns, which may be the objective of the designer of the system.

management aspects of systems with several branches working together toward a common goal.

### ► 6.3.4 Principles of agent communication

One central aspect of interaction in MAS is *communication* between agents. Agents may need to communicate to, for example, negotiate for resources, offer services, or exchange capabilities. According to Bellifemine et al. (2003), three distinctive features of the communicational model in MAS are:

- Agents are active entities with the ability to say "no". Their communication is message-based and asynchronous. Instead of remote-procedure-calls in the object-oriented programming paradigm agents can simply send a message to other agents.
- Communication in agent-based environments becomes a special form of an action. Internal actions and external communication are handled on the same level. Effects and preconditions of each communicational act need to be clearly defined to make communication plannable for an agent.
- Communication carries a semantics meaning, which needs to be mutually understood by the group of agents it is in communication with.

Agent communication involves different layers of abstraction (Vlassis, 2003). On the lowest layer the messages exchanged between agents need to be transported safely and in a timely fashion. The functionality needed on this level is usually provided by the agent simulation environment using some lower-level protocol (run-time, SMTP, TCP/IP, IIOP, HTTP).

On the next layer of abstraction a common language is required to let the agents understand each other in a systematic way. Many different *Agent Communication Languages* (ACL) have emerged. An ACL provides agents with a means to exchange information and knowledge (Labrou et al., 1999). The most prominent and visible agent communication languages are FIPA-ACL and KQML (Luck et al., 2004).

On the highest layer, the application layer, communication enables agents to solve standard problems like coordination and negotiation. The concept enabling such complex communication patterns is called an *ontology*. In the following subsection we further elaborate on the concepts behind ACLs and ontologies.

### ► 6.3.5 Agent communication languages and ontologies

Agent communication is characterised by a mutually understood *Agent Communication Language* (ACL) for exchanging messages between agents. The ACL language provides language primitives that implement the agent communication model. ACLs can be described as wrapper languages since they are usually unaware of the choice of content language and ontology specification mechanism but focus on the encapsulation of the content, pragmatics, and other constituents of any communication requirement between two agents (Staab and Studer, 2004). An example of the structure of an ACL message is given in Figure 6.2. An ACL is more than the Remote Procedure Call (RPC) used in

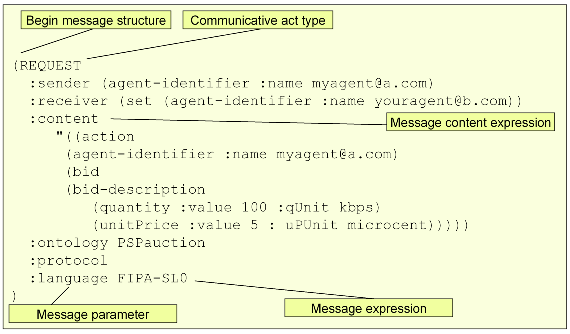



**Figure 6.2:** ACL example message using the FIPA ACL standard.

object-oriented languages because it can handle propositions, rules and actions instead of simple objects. The retrieval of a message can invoke very different consequences in different agents depending on its internal state and behavioural implementation. Traditionally, ACL messages are understood as *speech acts* or *communication acts* which are usually accounted for in terms of *beliefs*, *desires*, *intentions* (BDI). Each speech act carries a communicative primitive as a single action that updates and alters the knowledge of an agent (Vlassis, 2003). Standard types of speech acts are INFORM, QUERY, COMMIT, CONFIRM, and AGREE.

The content of the messages needs to be translated into the language of the agent in order to process the message and to derive actions from the message (Noy and McGuinness, 2001). Therefore, a common vocabulary is needed to build up a mutual understanding about the elements and attributes exchanged between agents. This vocabulary is commonly called an *ontology*. The term, which has a long history in philosophy, where it is defined as a *systematic account of Existence*, is used differently in Artificial Intelligence and agent-based systems. In short, ontologies in the domain of agent-based communication can be defined as a *"specification of a conceptualisation"* (Gruber, 1993). An ontology is therefore a description of concepts and relationships that can exist for a community of agents. It makes use of a conceptualisation, which can be described as an abstract, simplified view of the world. An ontology in this sense never describes a domain exhaustively but focuses on the relevant elements needed for the agent communication (Cranefield and Purvis, 2001).

### ► 6.3.6 Application areas for multi agent systems

Over the last decade, MAS have been applied in many disciplines ranging from robotics over computational economics to applications in medicine and sociology. Rapid advancements in computational power allows the design of increasingly complex environments, which considerably widens the application horizon of MAS. With an increasing maturity of agent-based architectures, more and more research areas are starting to explore the advantages of agent-based systems. Wood and DeLoach (2001) distinguishes four different

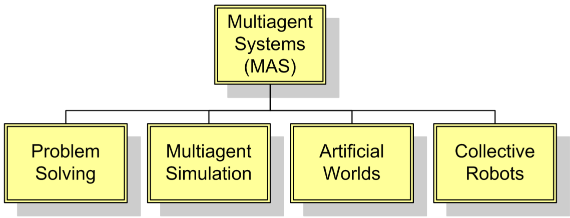



**Figure 6.3:** MAS application areas. Adapted from Wood and DeLoach (2001)

categories of MAS applications (Figure 6.3).

The first category subsumes all MAS applications concerned with *distributed problem solving*, which is also the field in which MAS was first used. The domain of distribution can differ from application to application. One area is the distribution of knowledge over several entities, each with complementary expertise and without central coordination. Agent-based systems can provide a way to let these experts exchange information and commonly solve a problem.

The second category of MAS applications are *multi agent-based simulations* (MABS). While mainly influenced by MAS, MABS is a inter-disciplinary field also influenced by the agent-based social simulation community (ABSS) (Ferber, 1999). MABS has been applied by many disciplines such as biology, sociology, and economics. While in ABSS the focus is on simulating and synthesizing social behaviour to better understand real social systems with properties of self-organization, scalability, robustness and openness, the focus of MAS is on the solution of hard engineering problems related to the construction, deployment and efficient operation of agent-based systems.

Another application area of MAS is the design of *artificial worlds*. They can be used to describe specific interaction mechanisms and analyse their impact at a global level in the system (Ferber, 1999). The design of the agents in such synthetic universes is often inspired by human or animal behaviour. Applications in this area can demonstrate the development of a population over time. Another aim is often to create societies of agents that are very flexible and can adapt even in cases of individual failure.

The last main category of MAS applications deals with the design of *collective robots*. In an agent-based robot system each subsystem has a specific goal. By solving each individual goal the common task set on the macro-level is accomplished as well.

One of the most challenging environments for multi agent systems is the Internet (Nwana and Ndumu, 1999). While a common transportation protocol exists (the Internet Protocol (IP)), the information and services available on the higher layers form a complex and heterogeneous environment. The main challenge is to let agents understand different languages and semantics to, for example, negotiate for resources, make purchase decisions based on several offers or retrieve information on behalf of a user. Auctions on the Internet and electronic commerce are examples for current and future uses of autonomous agents to relieve users from manual tasks of, for example, comparing competing offers.

### ▸ 6.3.7 Multi agent systems as basis for simulation

Simulation as a research method has a long history in many academic and professional areas. For example, simulation is often used in physics, electrical engineering, robotics, economics, logistics and production planning.

As described in the last section, MABS has evolved as one subfield of MAS as a way of going beyond the limitations of traditional approaches. One key difference to alternative simulation systems is that agents introduce heterogeneity on a micro-level rather than homogeneity of behaviour for all entities of the system (Davidsson, 2002). The agent-based approach can be described as a bottom-up approach of modelling each individual entity and to derive macro-results by summarising individual achievements. The approach has provided new insights in the connection between the micro-level and the macro-view and how the individual behaviour of an entity influences the overall dynamics of systems.

In economics, the MABS approach finds its roots in agent-based computational economics (ACE), which studies *"economic problems modelled as dynamic systems of interacting agents"* (Tesfatsion, 2006). Examples of possible agent types are individuals, social groups, institutions, or physical and biological entities. The designer of the system specifies the initial knowledge of the agents, their behavioural methods, and the degree of observability by each agent.

In ACE four different aspects of research objectives can be distinguished (Tesfatsion, 2006). One predominant objective is to develop an *empirical understanding*. Important questions addressed are why particular regularities have evolved or why a system has found a stable equilibrium at a certain point. Agents are designed in a way that they closely map the real world entities.

A second objective is that of *normative understanding* to identify a good economic design from simulation experiments. Research with the objective is interested in evaluating new economic designs in terms of its performance and stability over time. Important aspects of the analysis are if the outcome is efficient, fair and well-behaved despite the attempts of agents to gain advantage by strategic behaviour. For example, introducing a new tax into an economic system could be simulated in a MABS system. The results for the society as a whole can be analysed and parameters can be varied to understand the sensibility on the overall result. The third objective of ACE is to gather *qualitative insight* and *generate new theories* from the results of the simulation experiments. This objective is primarily concerned with improving the understanding of economic models through a systematic analysis of its behaviour in different situations and alternative initial conditions. One central intention is to better understand the self-organising capabilities on a micro-level without central control. An ACE model based on key properties of the market (limited information of agents, different production functions, buyer preferences) and privately motivated buyer agents with learning capabilities can help to grasp the working of such a complex system in a laboratory setting.

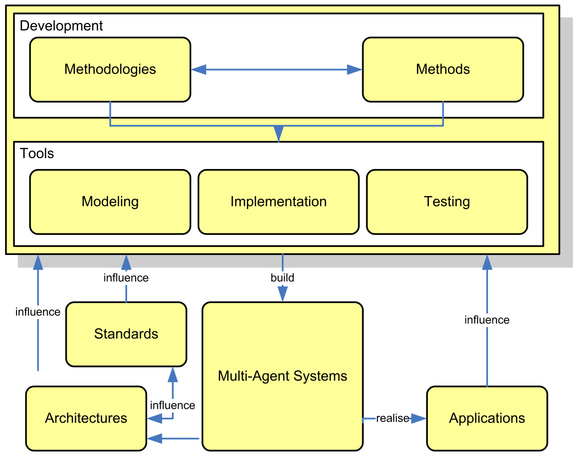



**Figure 6.4:** Overview of Multi Agent System (MAS) research elements. Source: JADEX Research Group (2005)

Finally, ACE can be used for *methodological advancement*. Models in ACE can only be as good as the underlying assumptions about the structural, institutional and behavioural characteristics of the economic systems and the entities within. In order to produce valid and compelling results from agent-based simulation experiments, researchers need methodologies and tools to generate rich models, build a solid experimental design and to validate their models against real-world data. To meet these requirements many different techniques have been developed ranging from careful consideration of methodological principles to the development of validation tools to compare simulation results with data collected in empirical research.

## ▸ 6.3.8 The MAS research framework

A considerable number of agent-oriented methodologies, architectures and tools are available today (Luck et al., 2004). When developing a specific MAS application one has to decide on the appropriate framework to use in a particular context. Figure 6.4 provides an overview about the different elements of research connected with the agent-based research domain.

*Methodologies and methods* guide and support a structured specification and development process of an agent-based application. A software methodology usually consists of a *modelling language*, which is used for the description of the models, and for defining the elements of a model together with a specific syntax, and a *software process*, which defines the development activities and the interrelationship between the activities. Most concepts and approaches supporting the software engineering of agent-based systems were inspired by models originally developed in the context of object-oriented programming. Many frameworks extend object-oriented techniques, such as the *Unified Modelling Language* (UML) and build agent-specific extensions on top of these approaches. Similar to UML, which defines nine different modelling diagrams to support the object-oriented software development process, an agent-oriented modelling language should ideally provide different diagram types for the different phases and views of an agent-based system. However, the underlying concepts of agent-based systems are often richer and more complex and the UML language does not natively support them. The *Agent UML language specification* (AUML), which has been initiated and partly supported by the *Foundation for Intelligent Physical Agents* (FIPA), aims to develop an extension of UML to support agent-based developments. It defines model extension to, for example, model agent ontologies or to visualise the agent communication process.[3] However, many specifications within AUML remain incomplete and research efforts have been stopped. Other agent based modelling languages are often associated with specific agent development tools or methodologies such as *Prometheus*, *Tropos*, or *the Gaia modelling language* (Chan et al., 2004). Many of these approaches follow a certain agent development paradigm, such as the Belief-Desire-Intention agent design.

Another stream of agent-oriented methodology comes from the domain of knowledge engineering, which develops *knowledge-based systems* (KBS) for the system description. The agent methodologies extend the different KBS models such as the organisational model, the task model, or the knowledge model.

Another perspective of agent-oriented systems are the *tools* used in various phases. A continuous tool-support considering all stages of the development is needed. The design artifacts created in one development phase should be the base for the subsequent phase and should lead in a natural way to an executable system specification (JADEX Research Group, 2005). This concept should also extend to the application run-time to test the system against the design concept originally developed. Agent-related standards with the greatest impact and visibility were introduced by the *Foundation for Intelligent Physical Agents* (FIPA), the *Knowledge Sharing Effort* (KSE), and the *Object Management Group* (OMG) (JADEX Research Group, 2005).

Agent-based standards allow for interoperability between different agent-based systems and influence all different levels of the agent development cycle from specification to implementation. For example, in order to let agents communicate across application borders a common agent communication language needs to be defined to let agents understand the messages exchanged. Many initiatives to develop a general agent reference model have been started. However, on the architecture level very different approaches have been proposed but none of them covers all aspects of agent architectures nor has a high acceptance level.

*Agent architectures* define essential data structures, relationships between these structures, the processes or functions that operate on these data structures, and the operation cycle of an agent (Luck et al., 2004). According to Wooldridge and Jennings (1995) three different types of agent architectures can be distinguished: *deliberative architectures*, *reactive architectures*, and *hybrid architectures*. In a deliberative architecture the world is represented in an explicit and symbolic way and decisions of agents are based on logical

[3]For more information see http://www.auml.org

reasoning derived from methods such as pattern matching and symbolic manipulation (JADEX Research Group, 2005). It is assumed that agents have a model of their environment which can be used to generate an "intelligent" action based on logical reasoning together with the signals received from the environment. In contrast, reactive architectures consist of simpler agents which only react to changes in their environment but do not initiate actions based on complex behavioural structures. One idea of reactive agents is that many routine activities can be done with very little abstract reasoning. Once the ideal behaviour has been learned the task can be accomplished with little variation. Hybrid architectures try to define a combination of the advantages from the other two approaches with the aim of an integrated effective and efficient agent behaviour. Agents are equipped with some kind of artificial intelligence but also implement basic reactive features.

Finally, all the different research domains of the MAS framework (see Figure 6.4) are used to build a particular agent-based application. As already described, a variety of different application areas exist which all require different types of agents. Additionally, case studies of existing systems can help the designer of a new system to identify the strengths and weaknesses of the different approaches and to learn from mistakes made in former projects. An extensive body of such reports could help to build a general framework allowing the selection of appropriate agent-based solutions to problems of specific application areas.

### ► 6.3.9 Delimitation of multi agent simulation from other concepts and paradigms

Multi agent simulation is not a completely new concept but builds upon existing concepts of other disciplines such as object-oriented simulation or parallel and distributed discrete event simulation. An important question is how agent-based techniques can be delimited from existing concepts.

Object-orientation as a programming paradigm has achieved much success and offers a valuable abstraction for the design of complex systems (Luck et al., 2004). The agent concept extends the object-oriented design approach by making an object active. Software agents, which in their core are nothing more than an object with several implemented methods, have their own thread of control, localising not only code but their invocation as well (Odell, 2002). Each agent can have its own individual rules and goals which allows it to act autonomously. Each agent can decide individually whether to participate in a computational activity, or whether to perform the desired operation. Because of these capabilities agents cannot be directly invoked like objects. Nevertheless, they usually follow the same object-oriented programming principles of the object-oriented design. In short, agents can be described as "active objects with initiative" (Luck et al., 2004) or "objects with an attitude" (Odell, 2002) since each agent can implement different behaviour and can also change its type of behaviour depending on external or internal events during the run-time of the simulation.

Another distinguishing factor of agent-based systems is the flexibility of communication. In the object-based model communication is synchronous and a message usually invokes only one pre-defined action, whereas the agent communication model is richer and can handle asynchronous communication. There is no predefined flow of control between entities in the system but rather an open communication architecture exists to allow for flexible communication with any other entity in or outside the community.

Current object-oriented languages let an object learn only about another object's supported interfaces. With the agent-based approach, agents can advertise their interfaces to other agents or through a brokering agent. In this way, agents can learn about their environment and can adapt their behaviour according to the available services or resources. For example, an agent in a network simulation may learn about available network resource and may then make its decision how to distribute its demand among the available options.

## ► 6.4 The JADE agent platform

A simulation environment has been developed in JAVA JDK5.0, using the *Java Agent DEvelopment (JADE) Framework*[4]. JADE is open-source middleware developed jointly by CSELT (Centro Studi e Laboratori Telecommunicazioni) in conjunction with the Computer Engineering Group of the University of Parma and is used in numerous academic and industrial applications worldwide. JADE is well documented with many programming tutorials and code examples available from different sources. An active user community of over a thousand members from academic environments as well as from R&D centers of world-leading companies such as Motorola, HP, Siemens and Rockwell Automation contributes to the project by providing library extensions and add-ons.

### ► 6.4.1 Features of the JADE middleware

JADE simplifies the implementation of multi agent systems through a predefined middleware concept, which fully complies with the FIPA reference model (Bellifemine et al., 2005). The environment itself does not contain specific agent architectures but provides a basic set of functionalities, which are needed in an autonomous agent implementation. JADE defines a common agent base-class for creating user-defined agents, extending the standard functionality. JADE includes both the libraries required to develop application agents and the run-time environment that provides the basic services and that must be active on the device before agents can be executed. Each instance of the JADE run-time is called a *container*. Applications which are extending the generic agent class operate within these containers that manage the agent's life cycle as well as the status of the agent. The set of all containers is called a *platform* and provides a homogeneous layer that hides

[4]http://jade.tilab.com

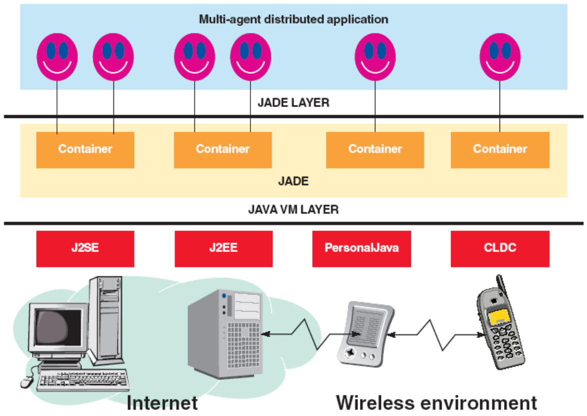



**Figure 6.5:** The JADE agent architecture. Source: Bellifemine et al. (2005)

the complexity and the diversity of the underlying layers such as the specific hardware or the operating system. The basic architecture model of the JADE platform is shown in Figure 6.5.

Agent tasks in JADE are implemented as behaviours extending standard behaviour classes such as `OneShotBehavior` or `CyclicBehavior`. Once an agent has been created in a container and is set active, behaviours are executed based on a round-robin nonpreemptive scheduling policy (Bellifemine et al., 2005). The invocation of the implemented behaviours can be triggered in different ways. For example, a behaviour can be started when a message matching certain criteria arrives at the inbox. Another possibility is to define a predefined time period after which the behaviour is started. Behaviours can also be nested or can contain sub-behaviours that are executed in parallel or serial form. In this way an agent can either be reactive (taking an action as soon as an external event is observed) or active (starting an action based of a change of an internal state).

The *communication* between agents is based on the FIPA-defined Message Transport Protocol (MTP) over which Agent Communication Language (ACL) messages can be exchanged. The format of the message content can implement an application specific ontology based on FIPA specifications for content languages (CL). Several generic JAVA interfaces can be implemented to define an application-specific ontology for agent communication. JADE provides many different transportation encodings such as XML, RDF, or proprietary bit-efficient technologies. Additionally, messages can be secured by several encryption techniques which may be important in a distributed system spanning several public networks.

To enable *asynchronous communication* between JADE agents, each agent possesses a private mailbox for storing messages from other agents. This mailbox serves as inbox for incoming messages until they can be processed by the agent. The agent can filter the inbox and can process only messages of a certain type or from a defined sender. For example, if an agent runs several parallel behaviours, each task can process only messages with a particular characteristic and leave other messages untouched. Also, the agent can decide the order of the processing to, for example, prioritise messages by certain criteria.

As defined in the FIPA specifications, JADE provides high-level services for interacting with the standard platform. An example of this is the *Directory Facilitator (DF) agent*, which provides *yellow book services* for service advertisement and discovery. Custom agent implementations can extend this service and implement an application-specific service directory.

One further feature of JADE is the support for *distributed processing* of agents in a network, which allows for scaling of the simulation environment as well as the introduction of agent mobility at run-time. Agents can be moved between different containers running on different devices and which are distributed over the network.

An extension of JADE, called LEAP (Lightweight Extensible Agent Platform) allows porting JADE code to limited-capability devices such as PDAs and mobile phones. In a possible extension of this research the agents developed within the simulator could be reused for building up a pricing test-bed running on mobile devices in a real wireless network.

### ▸ 6.4.2 The JADE ontology concept

By conforming with the FIPA standard for agent communication, the JADE environment comes with a certain degree of standard commonality. An application-specific ontology in JADE describes the content elements that can be exchanged between agents. The concept follows standard object-oriented principles by extending the abstract ontology class provided by JADE. Three main interfaces are defined that can be implemented by the specific application: `Concept`, `AgentAction`, and `Predicate`. Concepts define the data structures in the form of objects. For example, a `BidderValuation` object describes the characteristics of the valuation function used by a specific agent. Concepts define complex data structures consisting of more than one primitive type. They are usually used within an `AgentAction` to allow for exchanging such data structures between agents. The interface `AgentAction` refers to an action invoked by an agent, which becomes the content of the message. For example, if agent $A$ wants to send a new bid to another agent $B$ it can submit a `NewBid` object in the ACL message. Finally, a `Predicate` object describes if an agent's proposition is true or false. The content of the message must be the object representing the proposition to check. For example, an agent can send an object `CheckActive` to a seller agent to check if this particular seller is currently active. Beside the three generic interfaces JADE also allows for the definition of atomic elements such as Strings, Integers, or Floats that generally constitute the slots of the abstract objects.

## ▸ 6.5 The architecture and implementation of the simulation environment

The following section describes in detail the developed simulation environment for dynamic pricing in wireless access networks. We first provide an overview about the generic ontology model and the agent architecture underlying both software applications. We then present a description of the implementation and required extensions of the *PSPSim* platform and the *AdSim* platform. With this design we follow the standard JAVA programming paradigms of class inheritance and class extension. We can create a generic simulation environment, which provides the basic agent structure to implement arguably any kind of communication protocol.

### ▸ 6.5.1 Generic agent architecture and ontology

In this section we present the principle agent architecture and the ontology model of the generic simulation environment. The described framework serves as a basis for both simulation platforms (*PSPSim* and *AdSim*) and defines the basic object model for dynamic pricing simulation in wireless networks.

#### The agent architecture

The generic agent architecture defines three groups of different agent types resembling a standard three-tier architecture (Figure 6.6). The first group, called *Boundary Agents*, provides the graphical user interface for the simulation setup and monitoring. Two generic agent types have been implemented, the `GUISetupAgent` and the `GUIMonitorAgent`. The setup agent serves as the main configuration tool for defining a specific simulation scenario and instantiating the ontology. It supports reading and writing of XML to load and save such scenarios. Each implementation has to extend the classes to create the specific GUI for filling the respective objects with content.

The monitoring agent type provides a mean for visualising simulation output to the user. In its generic form it supports agent communication over the `WriteGUIEvent` object but does not yet implement any specific functionality.

On the second layer are the *Management Agents*, which are responsible for creating and parameterising the actual agent population for the simulation. A second task is the creation of agents with mediation functionality such as logging agents or agents providing yellow page services.

The `NetworkUserManagementAgent` creates the `NetworkUserAgent` and sends the agent parameterisation as a `NetworkUserProfile` object. Respectively, the `BaseStationManagementAgent` creates all agents representing a wireless base station and transmits the profile via the `BaseStationProfile` object. After that the management agents become inactive until the platform is shut down and they invoke the shut down of all single agents.

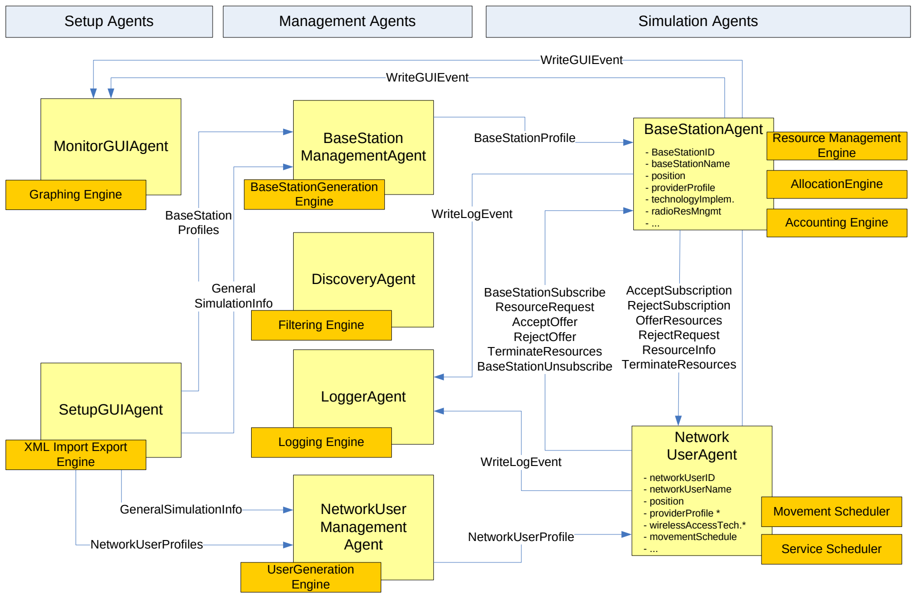



**Figure 6.6:** The generic agent architecture of the simulation environment.

Finally, the third layer holds the *Simulation Agents* involved in the experimental activity. Two principle agent types have been designed; the `NetworkUserAgent` represents users with certain usage patterns (defined in the `ServiceSchedule`) and the `BaseStationAgent` represents a physical access point offering a certain amount of resources to all network users in range. In the generic implementation such agents implement only functionality that is unspecific to any pricing mechanism. Specifically, the `NetworkUserAgent` implements a behaviour to discover base station agents covering the current position of the agent, basic setup and termination behaviours to open a communication channel to `BaseStationAgents`, and mobility methods needed to migrate agents to remote platforms.

Another important functionality implemented on this level is the random generation of service requests by the agent. For each possible `ServiceType`, a `NetworkUserAgent` can have two states, active or inactive. The time slots of activity and inactivity can either be deterministically set or stochastically generated, depending on the configuration of the `ServiceSchedule` object. If the time slots are modelled stochastically, a random inactivity time slot is created at the agent startup time and a behaviour is called, which is executed at the end of this time slot. This behaviour then starts the resource request process, which is left empty in the generic platform. At the beginning of the agent activity a random activity time slot is generated and, again, a behaviour is started, which executes at the end of the time slot to deactivate the service type and terminates the resource allocation after which the process of activation is started again.

The design of the behaviours for resource requests and resource allocation using some form of pricing mechanism is part of the actual implementation.

### The ontology model

The generic ontology model defines the main elements needed to define the agents and their properties. We distinguish between `Concepts`, which describe the internal design and data structure of agents, and `AgentActions`, which define the content of the agent communication. As described, `Concepts` can become part of `AgentAction` so agents can communicate complex data structures to other agents.

Protégé 3.1[5] has been used as the software tool for the ontology development and refinement. Protégé is a widely used tool for designing general ontologies for various purposes including agent-based platforms. An extension of Protégé allows the direct export of the ontology structures into the necessary class files for the JADE simulation environment.

Figure 6.7 shows the hierarchical relationship of all the Concept ontology elements. The two central elements are the `NetworkUserProfile` and the `BaseStationProfile`, which are described by the elements below. The elements in the middle are used by both main Concepts, while the elements on the left and right are exclusive for describing the respective agent profile.

A `NetworkUserProfile` is therefore described by a `MovementSchedule`, which describes an agent's moves in space and time, a `ServiceSchedule`, which describes the frequency of service requests of a certain `ServiceType`, and a `UserType`, which describes how it obtains resources and how it values such resources. Additionally, a `NetworkUserAgent`, is affiliated with one or more providers (which own the base stations), has a `Position` in space, and implements one or more `Technology`(-ies), which describes the technical interface.

The `BaseStationProfile` is described by a `RadioResourceManagement` object, which defines how the base station manages its wireless resources. In the simplest case this may be done on a first-come first-served basis until no resources are available. A base station implements a `PricingStrategy`, which describes how it prices for offered services. The pricing strategy, for example, can be an auction or a simple take-it-or-leave-it algorithm in which prices are formed. The `AccountingStructure` describes how a base station collects data about the used resources of active network users. The `BaseStationProfile` shares the Position object with the `NetworkUserProfile`, which defines the location of the base station in space. It also implements a `TechnologyImplementation`, which is a subclass of Technology, and further defines the characteristics of the technical setup of the base station. While the network user has only general information about the wireless technology, the base station needs more details on the specifics of the setup.

In Figure 6.8, the `AgentAction` ontology elements are shown for the two agent types `NetworkUser` and `BaseStation` which are sorted by the different agent communication acts.

To open a communication channel, a network user sends a `BaseStationSubscribe`

[5] For more information see http://protege.stanford.edu.

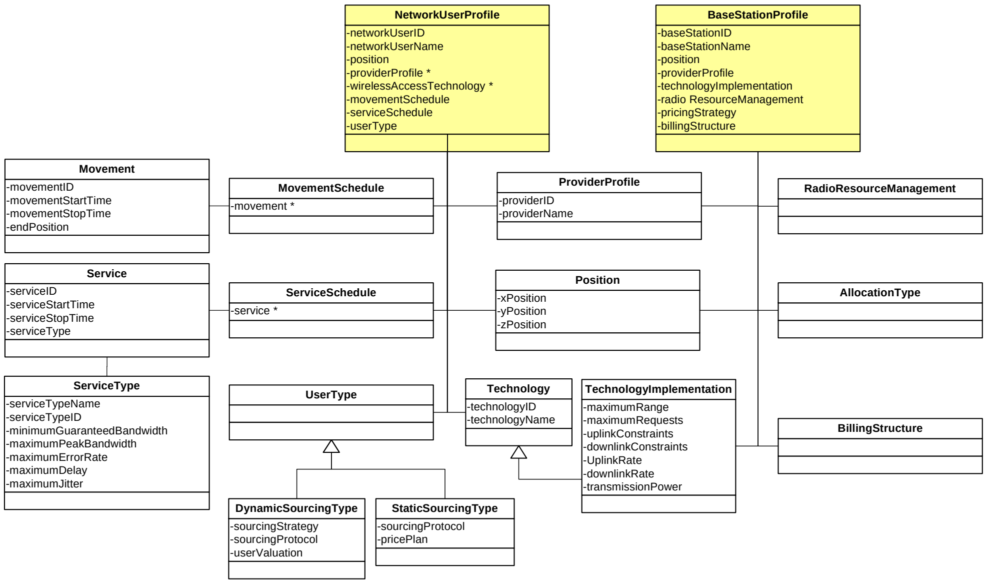



**Figure 6.7:** The main `Concept` classes defined in the generic simulation ontology.

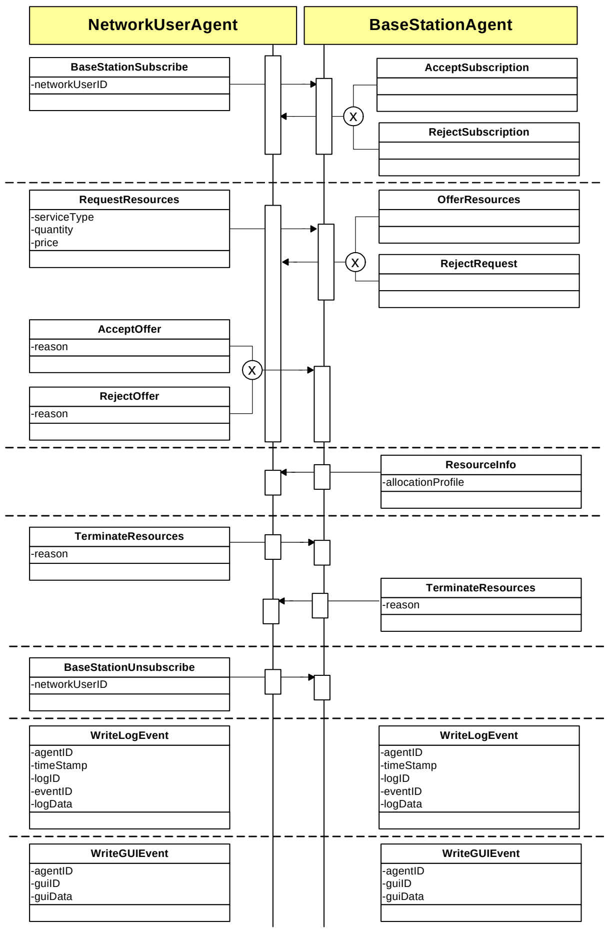



**Figure 6.8:** The main `AgentAction` classes defined in the generic simulation ontology.

object, which the base station can either accept or reject by using the `AcceptSubscription` and `RejectSubscription` objects, respectively. In the next step the user may request network resources by submitting a `RequestResources` object, which further defines his request depending on the specific implementation. The base station either sends a `OfferResources` or `RejectRequest` object and the network user can then use the `AcceptOffer` or `RejectOffer` for communicating the final decision. During the time of active services a provider may also be required to send `ResourceInfo` messages about the state of the connection or to, for example, inform the user about price changes or changes in the allocation of resources. To end a connection or the subscription, the objects `TerminateResources` and `TerminateSubscription` have been defined.

Additionally, to the basic objects for agent negotiation we have defined two generic messaging objects, `WriteLogEvent`, and `WriteGUIEvent`, which can be used by any agent to add a new entry into the database or to update GUI information about its state.

The model described above defines a general frame for the simulation environment. If the concrete implementation needs more information in certain classes it can simply extend this class. If certain elements are not used, for example, if no agent movement is needed, it simply leaves such elements empty.

### Common functionality

In the generic environment we have also specified and implemented some common functionality required for both simulation platforms (*PSPSim* and *AdSim*). The three main components are the event logging service, the discovery service providing a yellow page service about all agents present in the platform, and the generic data graphing tool.

The `LoggerAgent` implements the functionality to write a log file to a database. The table structure of the database has been predefined with a general layout to allow for fields describing the event (by ID, time stamp, and agent ID) and several generic fields for additional data created by the logging event. For example, a base station agent can log all resource reservations together with the price and the quantity of resources reserved. The log message can be sent by any agent in the environment by transmitting a `WriteLogEvent` object.

The `DiscoveryAgent` provides a yellow page service for all agents in the system, holding information about any agent in the environment. Any agent can inquire about agents of a certain type and the parameters the agent is subscribed with.

The generic graphing tool is part of the `MonitorGUIAgent`. It accepts messages of the type `WriteGUIEvent` containing data in the form `agentID`, `graphID`, `xValue`, `yValue`. Using this object an agent can send data points to the graphing application, which visualises this data graphically in a time-dependent graph. A base station agent can, for example, use this functionality to create a graphical output of prices or resource utilisation during the simulation. The graphing tool provides additional functionality to manipulate the graph, such as filter and scaling functions.

### ▸ 6.5.2 The implementation model of the PSPSim simulation platform

To create a suitable simulation platform for the PSP auction and for implementing different bidding strategies, the generic agent environment needed to be extended in several directions. We first describe the architecture extensions to the actual agents and then elaborate on the extensions to the generic ontology.

#### Internal design of the agents

First, a subclass of the `BaseStationAgent` called `PSPAuctioneerAgent` has been created with behaviours to accept bids and to execute the PSP auction algorithm. In particular, such behaviours needed to regularly check the message inbox for new bids, store the existing bid profile, execute the auction algorithm, and report the results back to the agents. Additional behaviours were implemented for managing the process of user agents joining or leaving the auction.

Second, a subclass of the `NetworkUserAgent`, named `BidderAgent`, extended the basic functionality by implementing a bidding behaviour to receive auction results and send updated bids to the auctioneer agent. The different bidding strategies are implemented by separate methods. Agents implement one of such methods as their way to act in the auctions depending on their parametrisation. To evaluate results and new bids a method called `EvaluateBid` implements the parabolic valuation function. For each bid the resulting utility and surplus is returned.

#### Extensions of the generic ontology

To accommodate the specific requirements of the PSP auction the generic ontology needed to be extended in several directions. Figure 6.9 depicts the main extensions made to the `Content` objects of the ontology. The original structure of the Content ontology is schematically shown in the middle (see Figure 6.7 as reference). Only a few extensions are needed on the side of the `BaseStationProfile`. The subclass `PSPAuction` extends the `AllocationType` class to tell the agent the type of mechanism used for resource allocation.

The main extensions are needed on the side of the users requesting services. Since different bidding strategies need to be implemented we need to reflect this in the ontology. Therefore, the PSP ontology implements and further details the `DynamicSourcingType` class, which describes the `NetworkUserAgent`. The `SourcingStrategy` is further detailed by the subclass `PSPBiddingStrategy`, which again has subclasses providing the different bidding strategies. In this way a complex substructure is created, which allows the dynamic configuration of each agent at run-time. At startup an agent reads the instances and activates the respective behaviour.

The `DynamicSourcingType` class is further detailed by the `SourcingProtocol`, which defines the format of the message exchange, and by the `UserValuation` class, which has a subclass defining the `ParabolicValuation`.

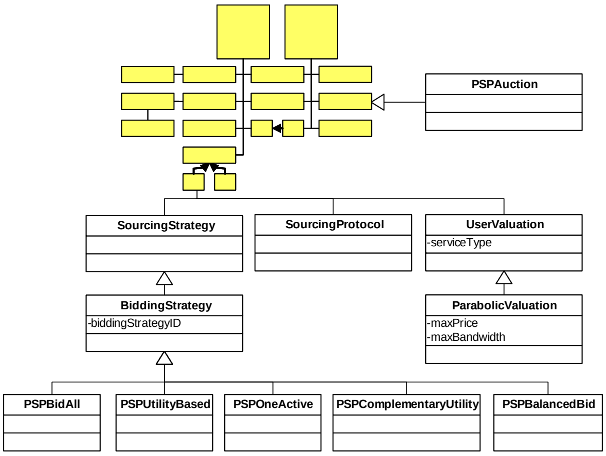



**Figure 6.9:** The ontology extensions of the PSPSim environment.

No changes needed to be made to the `AgentAction` classes as all elements for the agent communication are already available. Besides the classes for setting up and taking down the communication channel only the `RequestResources` and `ResourceInfo` classes are used for sending a bid to the base station agent and receiving auction results back.

## The PSPSim graphical user interface

To simplify the setup of the simulation scenarios we have created a simple graphical user interface to input such data or to load it from predefined files. The *PSP Setup GUI* is shown in Figure 6.10 and is divided by three main tabs. The first tab defines the general simulation parameters such as simulation length, the inactivity time of agents between submitting bids and the general auction parameters. The second tab allows the user to define and parameterise the setup of the auctioneer agents representing the mobile base stations. The third tab (shown in the picture) defines the parameters of all user agents such as position, service schedules and bidding behaviour. To have full control over the user profiles being generated, the GUI allows to define the specific setup of each individual agent.[6] Finally, the last tab lets the user define the valuation function of each agent as definedThe JADE agent architectur in the `UserType` object. All bidder profiles can be automatically loaded from a saved configuration file.

As soon as the simulation has been started the *PSP Monitor GUI* is opened. It visualises the simulation progress (upper box) and the current allocation of a bidder agent selected (in the drop-down box). Additionally, the GUI allows the experimenter to export log data into Excel format for further analysis. A predefined Excel sheet serves as a template in which up to six data sets are automatically visualised.

[6] An alternative way of defining the agent profiles is to create agent groups with identical properties or to randomise certain variables within a group in a given range.

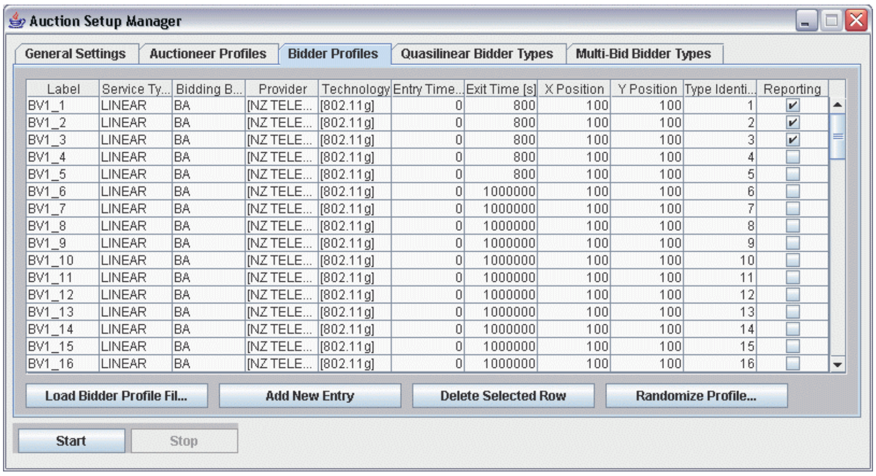



**Figure 6.10:** The graphical user interface of the PSP auction created for parameterising the experiments.

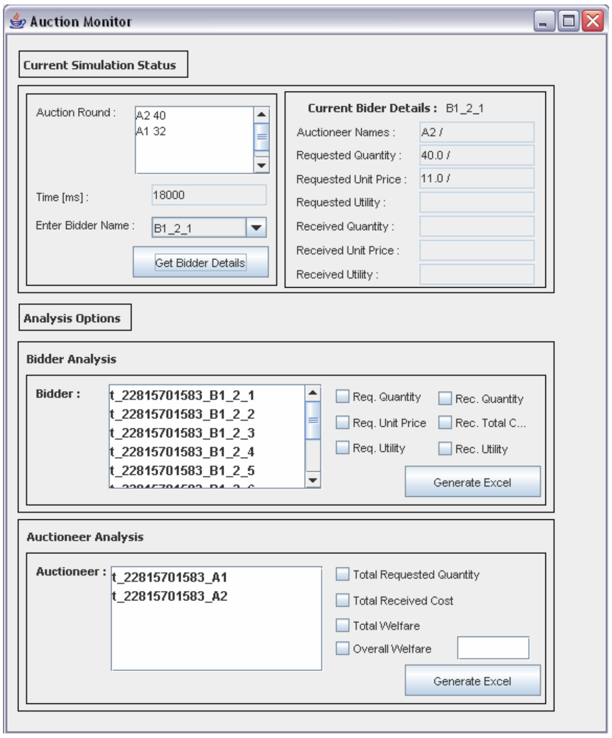



**Figure 6.11:** The graphical user interface of the PSP auction allowing for monitoring and customised report generation.

### ▶ 6.5.3 The implementation model of the AdSim simulation platform

The *AdSim* simulation environment differs from the *PSPSim* environment in one important point. With PSP the actual complexity is on the bidder side to calculate the optimal response to a received bidding profile. Also, to configure bidder agents in *PSPSim*, a considerably complex ontology was needed. In contrast, the complexity in *AdSim* is on the base station side, which needs to come up with the optimal pricing and serving strategy to maximise prices. In addition, in *AdSim*, the base station needs to run rather complex mechanisms for the radio resource management to decide on the acceptance or rejection of newly arriving service requests. Users in the *AdSim* environment only have to perform regular service requests and a simple comparison process if received offers are within the willingness-to-pay of the agent.

As with *PSPSim* we review the required changes and explain the main extensions to the generic simulation platform.

#### Internal design of the agents

To implement the required functionality we created sub classes of all simulation agents. The implementation of the `AdBaseStationAgent` was the most complex task in customising the generic architecture. The agent needs to keep track of all admitted and committed service requests as well as to maintain a data structure which stores all historic events to generate estimations of the service arrival rate and service duration. Furthermore, it needs to capture all events of service rejections to learn about the competitive situation in its transmission range.

To process new service requests the `AdBaseStationAgent` implements a `NewService-ArrivalBehaviour`, which performs a capacity check according to the technical constraints defined in its profile and the current load given by the active services. If resources are technically available, a second behaviour, the `ServiceQuoteBehaviour`, is started, which decides if the user should be admitted, given his distance from the base station, and what price should be offered. To reduce the computational workload the recalculation of the pricing strategy is not performed for each new request but only in a certain frequency. Otherwise, the results from the last optimisation round are reused for the new request. If from an economic view the user should receive an offer, the offer is sent out and the required resources are reserved at the base station.

To handle the user reactions to the service offers a behaviour called `ServiceOffer-Behaviour` processes all messages containing the `AcceptOffer` or `RejectOffer` objects. In the case of a positive response the status of the resource reservation is set to active. Otherwise, the reservation is canceled. In order to not reserve resource infinitely a `ReservationWatchdogBehaviour` regularly checks the age of the open reservations and deletes reservations after a predefined time window. If users terminate their active services the `TerminationBehaviour` deletes the active reservation.

On the user side we have extended the `NetworkUserAgent` by two subclasses. The

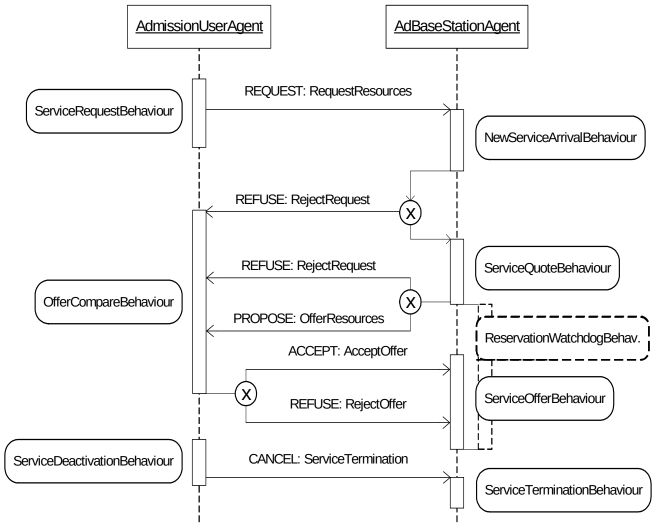



**Figure 6.12:** The sequence chart of the communicative act between the `AdmissionUserAgent` and the `AdBaseStationAgent`. The boxes show the behaviour activated by the message retrieval. The arrows describe the message performative and the message content consisting of an `AgentAction` object.

`AdmissionUserAgent` implements the `OfferCompareBehaviour` to compare offers from different providers and to select the offer with the lowest price. In contrast, the `SubscriptionUserAgent` only sends a request to a base station of the subscribed provider and checks if he is accepted.

The sequential diagram in Figure 6.12 shows the communicative act between the `AdmissionUserAgent` and the `AdBaseStationAgent` and the corresponding behaviour invoked upon message receipt.

### Extensions of the generic ontology

The *AdSim* ontology extended several main `Concept` classes to implement the details needed to define the simulation scenario (Figure 6.13). The `BaseStationProfile` was further defined by different `DynamicPricingStrategy` classes, which contain the definition of the optimisation model to be executed by each base station.

Furthermore, the `TechnologyImplementation` class needed to be extended for the admission control function of the agent. The `UplinkConstrains` and `DownlinkConstraints` classes further define how the admission control is performed. For example, the `DownlinkConstraints` class has two describing classes, which define that a `PowerConstraint` and a `RateConstraint` applies. An agent reading the profile can then activate the respective behaviours to check resources in the predefined way every time a new service request arrives.

On the user side both the `DynamicAccessType` and `StaticAccessType` classes, which are part of the `UserType` class, have been extended. The dynamic user is further described

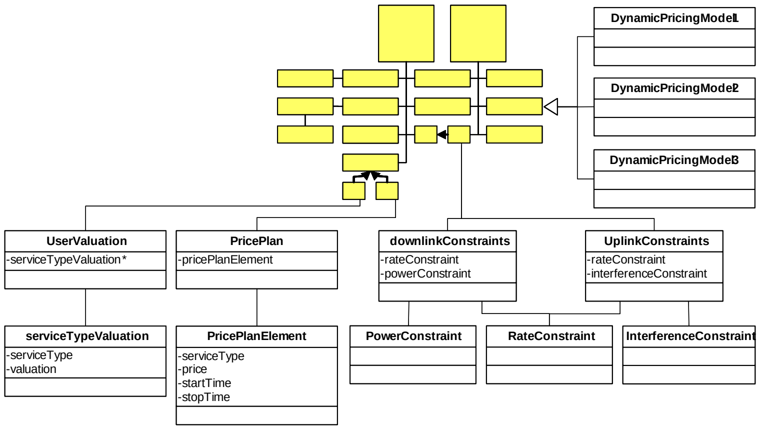



**Figure 6.13:** The ontology extensions of the *AdSim* environment.

by the `UserValuation` class defining his valuation for different service types. The static user can access the `PricePlan` class, which contains details about the predefined pricing structures with the respective provider.

As with the *PSPSim* ontology no changes were required to the `AgentAction` classes as they were sufficient to carry the respective information between agents for the negotiation process. The details of the communicative act between the user agents and the base station agents is depicted in Figure 6.12.

### The AdSim graphical user interface

For the *AdSim* environment four different graphical user interfaces have been created. Figure 6.14 shows the *AdSim* Setup GUI, which, as for the *PSPSim* case, lets the user define the parameters for the base stations, the admission-based user groups (shown in the figure), and the subscription-based user groups. Based on this information an instance of the *AdSim* ontology is created and used for the distributed setup of the simulation over the management agents.

The main visualisation tool of the *AdSim* environment is the *Visualiser* tool (Figure 6.15). It allows the experimenter to see the experimental setup with all base stations and all network users. As an additional option, it visualises the progress of the simulation by several filtering methods. First, the user activity is shown by different colours. Green indicates that the user has been admitted and is actively using resources. Red indicates that a service request has been blocked. A blue dot means that the price offered by all available base stations has been above the user's valuation. Yellow tells the experimenter that the evaluation process is currently in progress. The user activity visualisation can be filtered by service type since agents can be concurrently active or inactive in several service classes. A second visualisation option monitors the state of the base stations in the same window. The maximum transmission range is shown as a black circle. Additionally, we can picture the current optimisation parameters, price and maximum distance. The

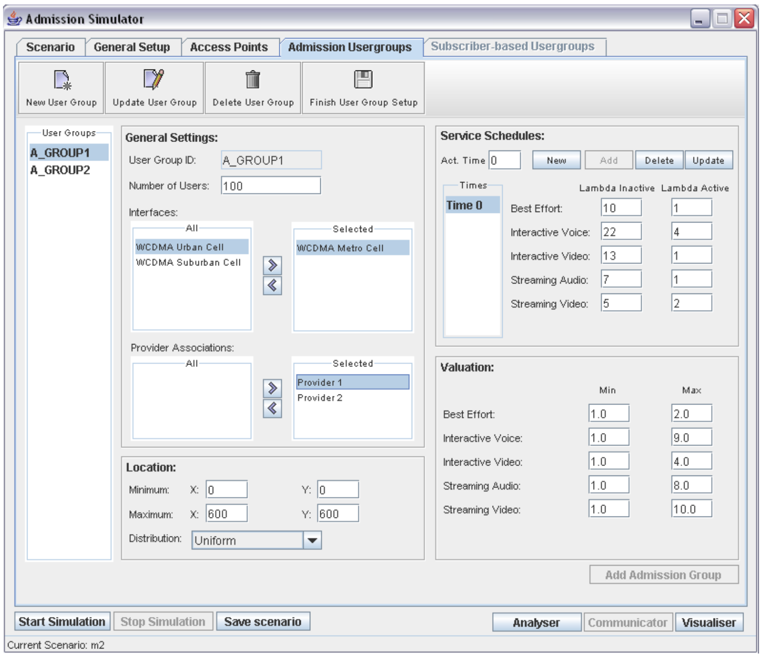



**Figure 6.14:** The graphical user interface (showing the tab for configuring an admission-based user group) created for defining the simulation scenarios and to load and save existing scenarios in XML format.

price is shown by the color of the transmission range of the base station. Green indicates a low price (close to the lowest recorded valuation in a certain time window) and red indicates a current price equal to the highest measured user valuation. The size of this coloured circle indicates the area the base station is currently serving. The visualisation option can be filtered by service type since prices and maximum distance may be different in each service class.

Besides the *Visualiser* tool the experimenter can make use of two other functions. The *Analyser* tool has been created as an extension of the graphing tool implemented by the generic platform. It allows the experimenter to plot a graph of a variable over time (Figure 6.16) from any agent implementing the respective method. Several selection and filtering functions have been added to allow the concurrent graphing from multiple sources on different scales.

Finally, the *Communicator* tool has been designed as a generic GUI to print custom messages from any agent in the system, which implements the method for publishing information on this tool. For example, a base station can regularly send information about network utilisation, prices in the different service classes, and blocking rates. The `WriteGUIEvent` ontology object is used to transfer the information between agents and the Communicator.

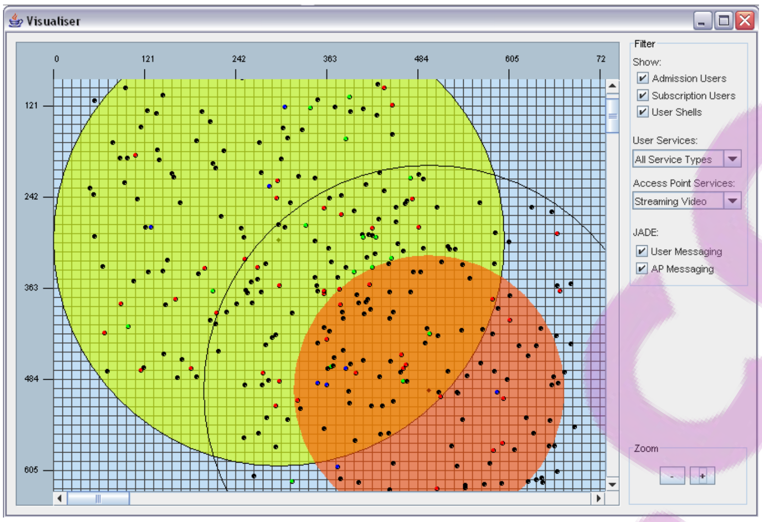



**Figure 6.15:** The visualisation tool for displaying the setup of network users and base stations as well as the activity of users and the resulting price and range optimisation of the base stations.

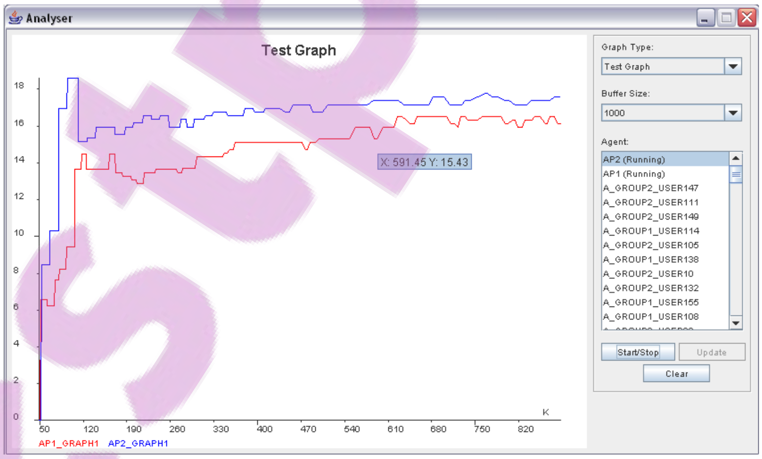



**Figure 6.16:** The analysing tool for displaying time-dependent information from any agent implementing the analysing behaviour. The graph shows the price formation of two competing base stations.

## ▶ 6.6 Chapter Summary

In this chapter we have presented the simulation environment, which has been created as part of the research work. We have introduced the reader to the main concepts of agent-based modelling and simulation and have explained the MAS research framework. We have then provided the details about the generic ontology and agent-based architecture, which has been developed to implement two simulation platforms, the *PSPSim* platform, and the *AdSim* platform. We provided an overview of the extensions to implement the required functionality and the GUI tools for parameterising and monitoring the simulation experiments.

# Chapter 7

## Conclusions and Future Research Directions

The overarching research objective of this thesis was to advance the knowledge of a market for wireless resources, in which resources are sold by multiple competing providers on-demand. Such a scenario could become a reality with the stepwise introduction of Internet-Protocol (IP) in wireless networks which enables seamless access to wireless resources independent from the underlying wireless technology. Together with high-performance mobile devices, which will be capable of dynamically switching between networks and negotiating for wireless resources at the time of demand, the new wireless environment will offer a myriad of new services and business opportunities.

While the technological migration can already be foreseen, the corresponding business models, which are supporting such a vision, have still not fully evolved. New forms of business are expected to grow, which will be based on short-term access in terms of seconds or minutes and which will partly replace the idea of long-term customer relationships. To support such business models from the economic perspective, new approaches to resource allocation and pricing of wireless resources are required.

The existing literature is rich on innovative resource allocation schemes, which make use of advanced pricing concepts for reaching certain design objectives. A central assumption of many models is that the network selection is long-term based on an explicit user decision. In contrast, we believe, that in the future scenario, the network selection process will become dynamic on the time scale of seconds or minutes. Network selection and resource negotiation may become fully automated and users may not be aware of frequent changes to optimise both the technological and the economical aspects of the active connection.

Most of the existing models, which take a multi-provider setting into consideration assume a cooperative approach, in which providers jointly optimise their actions to reach joint goals. In contrast, we believe that, in the future scenario, wireless transport services will be provided by competing firms, which do not cooperate but act selfishly to reach their individual goals.

Under these two main assumptions of short-term decision making and multi-provider competition we have focussed on understanding possible settings and feasible market institutions for a scenario on which customers have multiple options to join a network, while on the other side, providers have to cope with competition on a time-scale of seconds or minutes. We have presented two different approaches for resource allocation through pricing. The selected approaches differ in several aspects such as the design objective, the assumed market setting, the time-scale of pricing and the type of transport services supported. Our focus was on understanding the optimal behaviour of rational entities in a wireless network environment under defined rules of interaction. While in the first model our focus was on the user side, in the second approach, we have concentrated on developing optimal pricing strategies for a profit-seeking network provider facing direct price competition.

In the following we summarise the previous chapters, describe the contributions of the thesis and provide an outlook on future research which has been identified by the work in this thesis.

## ► 7.1 Chapter Summary

**Chapter 1** has explained the motivation for this research and has introduced the reader to the topic and the focus of the research. This included a short introduction to the main concepts of next-generation wireless networks as well as the basic underlying economic principles.

**Chapter 2** has formally defined the research objectives and the research methodology, namely mathematical modelling using non-cooperative game theory and agent-based simulation. We have elaborated on simulation as a valid research methodology in MIS and have presented a brief overview of the main simulation concepts.

**Chapter 3** has developed a classification framework for categorising existing pricing models into different time-scales. While similar classification concepts have been used in previous work, we have adapted the model to reflect the special characteristics of mobile and wireless networks. The chapter has also provided a comprehensive overview of pricing concepts in the existing literature and has briefly summarised the approach and findings of studies most relevant to our research.

**Chapter 4** has presented a resource allocation model based on congestion pricing, which allocates resources on the flow-level. We see wireless resources as a public good, which needs to be efficiently allocated among users. We therefore have implemented a second-price auction at each seller and let customers bid for bandwidth. Among several bidding strategies we present the *BalancedBid* strategy, which has been shown to be the truthful best-reply for a bidder faced with multiple second-price auctions. We describe the results of an extensive simulation study to understand the properties of *BalancedBid* in several network setups and to compare the performance of the bidding strategies. We also present the results of a seven-cell, three provider simulation experiment and compare the performance of *BalancedBid* against alternative allocation mechanisms.

**Chapter 5** has presented a model in which wireless resources are provided by multiple, profit-seeking providers and prices are formed at the time of the service request. The main objective of each provider becomes to set prices at admission time so that revenue from resource allocation is maximised. We have modelled the competitive situation between the providers as a non-cooperative game, in which certain information is private information to each provider. Because of the difficulty of identifying the equilibrium strategies in explicit form we have used agent-based simulation to learn from the steady-state characteristics of the market. In the simulation we have presented an approximation heuristic, which uses a fixed grid to identify the optimal price/cell-radius combination given that each base station can collect information about customer behaviour and the competitive situation and adapts its actions accordingly so that revenue is maximised.

**Chapter 6** has presented the agent-based simulation platform, which has been an integral part of this research. We have explained the main concepts of agent-based simulation and the multi-agent system approach. In the second part of this chapter we have presented our simulation platform; we have described the details of the developed architecture, the agent ontology, and the graphical user interfaces to parameterise and monitor the simulation experiments. Finally, the chapter outlines the extensions of the ontology and the agent architecture, which were needed to implement the two models developed in Chapter 4 and 5.

## ▸ 7.2 Contributions of the research

In this section we reflect on the research contributions of the material presented in the previous chapters and relate back to the research questions asked in Chapter 2 of this thesis.

**Q1: What is a suitable categorisation framework for pricing in wireless communication networks, with which the existing literature can be classified?**

In a first step we have developed a simple classification framework for dynamic pricing in wireless networks by using the time-scale of the pricing decision. While similar approaches have been identified in the literature which have been used to classify existing work in fixed networks, we have added additional components to reflect the time-scales usually used in mobile and wireless networks, namely the admission time-scale and the subscription time-scale. For each scale we have provided a short description of the main characteristics.

**Q2: What are relevant studies and research articles on pricing of wireless resources and how do they fit into the developed categorisation framework?**

In the next step the developed classification framework has been used to categorise the existing work on dynamic pricing in mobile and wireless networks. For research studies most related to our work we have provided a brief summary of the approach and the research findings on each time-scale. Additionally, we have reviewed layer-spanning pricing approaches and have looked into existing studies describing resource allocation and pricing in multi-provider settings.

**Q3: What is the optimal behaviour of a rational user with the possibility to connect to multiple wireless networks, when faced with competition from other customers?**

The first contribution of this research stream was the definition of a multi-provider market, in which each base station sells bandwidth as a divisible good and uses the Progressive-Second-Price auction as the allocation tool. For a situation in which mobile terminals can bundle resources we have developed the *BalancedBid* strategy, which has been shown to extend the concept of the truthful best-reply in the multi-auction market. The *BalancedBid* strategy is therefore the utility-maximising way for a myopic bidder to behave and to distribute their demand among the auctions. We have also shown that a Nash equilibrium for the entire market exists, in which resources are efficiently allocated among bidders. Simulation results contribute to the understanding of the *BalancedBid* strategy in different setups, which could not be explained analytically. This included an analysis of settings, in which only some bidders had access to multiple auctions. We have also shown the effects on convergence and efficiency when different bidders have multiple-access to different auctions.

In addition to the main strategy we have also developed alternative bidding strategies if bidders are constrained in the behavioural options. We have shown by simulation that the alternative bidding strategies could not combine the properties of *BalancedBid*. Either, the strategies do not lead to efficient market outcomes, or the auction market does not converge to equilibrium.

To build our intuition on how PSP could be used in a more complex setting we have compared the performance of the PSP market with other centralised allocation schemes. While certainly the outcomes of this experiment strongly depend on the chosen setup, the results have shown how the developed distributed allocation mechanism could be used in a wireless broadband setting.

**Q4: What is the behaviour of a revenue-maximising wireless provider when faced with price competition from other wireless providers partly or fully covering its service area?**

The contribution of the second research stream consists of a centralised pricing model, in which multiple profit-seeking wireless network providers offer network resources in a market and pricing is determined at the time of the customer request. We have presented a game theoretic discussion for the two-provider situation, in which cells fully or partly overlap. In the game of complete information, in which the payoff functions of providers are common knowledge and in which customer demand is sufficient to utilise all resources in both network cells, we could show the existence of a Nash equilibrium.

We have also described the game of incomplete information, in which a part of the cell setup is private information to each player. In this setting players needed to form beliefs about the setup of the other player to form its pricing function. We have employed several analytical techniques to find an equilibrium pricing strategy but have not been successful finding an analytical solution. We have used simulation to gain an understanding about steady-state, price/cell-radius combinations chosen by the providers if we let providers learn about the prices set by the provider and the share of customers having access to multiple networks. Several experiments with different setups have shown that steady-state price/cell-radius combinations exist. We could also show how these price/cell-radius combinations change when gradually varying a single input parameter such as the cell overlap or the maximum cell size of a single provider.

For conducting the simulation experiments we have developed a heuristic approximation framework to identify a near optimal price/cell-radius combination without explicitly solving the constrained maximisation problem given for the provider. We have used this method to implement the game of incomplete information by letting providers form estimator functions about user demand structures and the competitive situation.

**Q5: How can a simulation platform be developed, which allows us to experiment with different pricing mechanisms in a wireless multi-provider network?**

Besides the main contributions of developing behavioural strategies in a multi-provider wireless market for different market settings and for reaching different design goals, we have developed a general software architecture based on agent technology, which allows us to simulate resource allocation in wireless networks from a microeconomic perspective. The contribution in this area consists of an agent-based architecture and a general ontology, which have been kept modular and extentable so that other market mechanisms and pricing protocols can easily be implemented. In the simulation environment each entity is represented by an individual agent, which implements its preferences, behaviour and strategies. While we have only implemented myopic strategies with our research the simulation environment is also capable of supporting advanced strategies, which involve complex learning algorithms or which require intelligent decision-making by inferring future actions of other market participants.

### ► 7.2.1 Methodological approach of the thesis

The concurrent use of non-cooperative game theory and agent-based simulation has resulted in a methodological contribution. While the use of both methodologies in separation is common to the field, the feedback cycle between the mathematical model and the simulation experiments has not been extensively described in the existing literature. This feedback cycle has been used in both directions: first, we have transformed the formal model into a simulation approach. Second, we have inferred back from the simulation results to the formal model.

## ► 7.3 Possible implementation scenarios for dynamic pricing

After having developed a detailed theoretical basis on dynamic pricing under direct competition and having shown two different modes of pricing, namely flow-based pricing and access-based pricing, we now turn our attention to some aspects of a potential practical implementation. We briefly describe two different application scenarios: a campus WLAN network using dynamic pricing to control demand, and an implementation of dynamic pricing on a national level. In both cases we limit our attention to the access case as the more practical way of

### ► 7.3.1 Implementation of dynamic pricing with a campus WLAN

Universities around the world are more and more implementing Wireless LANs around the campus to provide access to intranet and internet resources for students and staff. Such networks usually span all major buildings as well as highly frequented outside areas where students may want to gain access. The main device type used in this setting is the laptop, but more and more students also use such networks with alternative mobile devices.

In most settings access is provided on a free basis and costs are partly recovered from student fees. Students log in with their user name and password and have full access to all university resources as well as the entire internet (maybe with some filtering of inappropriate content). No QoS takes place and users share the available bandwidth at a certain location on a best-effort basis.

At the University of Auckland a different model is applied. Users are offered two service classes: a best-effort model with high-speed access to the intranet and dial-up speed to all internet resources. This service class is free of charge. The second class offers high-speed access also to the internet with a monthly charge of USD 2 per month (and a data limit of 200MB). This ensures that students with a higher valuation for such resources have the possibility to gain faster access. However, resources are still shared on a best-effort basis.

An alternative to this model would be a dynamic pricing model to prioritize demand on the campus. As basic access options users may be offered dial-up speed as of today. If demand in a certain location is very low users may gain high-speed access without additional costs. If demand rises, e.g., during the morning hours, the network may impose additional charges for high-speed access if new users want to join the network at this location. Such prices would be signalled to the newly arriving users and they could decide if their valuation for such resources is higher than the price. Alternatively, they could move to a different, less busy location on the campus to gain access at potentially lower or no costs.

Such a scenario could be implemented by offering a small client application for download, which contains a software agent acting on the user's behalf. This agent could learn about the individual willingness-to-pay of the user and could take over the signalling to the network. This may mean that a user may either gain full-speed access or may fall back to dial-up speed.

As already existing today, in some locations multiple networks from different external providers (such as Vodafone, Woosh, or Easy Internet) may be available. The client application could easily be extended to accommodate for a free selection of the network provider based on the current price level in the different networks.

The described scenario would improve overall user satisfaction as resources are no longer shared on a best-effort basis but the valuation of users is taken into account. With increasing local competition students and staff would be given an alternative for internet access, which would bring down overall charges.

For a finer control of network resources the network designer may also decide to introduce additional time slots of price signalling instead of announcing the price more frequently than at the initial access. As users in such a setting usually stay connected over a longer time period this may enable a more exact control of resource distribution.

### ▶ 7.3.2 Implementation of dynamic pricing in a national economy on the example of New Zealand

We now turn our attention to the case of a larger implementation of dynamic pricing in a national economy. We can only sketch the idea of a potential implementation and provide some idea about possible consequences.

With today's industry structure wireless resources are usually provided by 2-5 large network providers (offering 2.5G, 3G or 3.5G services) covering a large area of the national economy or at least highly populated areas. Such coverage is complemented by smaller WLAN implementations or other access technologies, which are operated by various smaller (local) organizations.

Such a setting is also true for New Zealand. The main provider of cell-based technology are New Zealand Telcom (the national incumbent) and Vodafone. However, different technologies are used for providing service to customers. In addition, Woosh offers a UMTS-based data service across the main city centers (requiring special access devices in form of mobile modems) and various other local providers offer wireless internet services in highly populated areas such as Auckland or Wellington.

The current situation makes it difficult to think about a realistic scenario for competition-based pricing (e.g., on the access level) on a national basis. As for each access technology different devices are needed providers would not be able to directly attract customers from other providers as long as they are not supplied with multi modal devices. In addition, both main providers have an established customer base and it would not be attractive for them to offer contracts with higher flexibility to churn.

However, looking at the market for wireless data for business applications especially in highly populated areas of New Zealand, competitive pricing may become a reality much faster than in other areas. Since devices become more and more advanced and allow for connection to different wireless standards such as UMTS, WLAN or most recently, WiMAX, it becomes technically feasible to dynamically switch between technologies or even use different wireless standards concurrently. Local providers may offer small java clients residing on client devices, which intelligently select the wireless network with the lowest pricing.[1] While only attractive in highly populated areas, such providers could easily capture an important share of mobile data traffic, which is currently priced well above average OECD pricing levels[2].

To price their services such alternative providers would require a flexible pricing model, which would allow them to skim customers from established providers and would allow them to control demand in the local market. Since resources in such alternative networks would also be limited and customers would require certain QoS levels, the proposed dynamic access pricing model would be well suited to serve as a pricing basis. While alternative providers could set their prices according to current demand, established providers can only compete with the pricing defined in the subscription plans of their customers and will not be able to adopt their pricing on a short-term basis. It may even become attractive for residential customers to open up their private WLAN to such a market if suitable software is available and charging and billing processes have been established.

To gain back market share for mobile data traffic, established providers may decide

[1] One very recent example for such application is iCall for Apple's iPhone, which allows seamless roaming of voice calls between the GPRS connection and a WLAN

[2] According to the 2007 Telecommunications Market Monitoring Report (ComCom NZ)

to enter the game by either announcing lower subscription prices or by participating in the new model of access-based pricing and thus, compete directly.

The described scenario would bring down the overall price levels for mobile data services in certain areas of the network. While local providers can decide on region-based pricing, national providers may be limited in their choice due to regulation and legal equality rights. Thus, such local competition may lead to lower pricing levels nationally on a longer time-scale.

## ► 7.4 Future research

This thesis has produced an important contribution to the field of network pricing and has illuminated the aspect of competitive access in wireless networks when resources are sold on-demand. However, when looking at the big picture, we could only touch a few aspects for designing a feasible and practical pricing framework for next-generation wireless networks. Many aspects, such as the charging process, the design of billing relationships, or the integration of network transport pricing with pricing for mobile services and content needed to be left untouched. Furthermore, to enable us to focus on the main concepts of pricing we needed to make many simplifications and abstractions from the complexity of real wireless systems. While such simplifications are nothing unusual in the research field, they need to be overcome in the next step to create realistic and implementable pricing concepts and to proceed to a proof-of-concept.

Several avenues for future research can be deduced from our research. First, the identified pricing concepts do not yet consider the mobility aspects of mobile and wireless networks. Since mobility and handovers are an integral functionality in next-generation wireless networks, new concepts need to be developed to include such aspects in the pricing decision. A second aspect is the support of multiple service classes to develop an integrated framework for a multitude of different wireless applications.

We also see a potential in further developing the agent-based simulation environment. Since the JADE middleware used in the project is fully portable to any mobile device, and agents can communicate over a standard IP connection in a fully distributed environment, the developed architecture and ontology can be used in a possible proof-of-concept.

# Bibliography


Acemoglu, D., Ozdaglar, A. E., and Srikant, R. (2004). The marginal user principle for resource allocation in wireless networks. In *Proceedings of the 43rd IEEE Conference on Decision and Control*.

Ahn, G., Campbell, A., Veres, A., and Sun, L. (2002). SWAN: service differentiation in stateless wireless ad hoc networks. In *Proceedings of the IEEE INFOCOM 2002*.

Ahonen, T., Kasper, T., and Melkko, S. (2004). *3G Marketing: Communities and strategic partnerships*. John Wiley & Sons Ltd.

Alexopoulos, C. and Goldsman, D. (2004). To batch or not to batch? *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 14(1):76–114.

Alpcan, T., Basar, T., Srikant, R., and Altman, E. (2002). CDMA uplink power control as a noncooperative game. *Wireless Networks*, 8:659–670.

Alwis, P. M. D. (2005). Call admission control and resource utilization in WCDMA networks. Master's thesis, Department of Electrical and Computer Engineering, University of Canterbury, New Zealand.

Arabas, P., Kamola, M., Malinowski, K., and Malowidzki, M. (2002/2003). Pricing for IP networks and services. *Information - Knowledge - Systems Management*, 3:153–171.

Argon, N. T., Alexopoulos, C., Andradóttir, S., and Goldsman, D. (2006). Replicated batch means variance estimators in the presence of an initial transience. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 16(4):317–328.

Aufderheide, P. (1999). *Communications Policy and the Public Interest: The Telecommunications Act of 1996*. Guilford Press.

Axelrod, R. (2003). Advancing the art of simulation in the social sciences. *Japanese Journal for Management Information System, Special Issue on Agent-Based Modeling*, 12(3).

Azouzi, R. E., Altman, E., and Wynter, L. (2003). Telecommunications network equilibrium with price and quality-of-service characteristics. In *Proceedings of the ITC, Berlin*.

Badia, L., Lindstroem, M., Zander, J., and Zorzi, M. (2004). An economic model for the radio resource management in multimedia wireless systems. *Computer Communications - Applications and Services in Wireless Networks*, 27(11):1056–1064.

Badia, L. and Zorzi, M. (2005). Radio resource management with utility and pricing for wireless LAN Hot-Spots. *Wireless Personal Communications*, 34:127142.

Bahl, P., Bahl, P., and Chandra, R. (August 2003). MultiNet: Connecting to multiple IEEE 802.11 networks using a single wireless card. Technical report, Microsoft Research.

Bambos, N., Chen, S. C., and Pottie, G. J. (1995). Radio link admission algorithms for wireless networks with power control and active link quality protection. In *Proceedings of INFOCOM*, pages 97–104.

Basar, T. and Alpcan, T. (2003?). A hybrid systems model for power control in multicell wireless data networks. *Performance Evaluation, Elsevier Science Publishers B. V.*, 57(4):477 – 495.

Bellifemine, F., Caire, G., Poggi, A., and Rimassa, G. (2003). JADE - A white paper. *exp - in search of innovation, special issue on JADE*, 3(3):6–12.

Bellifemine, F., Caire, G., Trucco, T., and Rimassa., G. (2005). JADE Programmers Guide. Last Revision March 2005, available at http://jade.tilab.com/doc/programmersguide.pdf. Technical report, TILAB.

Beltran, F. and Roggendorf, M. (2005). A simulation model for the dynamic allocation of network resources in a competitive wireless scenario. In *Proceedings of the Second International Workshop, MATA 2005, Montreal, Canada, October 17-19, 2005. Lecture Notes in Computer Science 3744, ISBN 3-540-29410-4*, pages 54–64.

Beltran, F. and Roggendorf, M. (2007). Multiple equilibria in symmetric strategies for simultaneous auctions in next-generation bandwidth markets. In *To appear in the Proceedings of the 1st Workshop for Games in Communication Networks (GameComm 2007), October 22, Nantes, France*.

Berezdivin, R., Breinig, R., and Topp, R. (2002). Next-generation wireless communications concepts and technologies. *IEEE Communications Magazine*, 40(3):108–16.

Bircher, E. and Braun, T. (February 2004). An agent-based architecture for service discovery and negotiations in wireless networks. In *Proceedings of the 2nd International Conference on Wired/Wireless Internet Communications, Frankfurt, Germany*.

Bitsaki, M., Stamoulis, G. D., , and Courcobetis, C. (2005). A new strategy for bidding in the network-wide progressive second price auction for bandwidth. In *Proceedings of CoNEXT2005, Toulouse, France*.

Bulow, J. and Roberts, J. (1989). The simple economics of optimal auctions. *The Journal of Political Economy*, 97(5):1060–1090.

Cao, X.-R., Shen, H.-X., Milito, R., and Wirth, P. (2002). Internet pricing with a game theoretical approach: concepts and examples. *IEEE/ACM Transactions on Networking*, 10(2):208 – 216.

Chan, K., Sterling, L., and Karunasekera, S. (2004). Agent-oriented software analysis. In *Proceedings of the Australian Software Engineering Conference*, pages 20–27.

Chen, Q. and Niu, Z. (2005). A game-theoretical power and rate control for wireless ad hoc networks with step-up price. *IEICE Transactions on Communications, Special Section on Advances in Ad Hoc Mobile Communications and Networking*, 8:3515–3523.

Chonacky, N. and Winch, D. (2005). 3Ms for instruction: Reviews of Maple, Mathematica, and Matlab. *Computing in Science & Engineering*, 7:7–13.

Chu, K. and Altmann, J. (1999). Demand for different qualities of service for internet access: A review of INDEX findings. *Network Modelling in the 21st Century, Royal Society, London*.

Courcoubetis, C. and Weber, E. (2003). *Pricing communication networks: economics, technology and modeling*. Wiley-Interscience series in systems and optimization.

Cranefield, S. and Purvis, M. (2001). Generating ontology-specific content languages. In *Proceedings of the Ontologies in Agent Systems Workshop*, pages 29–35.

Cushnie, J. (2003). *QoS Charging For Internet Access Networks: The Wireless QoS Gateway*. PhD thesis, Lancaster University, England.

Das, S., Lin, H., and Chatterjee, M. (2004). An econometric model for resource management in competitive wireless data networks. *IEEE Network*, 18(6):20 – 26.

Dash, R. K., Jennings, N. R., and Parkes, D. C. (November/December 2003). Computational-mechanism design: a call to arms. *IEEE intelligent systems*, pages 40–47.

DaSilva, L. A. (2000). Pricing for QoS-enabled networks: A survey. *IEEE Communications Review*, 3(2).

Davidsson, P. (2002). Agent based social simulation: A computer science view. *Journal of Artificial Societies and Social Simulation*, 5(1).

Dramitinos, M., Stamoulis, G. D., and Courcoubetis, C. (2004). Auction-based resource reservation in 2.5/3G networks. *Mobile Networks and Applications*, 9:557566.

Dramitinos, M., Stamoulis, G. D., and Courcoubetis, C. (April 2005). Auction-based resource allocation in UMTS high speed downlink packet access (HSDPA). In *Proceedings of the First EuroNGI Conference Conference on Traffic Engineering (NGI 2005), Rome, Italy*.

Elayoubi, S.-E., Chahed, T., and Salahaldin, L. (2005). Optimization of radio resource management in UMTS using pricing. *Computer Communications, Elsevier*.

Falkner, M., Devetsikiotis, M., and Lambadaris, I. (2000). An overview of pricing concepts for broadband IP networks. *IEEE Communications Surveys*, pages 2–13.

Falomari, D., Mandayam, N., Goodman, D., and Shah, V. (1998). A new framework for power control, in wireless data networks: games, utility, and pricing. In *Proceedings of the Allerton Conference on Communication, Control, and Computing*, pages 546–555.

Feng, N., Mau, S.-C., and Mandayam, N. B. (September 2004). Pricing and power control for joint user-centric and network-centric resource allocation. *IEEE Transactions on Communications*, 52(9):1547–1557.

Ferber, J. (1999). *Multi-Agent System: An Introduction to Distributed Artificial Intelligence*. Harlow: Addison Wesley Longman.

Fishwick, P. (1994). Computer simulation: Growth through extension. In *Proceedings of the European Simulation Multiconference: Modelling and Simulation of Uncertain Dynamics. Barcelona, Spain*.

Fromm, J. (2005). Types and forms of emergence. working paper. available at http://www.citebase.org/cgi-bin/citations?id=oai:arxiv.org:nlin/0506028.

Fulp, E. W. and Reeves, D. S. (2002). The economic impact of network pricing intervals. In *Proceedings of the Workshop on Advanced Internet Charging and QoS Technology (ICQT)*.

Gibbons, R. (1992). *Game theory for applied economists*. Harvester Wheatsheaf. New York, London.

Gilbert, N. and Troitzsch, K. G. (2005). *Simulation for the social scientist*. Open University Press, UK.

Goodman, D. and Mandayam, N. (2000). Power control for wireless data. *IEEE Personal Communications Magazine*, 7(2):48–54.

Grossglauser, M. (1999). Control of network resources over multiple time-scales. *Networking Technologies - ERCIM News*, 37.

Gruber, H. (2005). *The economics of mobile telecommunications*. Cambridge Press.

Gruber, T. R. (1993). A translation approach to portable ontologies. *Knowledge Acquisition*, 5:199–220.

Harris, S. L. (1990). *Agents of Chaos: earthquakes, volcanoes and other natural disasters*. Missoula, MT: Mountain Press.

Hayel, Y., Maillé, P., and Tuffin, B. (2005). Modelling and analysis of internet pricing: introduction and challenges. In *Proceedings of International Symposium on Applied Stochastic Models and Data Analysis (ASMDA)*.

Hayel, Y. and Tuffin, B. (2005). A mathematical analysis of the cumulus pricing scheme. *Computer Networks*, 47(6):907–921.

Heidelberger, P. and Welch, P. D. (1983). Simulation run length control in the presence of an initial transient period. *Operations Research*, 31:1109–1144.

Heikkinen, T. (2001). Congestion based pricing in a dynamic wireless network. In *Proceedings of the 53rd IEEE Vehicular Technology Conference, 2001. VTC 2001 Spring. IEEE VTS*, volume 2, pages 1073–1076.

Heikkinen, T. M. (2002). On congestion pricing in a wireless network. *Wireless Networks*, 8:347354.

Holma, H. and Toskala, A. (2000). *WCDMA for UMTS*. Wiley, New York.

Huang, J., Berry, R. A., and Honig, M. L. (2004). Auction-based spectrum sharing. In *Proceedings of the 2nd Workshop on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt04), Cambridge, UK, March 24-26*.

Ileri, O., Samardzija, D., Sizer, T., and Mandayam, N. B. (2005). Demand responsive pricing and competitive spectrum allocation via a spectrum server. In *Proceedings of the first IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, 2005. DySPAN 2005. 2005*, pages 194– 202.

JADEX Research Group (2005). JADEX research group project web site (http://vsis-www.informatik.uni-hamburg.de/projects/jadex). Technical report, University of Hamburg, Germany.

Ji, H. and Huang, C.-Y. (1998). Non-cooperative uplink power control in cellular radio systems. *Wireless Networks*, 4:233–240.

Jiang, H. and Zhuang, W. (2004). QoS-oriented resource allocation for video traffic in the wireless internet. In *Proceedings of the Global Telecommunications Conference, GLOBECOM2004*, volume 3, pages 1660 – 1664.

Jin, N., Venkitachalam, G., and Jordan, S. (2003). Dynamic pricing of network resources. In *Proceedings of IEEE GLOBECOM 2003*.

Johari, R. (2004). *Efficiency loss in market mechanisms for resource allocation*. PhD thesis, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science.

Klemperer, P. (2004). *Auctions : theory and practice*. Princeton, N.J. : Princeton University Press.

Koutsopoulou, M., Kaloxylos, A., Alonistioti, A., and Merakos, L. (2004). Charging, accounting and billing management schemes in mobile telecommunication networks and the internet. *IEEE Communications Surveys & Tutorials*, 6:50–58.

Krishna, V. (2004). *Auction theory*. Academic Press.

Kurose, J. F. and Ross, K. W. (2005). *Computer Networking: A Top-Down Approach Featuring the Internet*. Boston : Pearson/Addison Wesley.

Labrou, Y., Finin, T., and Peng, Y. (April 1999). The current landscape of agent communication languages. *Intelligent Systems, IEEE Computer Society*, 14.

Laffont, J.-J. and Tirole, J. (2000). *Competition in telecommunications*. MIT Press.

Lataoui, O., Rachidi, T., Samuel, L., Gruhl, S., and R.-H., Y. (2000). A QoS management architecture for packet switched 3rd generation mobile systems. In *Proceedings of Networld & Interop*.

Law, A. M. and Kelton, W. D. (2007). *Simulation modeling and analysis*. Dubuque, IA : McGraw-Hill.

Lazar, A. and Semret, N. (1998). Design, analysis and simulation of the progressive second price auction for network bandwidth sharing. In *Proceedings of the 8th International Symposium on Dynamic Games and Applications, Maastricht, Netherlands*.

Le Bodic, G., Irvine, J., Girma, D., and Dunlop, J. (September 2000). Dynamic 3G network selection for increasing the competition in the mobile communications market. In *Proceedings of the IEEE Vehicular Technology Conference, Boston, USA*.

Lee, J. W., Mazumdar, R. R., and Shroff, N. B. (2002). Downlink power allocation for multi-class CDMA wireless networks. In *IEEE INFOCOM*, pages 1480–1489.

Lee, S.-B., Ahn, G.-S., Zhang, X., and Campbell, A. T. (2000). INSIGNIA: An IP-based quality of service framework for mobile ad hoc networks. *Journal of Parallel and Distributed Computing*, 60(4):374–406.

Liebowitz, S. J. and Margolis, S. E. (1994). Network externality: An uncommon tragedy. *The Journal of Economic Perspectives*, pages 133–150.

Lin, H., Chatterjee, M., Das, S. K., and Basu, K. (May/June 2005). ARC: An integrated admission and rate control framework for competitive wireless CDMA data networks using non-cooperative games. *IEEE Transactions on Mobile Computing*, 4(3):243–258.

Liu, P., Honig, M., and Jordan, S. (2000). Forward-link CDMA resource allocation based on pricing. In *Proceedings of the Wireless Communications and Networking Conference*, volume 3, pages 1410 – 1414.

Liu, P., Zhang, P., Jordan, S., and Honig, M. L. (2004). Single-cell forward link power allocation using pricing in wireless networks. *IEEE Transactions on Wireless Communications*, 3(2).

Luck, M., Ashri, R., and d'Inverno, M. (2004). *Agent-based software development*. Artech House.

M3I (January 2001). Market managed multi-service internet. Technical report, 5th Framework EU Project, IST Program, No. 11429.

Ma, H., Hausheer, D., Reichl, P., and Stiller, B. (2002). A simulation model for the Cumulus Pricing Scheme. In *Proceedings of Advanced Simulation Technologies Conference, San Diego, CA*.

MacKie-Mason, J. K. and Varian, H. R. (1995). Pricing the internet. *Public access to the Internet, MIT Press, Cambridge, MA, USA*, pages 269 – 314.

MacKie-Mason, J. K. and Wellman, M. P. (2006). Automated markets and trading agents. In *Handbook of Computational Economics: Agent-Based Computational Economics*, volume 2. Leigh Tesfatsion and Kenneth L. Judd (eds.), North-Holland.

Maglaras, C. and Zeevi, A. (2005). Pricing and design of differentiated services: Approximate analysis and structural insights. *Operations Research*, 53(2):242–262.

Maillé, P. (2003). Market clearing price and equilibria of the progressive second price mechanism. Technical Report 1522, IRISA.

Maillé, P. (2004). Auctioning for downlink transmission power in CDMA cellular systems. In *Proceedings of the 7th ACM international symposium on modeling, analysis and simulation of wireless and mobile systems*, pages 293 – 296.

Maillé, P. (2005). *Mécanismes d'enchéres pour le partage de ressources dans les réseaux*. PhD thesis, ENST-Bretagne.

Maillé, P. and Tuffin, B. (2003). The progressive second price mechanism in a stochastic environment. *Netnomics*, 5(2):119–147.

Maillé, P. and Tuffin, B. (2004a). Multi-bid auctions for bandwidth allocation in communication networks. In *Proceedings of the 23rd IEEE Infocom Conference, Hong-Kong, China*.

Maillé, P. and Tuffin, B. (2004b). Multi-bid versus progressive second price auctions in a stochastic environment. In *Proceedings of the 4th International Workshop on Internet Charging and QoS Technology(ICQT'04),Springer Verlag, Lecture Notes in Computer Science*.

Maillé, P. and Tuffin, B. (October 2006). Pricing the internet with multi-bid auctions. *IEEE/ACM Transactions on Networking*.

Marbach, P. and Berry, R. (2002). Downlink resource allocation and pricing for wireless networks. In *Proceedings of the IEEE INFOCOM 2002*.

Mirhakkak, M., Schult, N., and Thomson, D. (October 2001). Dynamic bandwidth management and adaptive applications for a variable bandwidth wireless environment. *IEEE Journal on Selected Areas in Communications,*, 19(10):1984–1997.

Mitchell, B. M. and Vogelsang, I. (1991). *Telecommunications pricing - theory and practice*. Cambridge University Press.

Murray, K. (2005). *Admission and handover management for multi-service heterogeneous wireless access networks*. PhD thesis, Department of Electrical Engineering, Cork Institute of Technology.

Musacchio, J. and Walrand, J. (2003). Game theoretic modeling of wifi pricing. In *Proceedings of the 41st Annual Allerton Conference on Communication and Control, Monticello, IL*.

Nahrstedt, K., Shah, S. H., and Chen, K. (2005). 1. In *Cross-Layer Architectures for Bandwidth Management in Wireless Networks*, pages 42–59. Springer.

Nigel, G. G. (1996). Simulation as a research strategy. In *Social science microsimulation*. K. G. Troitzsch, U. Mueller, G. N. Gilbert & J. E. Doran (Eds.), (pp. 448-454). Springer, Berlin, Germany.

Noy, N. F. and McGuinness, D. L. (March 2001). Ontology Development 101: A guide to creating your first ontology. Technical report, Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880.

Nunamaker, J. F., J., Chen, M., and Purdin, T. D. M. (1990). Systems development in information systems research. *Journal of Management Information Systems*, 7(3):89–106.

Nwana, H. S. and Ndumu, D. T. (1999). A perspective on software agents research. *The Knowledge Engineering Review*, 14(2):1–18.

Odell, J. (2002). Objects and agents compared. *Journal of Object Technology*, 1:41–53.

Odlyzko, A. (1999). Paris metro pricing for the internet. In *Proceedings of the 1st ACM conference on Electronic commerce*, pages 140 – 147, Denver, Colorado, United States.

Odlyzko, A. (2001). Internet pricing and the history of communications. *Computer Networks: the International Journal of Distributed Informatique*, 36(5-6):493–517.

Ormond, O., Miro-Muntean, G., and Murphy, J. (April 2006). Economic model for cost effective network selection strategy in service oriented heterogeneous wireless network environment. In *Proceedings of the 10th IEEE/IFIP Network Operations and Management Symposium (NOMS 2006), Canada*.

Ormond, O., Perry, P., and Murphy, J. (2005). Network selection decision in wireless heterogeneous networks. In *Proceedings of the 16th International Symposium on Personal, Indoor and Mobile Radio Communications*, volume 4, pages 2680– 2684.

Osborn, C. (2005). Customer retention: Can wireless data make em stay put? *The Future of Wireless: Business Strategies, Broadband Technologies, and Network Operations*.

Parkes, D. C. (2001). *Iterative Combinatorial Auctions: Achieving Economic and Computational Efficiency*. PhD thesis, Department of Computer and Information Science, University of Pennsylvania.

Parsons, J. D. (2000). *The Mobile Radio Propagation Channel*. Wiley, Second Edition.

Passas, N. and Merakos, L. (April 2003). Unified QoS provision in wireless access networks. In *Proceedings of the International Workshop on Wireless Mobile & Always Best Connected, Glasgow, UK*.

Plasser, E., Ziegler, T., Reichl, P., and Mittenecker, G. (2002). Internet services beyond best effort. *Elektrotechnik & Informationstechnik*, 119(2):30–36.

Reichl, P., Flury, P., Gerke, J., and Stiller, B. (June 2001a). How to overcome the feasibility problem for tariffing internet services: The cumulus pricing scheme. In *Proceedings of IEEE ICC 2001, Helsinki*, pages 2079–2083.

Reichl, P., Leinen, S., and Stiller, B. (1999b). Pricing models for internet services. Technical report, Deliverable for the Charging and Accounting Technologies for the Internet Project (CATI).

Reichl, P., Leinen, S., and Stiller, B. (May 2829, 1999a). A practical review of pricing and cost recovery for internet services. *Internet Economic Workshop Berlin, Germany,*.

Reichl, P. and Stiller, B. (2001). Nil nove sub sole: Why internet charging schemes look like as they do. In *Proceedings of the 4th Berlin Internet Economics Workshop IEW 2001, Berlin*.

Reichl, P., Stiller, B., and Ziegler, T. (2001b). Charging multi-dimensional QoS with the Cumulus Pricing Scheme. In *Proceedings of the ITCom 2001, Denver, Colorado, August 2001*.

Ries, M., Puglia, R., Tebaldi, T., Nemethova, O., and Rupp, M. (2005). Audiovisual quality estimation for mobile streaming services. In *Proceedings of the 2nd International Symposium on Wireless Communication Systems*, pages 173 – 177.

Roberts, J. (2004). Internet traffic, QoS and pricing. *Proceedings of the IEEE*, 92(9):1389 – 1399.

Roggendorf, M. and Beltran, F. (2006a). Flow-based resource allocation in a multiple-access wireless market using an auction. In *Proceedings of the Second International Workshop on Incentive-Based Computing (IBC), July 4-7, Lisbon, Portugal*.

Roggendorf, M. and Beltran, F. (2006b). A simulation-based approach to bidding strategies for network resources in competitive wireless networks. In *Proceedings of the Joint International Conference on Measurement and Modeling of Computer Systems, June 26-30, 2006, Saint-Malo, France*.

Roggendorf, M., Beltran, F., and Gutierrez, J. (2006). Architecture and implementation of an agent-based simulation tool for market-based pricing in next-generation wireless networks. In *Proceedings of the TridentCOM 2006, 2nd International Conference on Testbeds and Research Infrastructures for the Development of Networks and Communities, March 1-3, Barcelona, Spain*, pages 282 – 288.

Rulnick, J. and Bambos, N. (1997). Mobile power management for wireless communication network. *Wireless Networks*, 3(1):3–14.

Russell, S. J. and Norvig, P. (2003). *Artificial Intelligence: a modern approach*. Prentice Hall, 2nd edition.

Saraydar, C., Mandayam, N., and Goodman, D. (1999). Pareto efficiency of pricing-based power control in wireless data networks. In *Proceedings of the IEEE Wireless Communications and Networking Conference*, volume 1, pages 231–235.

Saraydar, C., Mandayam, N., and Goodman, D. (2000). Power control in a multicell CDMA data system using pricing. In *Proceedings of the 52nd IEEE Vehicular Technology Conference, 2000*, volume 2, pages 484–491.

Saraydar, C., Mandayam, N., and Goodman, D. (2001). Pricing and power control in a multicell wireless data network. *IEEE Journal on Selected Areas in Communications*, 19(10):1883 – 1892.

Saraydar, C., Mandayam, N., and Goodman, D. (2002). Efficient power control via pricing in wireless data networks. *IEEE Transactions on Communications*, 50(2):291 – 303.

Semret, N. (1999). *Market mechanisms for network resource sharing*. PhD thesis, Columbia University, Center for Telecommunications Research.

Semret, N., Liao, R., Campbell, A., and Lazar, A. (2001). Pricing, provisioning and peering: Dynamic markets for differentiated internet services and implications for network interconnections. *IEEE Journal on selected areas in communications*, 18(12).

Shah, S. H. and Nahrstedt, K. (2004). Price-based channel time allocation in wireless LANs. In *Proceedings of the 24th International Conference on Distributed Computing Systems Workshops*, Hachioji, Tokyo, Japan.

Shah, V., Mandayam, N., and Goodman, D. (1998). Power control for wireless data based on utility and pricing. In *Proceedings of the Personal Indoor Mobile Radio Communications Conference, PIMRC'98, Cambridge, UK*, pages 1427–1432.

Shakkottai, S. and Srikant, R. (1999). Scheduling real-time traffic with deadlines over a wireless channel. In *Proceedings of the International Workshop on Wireless Mobile Multimedia*, pages 35–42.

Shapiro, C. and Varian., H. R. (1999). *Information rules : a strategic guide to the network economy*. Harvard Business School Press, Boston, Massachusetts.

Sheikh, A. U. H. (2004). *Wireless communications: theory and techniques*. Kluwer Academic Publishers.

Sheldon, T. (2001). *McGraw-Hill's Encyclopedia of Networking & Telecommunications*. McGraw-Hill Companies.

Shin, S. and Weiss, M. (2005). Optimal pricing for broadband wireless internet access service. In *Wireless Telecommunications Symposium*, pages 240– 245.

Shor, M. (2006). "dynamic games," dictionary of game theory terms, game theory.net. Online at http://www.gametheory.net/dictionary/url_of_entry.html.

Siris, V. A. (2002). Resource control for elastic traffic in CDMA networks. In *Proceedings of the Eighth Annual International Conference on Mobile Computing and Networking, MOBICOM 2002, Atlanta, Georgia, USA*, pages 193–204.

Soldatos, J., Vayias, E., and Kormentzas, G. (2005). On the building blocks of quality of service in heterogeneous IP networks. *IEEE Communications Surveys & Tutorials, First Quarter 2005*, 7(5):70–89.

Sonnessa, M. (2004). *Modelling and simulation of complex systems*. PhD thesis, Department of Economics and Finance, University of Turin, Italy.

Soursos, S., Courcoubetis, C., and Polyzos, G. (2003). Pricing differentiated services in the GPRS environment. *ACM Wireless Networks Journal (WINET), WMI Special Issue*, 9(4).

Staab, S. and Studer, R. (2004). *Handbook on Ontologies*. Springer.

Stiller, B., Braun, T., Guenter, M., and Plattner, B. (1999). The CATI project: Charging and accounting technology for the internet. In *Proceedings of the 5th European Conference on Multimedia Applications, Services, and Techniques (ECMAST99), Madrid, Spain, May 26-28*.

Stiller, B., J. Gerke, H., Reichl, P., and Flury, P. (2001a). Charging for differentiated internet services. In *Proceedings of the ITCOM'01, Denver, CO*.

Stiller, B., Reichl, P., Gerke, J., and Flury, P. (2001b). A generic and modular internet charging system for the cumulus pricing scheme. *Journal of Networks Systems Management*, 9.

Stone, P. and Veloso, M. M. (2000). Multiagent systems: A survey from a machine learning perspective. *Autonomous Robots*, 8(3):345–383.

Sun, J. and Li, H. (2005). Price driven soccer video adaptation for mobile users. In *the IEEE Transactions on Circuits and Systems for Video Technology*.

Sun, J., Modiano, E., and Zheng, L. (2004). A novel auction algorithm for fair allocation of a wireless fading channel. In *Proceedings of the Conference on Information Science and System, Princeton, NJ, March,2004*.

Sun, J., Zheng, L., and Modiano, E. (2003). Wireless channel allocation using an auction algorithm. In *Proceedings of the Allerton Conference on Communications, Control and Computing*.

Sutton, R. S. and Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. Cambridge: MIT Press.

Tesfatsion, L. (2006). *Handbook of Computational Economics, Volume 2: Agent-based computational economics*. L. Tesfatsion and K.L. Judd, Handbooks in economics series, North Holland.

Tirole, J. (1988). *The theory of industrial organization*. MIT Press, Cambridge, Massachusetts.

Tuffin, B. (2002). Revisited progressive second price auctions for charging telecommunication networks. *Telecommunication Systems*, 20(3-4).

Turocy, T. L. and von Stengel, B. (2001). *Encyclopedia of Information Systems*, volume 2, chapter Game Theory, pages 403–420. Elsevier Science, San Diego, USA.

Tyan, H.-Y. and Hou, J. (2002). Design, realization, and evaluation of a component-based compositional network simulation environment. In *Proceedings of the 2002 Communication Networks and Distributed Systems Modeling and Simulation Conference, San Antonio, Texas*.

Varian, H. R., Farrell, J., and Shapiro, C. (2005). *The Economics of Information Technology: An Introduction*. Cambridge University Press, Cambridge, UK.

Vlassis, N. (2003). A concise introduction to multiagent systems and distributed AI. Course material, University of Amsterdam, The Netherlands, available at http://staff.science.uva.nl/ vlassis/cimasdai/.

Walke, B., Seidenberg, P., and Althoff, M. P. (2003). *UMTS - the fundamentals*. Jon Wiley & Sons, Ltd.

Wang, Q. (2006). Pricing and equilibrium in communication networks. In *Handbook of Optimization in Telecommunications*. Springer Science + Business Media.

Wang, Q., Peha, J. M., and Sirbu, M. A. (1997). *Internet Economics*, chapter Optimal Pricing for Integrated-Services Networks with Guaranteed Quality of Service. Joseph Bailey and Lee McKnight, MIT Press.

Wang, X. and Schulzrinne, H. (2000). An integrated resource negotiation, pricing, and QoS adaption framework for multimedia applications. *IEEE Journal on Selected Areas in Communications*, 18:2514–2529.

Wisely, D., Eardley, P., and Burness, L. (2002). *IP for 3G*. Wiley.

Wood, M. and DeLoach, S. A. (2001). An overview of the multiagent systems engineering methodology. In *Agent-Oriented Software Engineering, Lecture Notes in Computer Science.*, volume 1957. P. Ciancarini, M. Wooldridge, Springer Verlag, Berlin, Germany.

Wooldridge, M. and Jennings, N. (1995). Intelligent agents: Theory and practice. *The Knowledge Engineering Review*, 10(2).

Woolridge, M. J. (2002). *An introduction to multiagent systems*. New York: J. Wiley.

Wu, D. (2005). QoS provisioning in wireless networks. *Wireless Communications and Mobile Computing, vol. 5, pp. 957–969, 2005. Wiley*.

Xiao, M., Shroff, N. B., and Chong, E. K. P. (2001). Utility-based power control in cellular wireless systems. In *Proceedings of IEEE INFOCOM*.

Yaipairoj, S. and Harmantzis, F. (2006). Auction-based congestion pricing for wireless data services. In *Proceedings of the IEEE International Conference on Communications (ICC 2006), Istanbul, Turkey*.

Yang, J., Ping, Z., Zheng, H., Xu, W., Yinong, L., and Xiaosheng, T. (2006). Towards mobile ubiquitous service environment. *Wireless Personal Communications*, 38(1):67 – 78.

Zhang, W. (2005). Bearer service allocation and pricing in heterogeneous wireless networks. In *Proceedings of the IEEE International Conference on Communications*, volume 2, pages 1367– 1371.

Zhang, Y., Jiao, L., Chen, S., and Ni, Y. (2003). Information, incentive and efficiency in power market design. In *Proceedings of the 6th International Conference on Advances in Power System Control, Operation and Management*, pages 587 –592.

Zhou, C., Honig, M., and Jordan, S. (2005). Utility-based power control for a two-cell CDMA data network. *IEEE Transactions on Wireless Communications*, 4(6):2764 – 2776.

Zhou, C., Honig, M., Jordan, S., and Berry, R. (2002). Utility-based resource allocation for wireless networks with mixed voice and data services. In *Proceedings of the IEEE International Conference on Computer Communications and Networks, Miami,Florida, USA*, pages 485–488.

Zhou, C., Honig, M., Jordan, S., and Berry, R. (2003). Forward-link resource allocation for a two-cell voice network with multiple service classes. In *Proceedings of the IEEE Wireless Communications and Networking Conference*, volume 2, pages 1060 – 1065.

Zhou, C. and Honig, M. L. (2001). Two-cell utility-based resource allocation for a CDMA voice service. In *Proceedings of the IEEE Vehicular Technology Conference*, volume 1, pages 27–31.

Zhou, C., Qian, D., Pissinou, N., and Makki, K. (2004a). Resource allocation based on pricing for wireless multimedia networks. In *Proceedings of the IEEE Wireless Communications and Networking Conference*, volume 1, pages 477 – 482.

Zhou, C., Zhang, P., Honig, M. L., and Jordan, S. (2004b). Two-cell power allocation for downlink CDMA. *IEEE Transactions on Wireless Communications*, 3(6).

Zhu, H. and Chlamtac, I. (2005). QoS for multimedia services in wireless networks. In *Resource Management in Wireless Networks*, pages 2–34. Springer.