TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABREVIATIONS

3G	3rd Generation
4G	4th Generation
ASC	Agence Spatiale Canadienne
CDF	Cumulative Distribution Function
CDMA	Code division multiple access
CP	Content Provider
CS	Coalitional Structure
ÉTS	École de Technologie Supérieure
GB	Gigabytes
GCA	Geographic Coverage Area
HetNet	Heterogeneous Network
ICT	Information and Communications Technology
IEEE	Institute of Electrical and Electronics Engineers
KB	Kilobytes
Kbps	Kilobits per second
KKT	Karush–Kuhn–Tucker
LTE	Long Term Evolution
MB	Megabytes
Mbps	Megabits per second

MIMO	Multiple-Input and Multiple-Output
MMP	Multi-Provider Payoff
NTU	Non-Transferable Utility
OFDMA	Orthogonal Frequency-Division Multiple Access
PCF	Process of Coalition Formation
PDF	Probability Density Function
PED	Price Elasticity of Demand
QoS	Quality of Service
SLA	Service Level Agreement
SP	Service Provider
SFC	Selective Free Content
SISO	Single-Input and Single-Output
TDM	Time-division multiplexing
SW	Social Welfare
SWF	Social Welfare Function
TU	Transferable Utility
UMTS	Universal Mobile Telecommunications Service
UWF	Usage Willingness Factor
WCDMA	Wide Band Code Division Multiple Access
WiFi	Wireless Fidelity

LISTE OF SYMBOLS AND UNITS OF MEASUREMENTS

$\mathbb{N}_T$	Set of all users.
$\mathbb{N}_C$	Set of voice users
$\mathbb{N}_D$	Set of data users
N_x	size of the set $\mathbb{N}_x$
N_u	Number of new users joining the market in each period
Q_T	Total available bandwidth
Q_C, Q_D	Total available bandwidth to voice and data services
c_d	data volume cap
p_d, p_c	price of data and voice services
γ	general notation for data rate
γ_c,γ_d	offered data rate to voice and data users
β	Spectral efficiency
θ_d	leveling factor for user's utility
ω_d	valuation factor for the data plan
b_g	random variable of user budget for data service
η	expected size of data flow
λ_c	service request rate of each voice user
λ_u	scale factor for the rate of incoming data flows
λ_d	rate of flow requests for each user $\lambda_d = \lambda_u c_d$

λ_D	overall service request rate of all voice data
$\frac{1}{\mu_c}$	voice service time
B_c	expected blocking probability of voice calls
B_d	expected blocking probability of data flows
A_i	Coverage area of provider i
G_i	Normalized size of A_i
S_i	Normalized technology speed of provider i
I_i	Set of new users joining provider i in each period
p_i	Current price of provider i
P_i	Price strategy set of provider i
s_i^j	Satisfaction factor (SF) of user j with provider i
d_i^j	Amount of data used by user j with provider i
D_i^j	Maximum amount of data used by user j with provider i
$f_i(s)$	PDF of user satisfaction for provider i
α_i	Cost of providing one unit of data (Provider i)
$c_i(G_i)$	Constant cost for provider i
V_i^j	Maximum payoff of user j in the network of provider i
$\pi_i(k,m,.)$	Profit of provider i in each period with respect to the parameters k,m, etc
K	Shape factor of payoff function
CS	Coalition Structure

SW	Social Welfare
SWF	Social Welfare Function
UWF	WiFi Usage Willingness Factor
S_{WiFi}	Base satisfaction factor for WiFi
S_{3G}	Base satisfaction factor for 3G
S_{4G}	Base satisfaction factor for 4G

INTRODUCTION

Fourth generation homogeneous wireless networks (4G) fast approach the theoretical limits of radio link performance but still cannot provide the 4A paradigm (any rate, anytime, anywhere, affordable) due to economic and technological barriers. Nevertheless, further significant gains can be achieved by introducing advanced network topologies that also integrate heterogeneous technologies such as LTE and WiFi. One way to achieve such heterogeneity in the offered services is to form a coalition of providers. This approach, due to its unique characteristic of cost reduction in the phase of network expansion as well as the embedded heterogeneity, can reduce the overall price of the service and give the subscribers the choice to instantly change their access technology based on the service price or data rate preferences. From another viewpoint, the heterogeneity of service types in today's Internet and the way in which content is delivered define a bold line between the business model of content providers (CPs) and service providers (SPs). In recent efforts, several major SPs in the US market tried to form a coalition with CPs to deliver free content to the end users (AT&T Data, 2016; Verizon-Data, 2016; T-Mobile, 2016). However, the interaction of providers in such coalitions should be controlled in a careful way since they can violate the neutrality of the Internet. In all mentioned cases, the economics of Internet, especially in wireless markets, and the pricing schemes play a major role in forming any coalition. In this thesis we address these issues by studying first the current pricing schemes in the wireless markets and then by analyzing possible coalitions in both network and service domains. By doing so, we introduce new frameworks for inter-SP and SP-CP coalitions. In our approach, we consider the end users and their usage behavior as the core of analysis and try to improve their overall satisfaction from service while the profit of providers is increasing due to cooperation. Namely, we believe that due to the competitive market and high cost of implementing heterogeneous technologies by one operator, all operators will be driven to form coalitions that can provide not only significant economic advantages but also can offer a significantly better level of services with major improvements in coverage, throughput, reliability, and energy consumption. Nevertheless, to achieve that objective, three key issues have to be addressed: a) the role of current pricing schemes on the profit of providers, b) the effect of subscriber's usage behavior on provider's pricing strategy, c) the possible cooperations between major network providers based on the pricing methods and content usage trends. In the following sections we describe in short the main problems, our objectives, the work originality, the applied methodology, the thesis content, and the achieved results.

0.1 Problem statement

The wireless technologies have been under very fast development in the past decade. However, the statistics show that the wireless services are still not available to many users around the world or even if they were, the cost and quality of this services do not look reasonable to many users. Also, there exist multiple technologies in the market and for using each of them a user needs a separate contract with the related provider which can be quite tedious. Moreover, it is difficult for a single provider to offer all types of services that would be globally accessible and affordable for all users in the market. This issue brings the idea of forming the coalitions among service providers to reach a win-win solution for both providers and users. To investigate coalition formation, one should consider the heterogeneity of providers in the current market. To the best of our knowledge, three general types of heterogeneous networks can be defined. The first type is the topological HetNet which refers to a network that consists of equipment with different coverage ranges like picocells, femtocells, microcells, and macrocells. The second type is the technological HetNet which refers to a set of wireless access technologies with various speeds and coverage ranges (like WiFi, LTE, and UMTS) that would be offered in the same market by multiple providers. These two technologies can be accessed concurrently by multi-interface devices similar to the most smartphones nowadays. A traditional view of the first two HetNet types can be seen in Fig. 0.1 where there are two cellular and two WiFi providers that are competing in the same area, and as a result, they have coverage overlapping. The third type

HetNets is referred to as service-oriented HetNets that contains heterogeneity in the offered services. Two major entities can be indicated in this category. Namely, the group of providers that offer infrastructure and connectivity and the group of providers that provide content to the end users. All providers from the first two types of HetNets fall into the first group of the third HetNet type. Unlike the technological and service-oriented HetNets, the topological HetNets are usually defined on a single provider's network. Hence, we consider the technological and service-oriented HetNets in our studies for coalition formation.

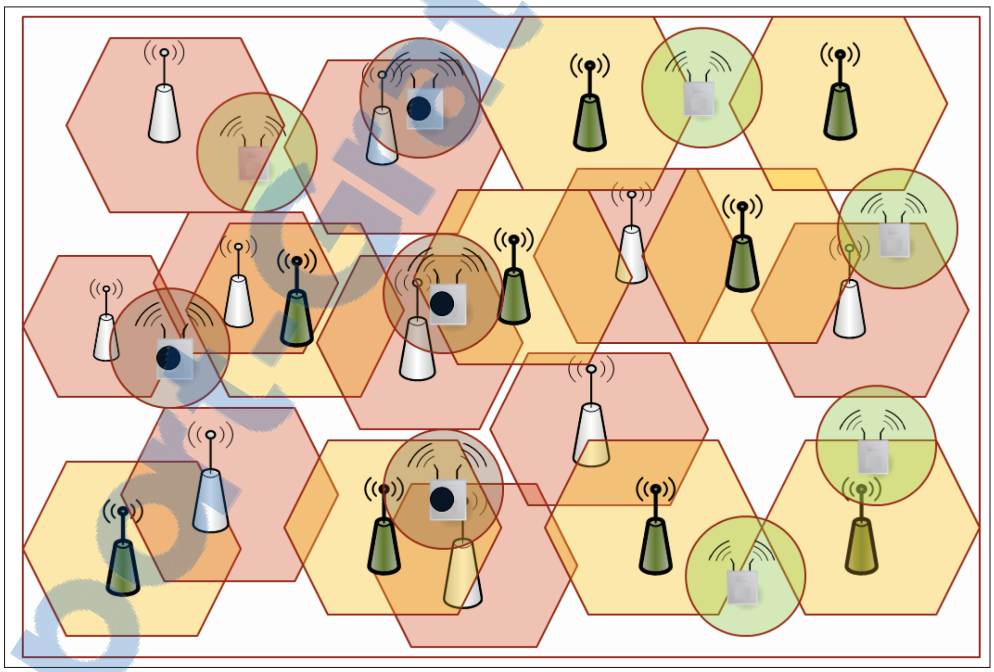

Figure 0.1 A heterogeneous network with two types of access technologies and four providers

We consider the problem of coalition formation as a part of economics analysis of the wireless market. Hence, the pricing strategies and their connection to the network parameters such as data rate and coverage play a key role in our study. From the market viewpoint, there are many internal and external parameters that force users to select a specific provider as their default SP or to choose a particular data plan within a provider's network. As with to any other analysis, one should focus on the key parameters and make a rational simplification of the real market conditions to be able to derive an acceptable and accurate economic framework. To achieve this goal, we start our analysis by focusing on the parameters that affect the decision behavior of subscribers: provider's coverage size, achievable data rate, and the pricing schemes.

0.1.1 The Case of Technological HetNets

For the case of technological HetNets we consider three main technologies in the current market: WiFi, 3G, and 4G cellular networks. Let us consider the general difference between WiFi, 3G, and 4G from the user's viewpoint as follows:

- **Coverage.** Among the mentioned three technologies, LTE and 3G have greater coverage than WiFi. To cover a macrocell coverage area, a WiFi operator needs to implement hundreds of access points. It should be noted that the current service coverage of LTE in the real market is smaller than 3G.
- **Data rate.** New standards of WiFi like 802.11n support up to 300 Mb/s which is twice LTE with 150 Mb/s speed, and it is much better than 3G which gives 15 Mb/s in the best condition.
- **Price**. In a real market, operators charge the highest price for the LTE data usage and the lowest price for the WiFi data usage.

Since users have different expectations of speed, price, and coverage, each user wants to select a provider that offers the best option from the user viewpoint. Most of the times, it is clear for the user how much traffic it is willing to use and how much it wants to pay for that. There is also a general perception about the coverage of each provider when a user wants to buy a service. Naturally, the most expanded coverage along with the highest speed and the lowest price seems the optimum service to all users. But providers have some constraints that prevent them from offering such an ultimate service. For example, a WiFi provider may offer the best speed and lowest price per unit of bandwidth in the current technologies. However, WiFi is using free ISM band as well as limited coverage transceivers which pushes providers to invest a vast amount of their financial resources to cover an area which can be served by just one macrocell (in cellular technology). The cellular technologies also have their issues such as the requirement for buying expensive dedicated spectrum and service equipment.

Since the profit of each provider is related to the number of users who are registered to its network and the average data usage; a provider doesn't have any choice except offering a reasonable service compared to others to attract the preferred number of users and make an incentive for them to utilize its network. Upgrading of services can be carried out in several ways:

- **Investment**. As a straightforward solution, a provider can expand the coverage and make its service better by implementing new equipment. This option may need massive financial resources and comes up with some issues such as making a balance between the service prices and costs generated by new investments. Many providers do not have enough monetary resources to develop their power in the market.
- **Serving users of other providers (roaming capability)**. The second choice for service providers is serving users of each other in the area where the users' primary provider does not have enough resources. This option is one of our considerations for coalition formation which needs the price and resource allocation strategies as well as side payment methods to be defined by the involving entities through a negotiation mechanism that leads to the coalition service agreement.
- **Pooling resources with other providers**. Another choice is turning a coalition into a unified provider by resource pooling. This case is more complicated in technical term because it needs a pricing strategy which is unique to all providers' users and the network

should be seamless even when the connected technology is changed through a vertical handover. Since, in a coalition, the coverage of its members is the same, some users may lose their preference in registering to one special provider's network that is a member of that coalition and this will be a roadblock to achieving an agreement with that provider. In this case, the negotiation is more difficult and the used technology needs more complex network protocols.

Concerning the technological HetNets, the main goal is finding a way for offering a better service to the users and making more profit for the providers by focusing on collaboration methods which can have a significant effect on upgrading network performance. Hence the general problem of our research is finding the ways of forming the coalitions with respect to network technologies, market types and users' preferred services and then reaching a general benchmark that shows provider's position as the consequence of its actions in the market with the existence of coalitions. Also, the stability of formed coalitions should be analyzed.

0.1.2 The Case of Service-oriented HetNets

In a service-oriented HetNet, the type of service can be the connectivity and infrastructure that SP offers to the end users and CPs or a specific type of content that a CP delivers to the end users through the SP's network. In both cases, the profit of providers is highly related to the network selection and content demand behavior of end users. Recently, several major SPs in the US such as AT&T, Verizon, and T-Mobile implemented sponsored data plans in which subscribers can see particular contents from selected CPs without being charged for their data transfer. We consider such plans as a cooperation between CPs and SPs that is in its early stage. However, such cooperation encountered several moral and legal issues regarding the neutrality of Internet. According to the critics, such plans can put the powerful CPs in a position that the smaller CPs cannot compete with. Also, the SPs would monitor the type of data that is being transferred which violates their neutral role based on the traditional philosophy of the Internet. SPs, on the other hand, argue that such plans benefit their subscribers who are simultaneously paying for their network connectivity as well as the content type that they demand. In this thesis, we tackle this problem by finding particular types of contents such as mapping services that can be delivered to the end users completely free of charge. Based on the gathered data from the market, such contents have low-usage pattern, yet they are highly valuable to the end users. This approach requires close cooperation between SPs and all the CPs who deliver the contents. We address this type of cooperation in Chapter 3.

0.2 Objectives

The goal of this research is to develop models that can be used to create more profit for providers by increasing the satisfaction of users. As already indicated, this goal will be pursued by focusing on forming coalitions among existing providers in the wireless market. Investigating cooperation between providers requires a general model of wireless market and users' preferences. This economic model should be developed with respect to all the parameters that can affect the provider's strategies for attracting users. Such a model should predict provider's profit and the number of users before and after forming coalitions. Furthermore, to analyze coalitions, we should develop a game-theoretical approach that predicts the optimal coalition for each provider and define the best strategies that it can choose to maximize its payoff in the stable or unstable coalition structures. The general objectives can be classified as follows:

a. **Defining a model for wireless market economics**

   The first objective is to define a model for wireless market economics. Since the characteristics of current wireless technologies and the economic strategies of providers play a crucial role in providers' profits, analyzing the relation between the wireless technologies and what the users want (like maximum speed and coverage area) requires a model of the market. This model should also cover the concepts of service types and pricing structures so it can estimate the market state in response to any strategy which would be chosen by each provider. In particular the general model of market economics should answer the following questions:

- What are the general market forms and how they change the behavior of providers when they are competing to attract users?
- What is the effect of providers heterogeneity in any specific type of the market?
- For a provider in a wireless market, what are the main decision factors that can affect the a provider selection by users?
- How the interaction between the provider and users can be modeled?
- What approach should be chosen to model the equilibrium state of the market?

b. **Defining a game theoretical model for coalition formation among the providers**

Our research will address the above questions.

The second objective in our research is constructing a framework that defines the process of coalition formation and describes the behavior of heterogeneous providers with different levels of market power, when they want to collaborate. This model should consider the competitive nature of the market and explain the best actions for each provider when it wants to select a desired coalition or deviate from a specific coalition. This model also should estimate the profit of the provider and consider the effect of contracts among coalition members. To achieve such a model, the following questions should be answered:

- What approach is most efficient and practical to form coalitions?
- Are game theory based approaches easy to implement or some alternatives including heuristics should be considered?

- What are the performance and revenue gains that can be achieved by forming coalitions with parallel connections?
- What are the coalition formation processes that can be suitable considering the nature of the wireless market?
- What are the formed coalitions in each defined process?

Our research will address the above questions.

0.3 Originality

The originality of our work relates to two aspects of market economics. The first aspect is studying the practical volume-based pricing strategy that is used by many providers around the world. In Chapter 1, we consider different access technologies to model the relation between spectrum assignment mechanisms and the profit of providers. We propose a method to relate the data usage pattern of subscribers to data rate and service availability. For the first time, to the best of our knowledge, we consider the available budget of subscribers as a random variable and introduce a mathematical framework for SLA-based volume-based data plans. We also model providers with multiple data packages and investigate the package renovation process for the subscribers during their monthly payment period.

The second aspect is the comprehensive study of coalitions in wireless markets in which we consider technological and service-oriented HetNets. Our work in Chapter 2 studies the impact of providers' cost functions on their strategy of coalition formation. We propose multi-provider utility functions for users to study their data usage behavior under cellular-cellular and cellular-WiFi cooperative providers. Unlike many other studies in this field, we build our analysis based the markets with negative externalities. In this way, the convergence of an existing coalition formation process is proved for wireless market. We model the role of regulatory units in coalition formation and analyze their best strategy based on the status quo of the Wireless HetNet market.

Considering the service-oriented HetNets, our work in Chapter 3 is the first work which considers a class of applications which can be offered free of charge to all users based on a cooperation mechanism between CP and cellular SP. To the best of our knowledge, this is the first work that considers an analysis of the payments' directions to provide a completely free access for several types of mobile services such as mapping applications and intelligent personal assistants. Our work shows that the directions of payments in today's cellular market can be altered to increase the satisfaction of end users at no profit-loss for both SP and CP. We also found the side-payment from CP to SP by using the concepts of the Nash bargaining solution as well as the Shapely value that also proves the possibility of our proposed method for selective free content delivery.

0.4 Methodology

This section describes the main methods used to develop and analyze the models of collaboration among wireless providers. These methods belong to three major categories that are related to the following issues: a) characteristics of wireless technologies, b) wireless market model, c) coalition models. In the following subsections, the methods that are useful to solve these issues and to achieve the objectives of this research are described.

0.4.1 Utility functions

Throughout this thesis, we use the logarithmic form of utility functions to model the data usage pattern of subscribers (Chapters 2 and 3). The main reason for to use this type of utility is that it follows the *law of diminishing marginal utility* that is, with extra usage, the slope of utility function decreases. This behavior is highly significant in economic studies since in a long-period analysis, with high data usage, the satisfaction level that subscriber experiences at the end of a period is not the same as it was at the beginning of the period. For the markets which a cap on their data volume is applied, we use the linear form of the utility function (Chapter 1). The reason is that the amount of transferable data in volume-based pricing is not high enough to diminish the marginal utility of the users. Hence, subscribers are always eager to have more data usage. The main parameters in our utility functions are the data usage level, price, coverage size and the given data rate in a service.

0.4.2 Wireless Market Models

Since the expected profit of each provider is affected by the form of its market, knowing the network economics will help to reach a proper model of each market form. These models will be used in the next stage to analyze coalition formation methods. Two major concepts in our model are market forms and pricing strategies which are defined as follows:

0.4.2.1 Market forms

In economic analysis, markets can be categorized as follows:

- **Monopoly**. This form exists when there is just one dominant provider in the market. In this market form, the competition is minimum and the monopolist has total power to set the prices in the market.
- **Duopoly**. In this market form, there are two dominant providers who have the greatest power in the market.

- **Oligopoly**. A small number of providers have the major power in this market form. Each of these entities is willing to know about the strategies of other providers to maximize its profit.

- **Perfect competition**. In this market form, there is no provider with enough power to set the pricing strategies for wireless services. This market form is optimal for the users since the service price is equal to the marginal cost which is equal to the marginal revenue. By this definition, the provider can get no more revenue than its economic cost which means there is no profit in a long run competition.

0.4.2.2 Pricing strategies

Another economic concept that has a significant role in profit calculation and provider's decision-making process is the pricing strategy for the services. Some of the major pricing methods in wireless market are as follows:

- **Flat fee pricing**. This is a pricing structure where provider charges users with a fixed price per data unit. We use this scheme to analyze the technological HetNets in Chapter 3.

- **Dynamic pricing**. In this strategy, the price is related to some other parameters like the time of day (several tariffs for different hours) or the amount of data which is used so far. We use a special case of this pricing scheme to model the volume-based data markets in Chapter 2.

- **Price discrimination**. The aim of this strategy is setting different prices for each class of users. These classes can be based on the users' capability to purchase a service or regional conditions.

- **Congestion pricing**. In this strategy, the users who have a usage higher than what they agreed to with the provider are charged in higher rates. This type of pricing will help providers to control their network utilization by hindering the users that have the data usage bigger than specified I the agreement.

Each of the above pricing strategies can lead to a different form of profit function for providers.

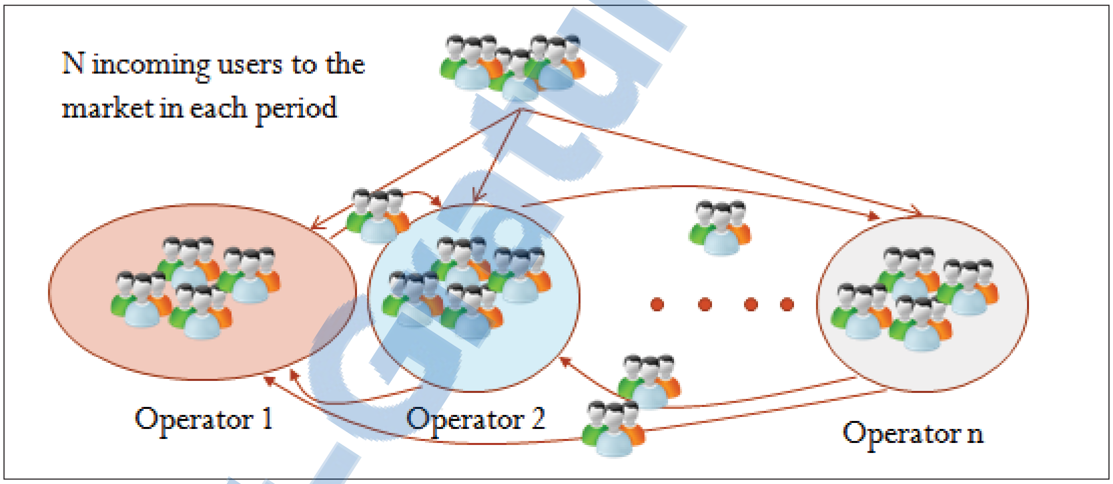

Figure 0.2 The user flow diagram in an unsaturated market

There are other market classifications that are important for this research. For example, the saturation level of the market can be another factor in network analysis. In an unsaturated market, there are a lot of newcomers to the market that should select their providers for the first time (Fig. 0.2). This is in contrast with a saturated market where each provider offers some incentives to attract users being currently with other provider. The saturated market requires analysis of the lock-in effect in the wireless market. This effect binds the user to its current service or provider and implies an extra cost to the user for abandoning its current service. For example, in today's market, many providers offer two-year contracts where the cell phone cost, or its part, is waved. Nevertheless, if a user wants to change its default provider before ending of the contract, there is a penalty fee imposed on the user.

0.4.3 Cooperative game theory

A cooperative game, especially from an economic viewpoint, is using coalition among decision makers to increase their profit. In a competitive market, forming coalitions changes the state of the market from individual competition to coalition competitions. There are two general forms of coalitional games. First, the *canonical cooperative games* (Saad *et al.*) where players want to form a coalition that consists of all players i.e. grand coalition. In this form, the profit division mechanisms that make the grand coalition stable are the main subject of study. The main usage of canonical games' concepts is in analyzing the profit division among the members of a coalition mechanisms such as the core, Shapley value, and nucleolus mechanisms. Since the grand coalition leads to a monopoly market, the regulatory entities do not allow formation of such coalitions. Therefore, it is not practical to investigate such cooperations for inter-provider cooperations. Hence, we exclude discussing such games in this thesis.

The second form of cooperative games is the *coalition formation games* where the structures and processes that force the players to a particular set of coalitions and the stability of these coalition structures are the research subjects.

While in the canonical coalition games the payoff is the most important factor, in coalition formation games the *network structure* and cost of cooperation (Saad *et al.*) are the main factors. A coalition formation game has the following characteristics:

- The game in not necessarily *supperadditive*, which means that the cooperation does not always lead to higher overall profit for the coalition unit. Also, the utility function can be in the form of transferable (TU) or non-transferable utilities (NTU).
- While the coalition forming can provide an additional profit for players, there is also a cost of formation.

- The grand coalition is not always the coalition with the maximum profit.
- Environmental changes like players' strength variation can change the best coalition (Saad *et al.*).

Coalition formation games can be divided into two major subcategories: *static coalition formation games* and *dynamic coalition formation games*. The former analyzes the effect of an external factor on the coalitional structure. The latter investigates the process of forming the coalitional structure (Saad *et al.*). The coalitional game considered in this research is, by its nature, a game with negative externalities, which means that players in the market with any *coalitional structure (CS)* try to reduce other members' profit and maximize their revenue. In our model, the coalition formation process can be *sequential* or takes place in a parallel manner for all negotiators.

0.5 Results

We achieved the following results in this thesis:

a. In Chapter 1, we modeled the markets with volume-based pricing and linked the data usage and price to the offered service data rate based on the utility of users and the available bandwidth to the provider. We found the optimal service parameters for providers with different access technologies such as OFDMA and CDMA. The relation between the available spectrum bandwidth of the provider and the offerable data rate to the users is investigated based on the data cap and price on a data plan. We considered the budget of subscribers as a key parameter and modeled their package renovation procedure. A model for service availability in the dynamic sub-carrier allocation method is proposed in which provider guarantees a data rate and service availability level to the users regardless of their distance from a base station. In this way, we built a mathematical framework that connects the cap of a data plan and its price to the optimal data rate and service availability.

The markets with multiple packages are analyzed in which the provider adjust the cap of each package to address a particular group of users based on their monetary resources. We considered the case of bandwidth splitting in which the provider can assign separated spectrum bandwidths for its voice and data subscribers. We showed the efficiency of our model in giving the optimal market parameters with the help of several realistic numerical scenarios.

b. In Chapter 2, we analyzed technological HetNets. We modeled the markets with flat-rate pricing and found the optimal values of data usage for subscribers based on the data rate, data unit price and coverage of providers. We analyzed several market forms such as monopoly, duopoly, and oligopoly and investigated the effect of competition on the service parameters. We showed that the providers' cost function affects their best strategy to enter a coalition formation process. In particular, we proved that in the case of linear cost functions, the providers are better off to expand their network without cooperating with their competitors. In special forms of exponential costs, providers need to form a coalition to increase their profit, otherwise, investing on network expansion is not profitable for them. We proved the form of a multi-provider utility function which is used to model the behavior of subscribers when they are under the coverage of a coalition of providers. We modeled the multi-provider utility functions for cellular-cellular and cellular-WiFi coalitional models. The profit of the providers is analyzed based on the usage patterns in the multi-provider model.

We used an existing coalition formation process which has a mathematical base for the markets with negative externalities. We proved that in wireless markets in which users consistently churn to newer technologies, a coalition formation process is always convergent. We provided a model for the role of regulatory units in coalition formation processes and proposed a method that finds the best cooperation strategies for increasing the social efficiency. The regulatory entity uses this method to allow or ban a coalition formation action. Various scenarios for coalition formation are analyzed in which we showed that the cooperation of small providers is an efficient way to compete with a monopolist. The efficiency is measured based on a social efficiency function that balances the overall payoff of the users as well as the profit of the providers.

c. In Chapter 3, we analyzed service-oriented HetNets. We investigated the statistics of real mobile markets provided by (Ericsson, 2016) and defined three categories of mobile applications based on the usage patterns of mobile subscribers. We found applications with particular business models that can be offered free of charge to all subscribers without being concerned about their data usage in those applications. We call this model a selective free content (SFC) program. Three categories of such applications are investigated: the Mapping applications along with intelligent personal assistants, the cloud-based IoT services, and the smart city and e-governance applications. We showed the difference between these categories by analyzing the direction of payments in each category that cause different business models. We showed the possibility of an SFC program by modeling the interaction between users, service and content providers in a wireless market.

   A three-stage Stackelberg game is introduced and solved by backward induction. Each stage of the game shows the best response strategy of one network entity. The side-payment from CP to SP is found based on the Nash bargaining solution as well as the concept of Shapely value. Our model showed that even with a linear profit model for the CP, an SFC program is possible. We found the profit threshold of SP in which an SFC program is possible. Several realistic numerical scenarios are analyzed, and the side-payment from CP to SP is found based on different bargaining powers of SP over CP.

0.6 Content

This thesis follows the paper-based format of ÉTS. Hence, each chapter represents one journal paper as follows:

a. Chapter 1 analyses the wireless market with volume-based pricing. The related paper is submitted to IEEE Transactions on Network and Service Management (TNSM).

b. Chapter 2 investigated the competition and cooperation in technological HetNets. The related paper is submitted to the Elsevier Computer Networks journal.

c. Chapter 3 studies the selective free content program in wireless markets and considers the coalition between a SP and a CP. The related paper is submitted to IEEE Transactions on Mobile Computing (TMC).

d. Finally, in Conclusion and Recommendations, we summarize the results of this thesis and proposed several directions as the possibilities of future studies.

0.7 Literature Review

Collaboration has emerged as a new paradigm that can have a significant effect on the network performance in several layers. In particular, the concepts from game theory, such as cooperative games and coalitions, are used to optimize and improve resource utilization while providing a fair distribution of the gains among game participants. Considering the theoretical works related to the concept of coalition formation, authors in (Hart & Kurz, 1983) created an endogenous framework for coalition formation. However their method does not consider the effects of externalities on CS. Another good example of coalition formation is (Yi, 1997) which considers both externalities and endogenous nature of coalition formation. However, they solely analyzed symmetric games. Since the wireless market is asymmetric, with negative externalities and different provided services, it seems that a method based on *relative power* (e.g., Shapley value) of providers is the best way for profit division within a coalition. In (Bloch, 1996), a sequential stationary game for coalition formation process is proposed. In this model, players announce their desired coalition based on an exogenously defined order. The players in the game can have asymmetric power. It seems that such model is the best option for analyzing cooperations in the wireless market. Hence, in Chapter 2, we use this model to investigate coalition formations in the wireless market.

In the context of wireless networks, the majority of the studies focused on cooperation among the nodes or users and mainly address the physical and link layers, e.g., (Mathur *et al.*, 2008), (Zhang *et al.*, 2013) and (Chan *et al.*, 2013), and the works that consider network layer as well, (Han & Liu, 2008). However, there is a scarcity of research in the area of network operators' collaboration. The pioneering work for non-wireless network operator coalitions is presented in (Gibbens *et al.*, 1991) where coalitions of national operators are analyzed in a game theoretic framework for international routing. The results show that capacity saving of the order of 20 percent is achievable by forming coalitions. Coalition of network operator is considered in (Sarkar *et al.*, 2008) and (Singh *et al.*, 2012a). In (Sarkar *et al.*, 2008) the spectrum pooling in wireless data access networks is studied in a given geographical region and the outcome shows that the grand coalition of all operators is stable and maximizes the profit. This work is extended in (Singh *et al.*, 2012a) to multi-hop networks with more relaxed assumptions allowing variability of channel characteristics and mobile locations. In both mentioned bodies of work, the main assumption is to form a grand coalition which simplifies the analysis, but it is not practical. The impracticality of the grand coalition comes from the reality of markets; in most countries the competition rules prohibit the formation of a grand coalition as it leads to a virtual monopoly. Therefore what needed is a study of how to form coalitions under realistic constraints where the optimal solution cannot be a grand coalition, but instead, it can be a set of coalitions where neither of them is a grand coalition. As already indicated, such games can be named coalition formation games as opposed to canonical coalitional games dealing with grand coalitions. To the best of our knowledge, there are no studies of coalition formation games in the context of cooperation between heterogeneous wireless network operators.

Considering the service-oriented HetNets, the majority of literature focuses on sponsored data in non-neutral networks which eventually leads to a type of CP-SP cooperation. (Lotfi *et al.*, 2016) investigates the profitability of non-neutral networks and shows that in certain scenarios a non-neutral network is nonprofitable. Also, it shows that when the market power of an SP is small, the end users can obtain a better overall payoff in a non-neutral regime. (El-Delgawy & La, 2015) analyses the interaction between a CP and SP when SP agrees to offer a better QoS to CP's service. This can be seen as a coalition of the two providers which is achieved by an agreement in a bargaining game. The effect of bargaining power on the QoS and the social efficiency level are the two factors which are considered in this work. (Joe-Wong *et al.*, 2015) investigates the optimum amount of content that CP sponsors. It shows that with sponsored data applied, the utility of users increases more than CPs. (Andrews *et al.*, 2016) considers a case in which an SP proposes a sponsored data service to several CPs. In this case, SP aims to select one of the CPs for the offered service and to determine the service price that maximizes its profit. One of the main issues that is addressed in this work is the truthfulness of CPs when they report their network parameters. In this thesis, similar to above works, we consider the concept of non-neutral networks as a way to increase the profit of CPs and SPs. However, we propose a method in which certain types of contents can be delivered to all users free of charge without requiring SPs and CPs to limit the content usage of subscribers. Our method is based on the nature of such contents which carry highly valuable information with low usage pattern. We call this method a selective free content (SFC) program and will introduce it in Chapter 3.

CHAPTER 1

PRICING THE VOLUME-BASED DATA SERVICES IN CELLULAR WIRELESS MARKETS

Behdad Heidarpour [1], Zbigniew Dziong [1], Wing Cheong Lau [2], Shahin Vakilinia [3]

[1] Département de Génie Électrique, École de Technologie Superieure ,
1100 Notre-Dame Ouest, Montréal, Québec, Canada H3C 1K3
[2] Department of Information Engineering, The Chinese University of Hong Kong,
Sino Building, Chung Chi Rd, Sha Tin, Hong Kong
[3] Laboratory for multimedia communication in telepresence, École de Technologie Superieure, 1100 Notre-Dame Ouest, Montréal, Québec, Canada H3C 1K3

Submitted to IEEE Transactions on Network and Service Management in April 2017.

Abstract

Over the past few years, many major wireless providers restricted their unlimited data plans and replaced them with limited-size fixed-price data packages. While this could be perceived as a disadvantage for customers, it helps the cellular wireless providers to reduce the traffic intensity at their base stations. Then the lower traffic intensity leads to a better service quality and higher rates for concurrently connected users. Hence, there is a trade-off between the data volume and the data rates attributed to the users. Therefore, to avoid the adverse effect of service inaccessibility, the cellular providers should carefully set the size and pricing of their data packages. To do that, the providers need a model that, together with proper market information, would allow to set the best prices for volume based data and estimate the acceptable quantity of subscribers along with average data rate for them. In this paper, we propose such a model that quantifies the relationship between pricing and various market/system parameters such as data volume size, user budget, data rate and service blocking probability. In particular, we formulate a set of profit optimization problems for different spectrum assignment criteria like shared-carrier and dynamic sub-carrier allocation. Finally, several realistic scenarios are investigated in which the optimal network parameters are computed.

Introduction

With rapid growth of wireless markets in past twenty years, providers have been evaluating different pricing schemes to maximize their profit. In particular, the providers of packet-switching networks have considered many ways regarding pricing criteria to achieve a higher amount of net income. These schemes are highly diversified from simple flat-rate (Courcoubetis & Weber, 2003b) to dynamic pricing methods (Viterbo & Chiasserini, 2001). Schemes that consider Service Level Agreement (SLA) are not common for the end-users in data networks, yet they are essential in mid-level and high-level inter-provider contract based methods. Smart pricing methods (Sen *et al.*, 2013b), such as time-dependent pricing schemes (Ha *et al.*, 2012a), in which the provider sets the price based on congestion hours, or other network parameters, are studied in several works but they are not widely accepted by the wireless providers. This is not only due to the operational complexity of such dynamic pricing schemes but also due to the difficulties in managing customer expectations and educating them on complicated interaction between user behavior and pricing.

Before 2012, it was common among major players in the cellular wireless market (such as Verizon, AT&T, and Sprint) to offer low-price unlimited data plans. However, nowadays their pricing schemes are dominated by plans with a cap on data volume and calling minutes. In this approach, instead of unrestricted access to the data service, the subscriber pays a certain price to use up to a specified amount of data alongside the voice service for a particular duration. For example, the subscribers pay for one, two or more Gigabytes of data at a particular price. The available data in the package is being refilled mostly in monthly periods and the user-provider contract usually stays unchanged for over one or two years. We refer the readers to (Verizon-Data, 2016; AT&T Data, 2016; Sprint, 2016) for some examples of U.S wireless carriers and (Telus, 2016; Bell, 2016; Rogers, 2016) for Canadian providers. The universal dominance of this pricing scheme motivates us to investigate its characteristics and optimality under different wireless technologies with shared, dedicated or dynamic spectrum allocation policies.

Analytical framework

We aim to find a model that can predict the profit of providers with volume-based data plans. The key elements of our model are the market and network parameters such as data-volume size, price, data rate, service availability and user budget. The combination of these parameters affects user's subscription behavior. Similar to many real markets, the voice and data services are assumed to be offered in separated packages. The set of data users is a subset of voice users which means that if a user wants to have a data subscription, it needs to join the voice network as well. In this way, we present the market in two stages; firstly, the voice users enter the market out of a set of potential subscribers. Secondly, the newly joined users are offered to have a data package which eventually forms a set of data users. From the user side, the widely adopted utility theory (Duan *et al.*, 2013c; Gajic *et al.*, 2009; Acemoglu *et al.*, 2004; Niyato & Hossain, 2009) is used to model the behavior of users in package selection. The cap of data-volume, price, data rate and service availability are the variables in the utility function.

We start developing our formulation by analyzing the optimal profit of a provider with a single data package and shared-carrier system. For this case, we show the relation between optimal data rate and the size of data volume. We then extend the framework to support a multi-package provider model. In this part, the effect of users' monetary resources and their budget distribution on provider's profit are investigated. The proposed analysis enters the next level by defining a second method in which the provider dynamically assigns the bandwidth to the users to guarantee a constant predefined data rate. Since the spectrum is limited in cellular networks, the majority of analysis in this part is focused on calculating the blocking probability of data flow requests originated by spectrum shortage. Then, we introduce the provider's profit optimization problem in which the optimal values of data-volume, price of data plan, as well as the data rate, are calculated. Here, to have a precise structure, separated spectrum bands are defined for voice and data. Hence, the bandwidth splitting ratio is an additional decision value for the optimization problem. Lastly, several numerical studies related to the optimal values of data cap, price, and data rate as well as service blocking probability are provided. These scenarios present real world market situations with current 3G/4G technologies that support shared-carrier and dynamic bandwidth allocation methods.

Contribution

In summary, the technical contribution of this paper include:

- We formulate a profit optimization problem to model and quantify the effect of the size of data volume and its corresponding pricing on expected user data rate and provider profit.
- Our formulation covers two methods of spectrum allocation: shared-carrier method that covers the TDM and CDMA schemes, and dynamic spectrum allocation method that covers OFDMA systems.
- We consider user's available budget as a random variable and use it to derive the number of subscribing users. The optimization problem in this part covers multi-package markets. Also, the case of data volume renovation in which users extend the data-cap under their default plan is analyzed.
- We link the offered data volume size, price and service availability together with a utility function and calculate the optimal service data rate that maximizes provider profit.

Structure

The rest of this paper is organized as follows: Section 1.1 is a brief review of related works in wired and wireless network economics. The general system model is described in Section 1.2. Single and multi-package wireless markets regarding the shared-carrier model are analyzed in Section 1.3. Pricing in dynamic spectrum allocation method is the subject of Section 1.4. Section 1.5 includes several numerical studies similar to real markets based on shared-carrier and dynamic bandwidth allocation methods. Finally Section 1.6 concludes the paper.

1.1 Related Works

Economics of data networks is a well-established branch of network analysis in both engineering and economic divisions. Considering the Internet as a new opportunity to make wealth and capital, MacKie-Mason and Varian (MacKie-Mason & Varian, 1995) is one of the early works in this field investigating the Internet pricing schemes. Considering the cellular networks, (Wright, 2002) is an example of pure economic analysis of call pricing among competitive network carriers; in particular, the class of pricing for calls from fixed-network to cellular subscribers is analyzed. (Huang, 2008) is another work related to the pricing and call demands in mobile networks; this work considers different plans, each having a non-linear pricing scheme and the network demand is computed based on econometric models. The effect of cellular competition and the entrance of new competitors on pricing tactics is the subject of (Seim & Vaiard, 2011). The analysis is based on the data of U.S cellular market in the late 90s, and its main interest is the effect of market structure on pricing schemes and the optimality of diversified pricing methods on overall consumer satisfaction. (Zhang *et al.*, 2014) investigates the service pricing in two-tier small cell networks. This work proposes a paying mechanism in which the macro-cell providers incentivize the small cell owners to give access to macrocell users. The pricing strategies are based on a leader-follower dynamic game. Dynamic pricing of call rates is analyzed in (Dugar *et al.*, 2015). This work concerns two models of game: provider vs. provider and consumer vs. provider. Based on these models, the optimal strategies of network entities are analyzed. The dynamic price of the system is calculated based on several factors including the available bandwidth. However, the data service is not considered in this work.

With regards to the QoS of cellular networks, (Hou *et al.*, 2002) focuses on congestion control by combining pricing schemes and admission control algorithms. It considers an optimal call arriving rate which maximizes the provider profit and user utility and uses an adjustable service price based on network parameters. (Yilmaz & Chen, 2009) uses the admission control tactics to maximize the profit of provider by setting the optimal prices while QoS is set for each service class of the network. The main QoS parameter considered in this work is call dropping probability. Different admission control algorithms are also analyzed in this work.

For the current 3G/4G technologies, there are several pricing recommendations in literature, e.g., (Wallenius & Hämäläinen, 2002) which suggests QoS based pricing methods which are aware of service classes. For the further reading, we refer the readers to (Ezziane, 2005) and (Gizelis & Vergados, 2011) which are surveys on 3G and wireless networks pricing.

This work is different with all above research efforts by considering voice and data packages as separate options. It also formulates the relation between offered data size, its price, and network parameters as well as the number of potential subscribers.

1.2 Notation and System model

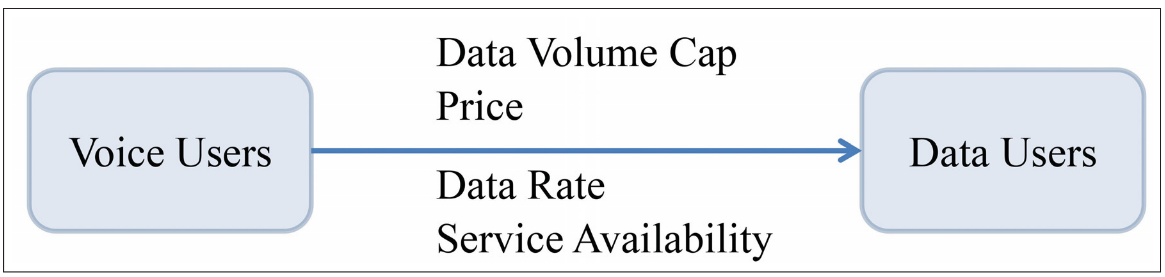

Figure 1.1 The subscription flow of the potential users. They join the voice cellular network and then would subscribe to data package

In this section, notation alongside the system model are introduced. We consider a cellular provider's network with several macrocells. The provider aims to offer a data plan in which a certain level of QoS, defined in an SLA, is provided throughout its network. Namely, in the busy hours of busiest macrocells, predefined levels of voice service availability and minimum data transfer's rate are provided. Based on the direct measurements of user distribution over the time, the provider expects a maximum of N_C voice users under the coverage of its busiest macrocells in the busy hour. N_C can be the maximum numbers measured or a threshold value that is exceeded with a small probability; this probability can be easily integrated into the SLA. The provider tries to attract a subset of voice users to purchase its data plan. We represent the number of data users associated with N_C as N_D. The service prices for voice and data plans are denoted as p_c and p_d, respectively. The cap of the data volume is represented by

c_d. The offered data rate is represented by γ_d. The main challenge is to model the behavior of users to optimize the network parameters, such as the size of data-volume, price and the offered data rate, to achieve the maximum profit. In particular, we want to maximize the profit function $\pi(c_d, \gamma_d, p_d) = N_D(c_d, \gamma_d, p_d) \times p_d$. We formulate this model of subscription backwardly. First, the user response is modeled based on the network parameters. Then, after finding N_D as a function of N_C, we find the final profit function of the provider. The perceptual importance of data service is modeled by assigning a random user data valuation number ω_d to each user. This random value emphasizes the significance of having data access for each user. For the sake of simplicity, it is assumed that ω_d is uniformly distributed in interval $[0,1]$, $\omega_d \sim U(0,1)$. In general, a Beta distribution can be assumed for ω_d. We will use Uniform as well as Beta distributions in our numerical examples. Each macrocell's coverage area is assumed to be circular with radius R_M and users are uniformly distributed in the cell area. Hence, for each user the random distance from the base station has the following distribution:

$$f(r) = \begin{cases} \frac{2r}{R_M^2} & r \leq R_M, \\ 0 & r > R_M. \end{cases} \tag{1.1}$$

In the rest of the paper, $P(x)$ represents the probability of x and $f(x)$ and $F(x)$ operators represent the probability distribution function (PDF) and cumulative distribution function (CDF) of the variable x respectively.

1.2.1 Data rate models based on channel access method

User satisfaction and preferences are playing a significant role in the package selection. To this extent, we need to model the fundamental parameters involved in the selection process. One of these parameters is the data rate (or the service speed) granted to the user. In this section we present data rate models for two different channel access methods along with admission policy behind them and their impact on user satisfaction. The analytical models for other channel access methods can be extracted from these two. First, we consider shared single carrier methods that covers schemes like CDMA and WCDMA. In this case, voice and data services can be offered by one carrier with time division or two separated carriers each having a fraction of total bandwidth. Second, we consider dynamic sub-carrier allocation in which provider tries to guarantee a predefined level of data rate to the users by allocating the proper portion of its available bandwidth. Here the main concern for the provider is the value of guaranteed data rate and also the blocking probability of data flows caused by spectrum shortage. Fig. 1.2 illustrates these two types of bandwidth assignment. Since the focus is on service pricing, we consider the high-level models of access methods which can be extended to different scenarios.

Considering the data rate, our model relies on a general path-loss definition; supposing the maximum achievable data rate by each user j in technology X in absence of other users is C_X^M, then the distance related rate is represented by:

$$C_X(r) = C_X^M \begin{cases} 1 & r \leq R_m, \\ \left(\frac{R_m}{r}\right)^{\alpha} & R_m < r \leq R_M, \end{cases} \tag{1.2}$$

where R_m is the maximum radius in which the highest rate can be achieved and $\alpha \geq 2$ is the attenuation factor. Generally, C_X^M is the technology dependent rate which is related to the carrier bandwidth Q_D by a *spectral efficiency* factor (ElNashar *et al.*, 2014), $C_X^M = \beta \times Q_D$. Using this data rate definition, in the next two subsections the achievable data rate for these two spectrum allocation methods are elaborated.

1.2.1.1 Shared single carrier

The first formulation is for shared single-carrier method where each user's data rate originates from code or time division multiplexing. In (Bonald & Proutière, 2003), a framework for CDMA networks is proposed That is based on representing a single-carrier macro-cell by a *processor sharing* model. In this model, if the arrival rate is λ and each user has a flow request

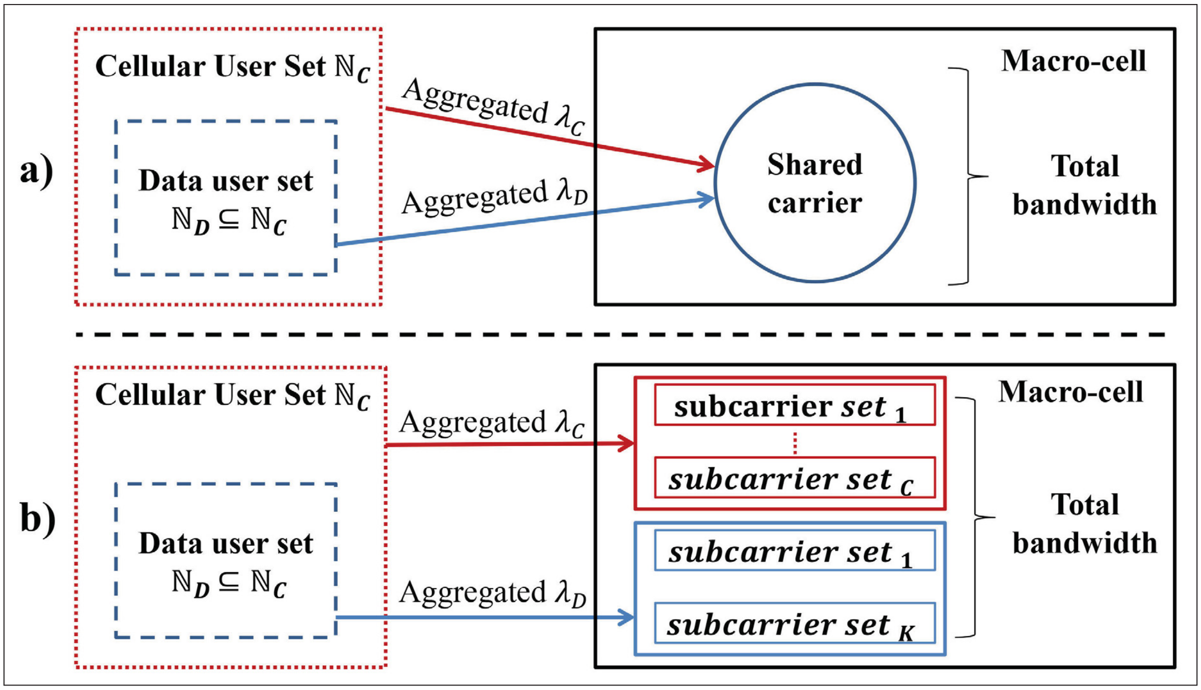

Figure 1.2 Two types of bandwidth assignment. a) Shared carrier model for voice and data. b) dynamically assigned sub-carrier sets for cellular voice and data

of expected size η, then the average service time $\frac{1}{\mu}$ for a user can be calculated as:

$$\frac{1}{\mu} = \int_0^{R_M} \frac{\eta}{C_X(r)} f(r)dr = \frac{\eta}{C_X^M} \underbrace{\left(R_m + \frac{1}{R_m^{\alpha}(\alpha+1)}(R_M^{\alpha+1} - 1) \right)}_{\zeta_1} = \frac{\eta \times \zeta_1}{C_X^M} = \frac{\eta \times \zeta_1}{\beta \times Q_D}. \quad (1.3)$$

In our model, we relate the arrival rate of data flows of each user (λ_d) to the offered data volume cap by a scale factor λ_u: $\lambda_d = \lambda_u c_d$. This simplification is mainly due to the rational behavior of users that try to utilize the network based on their package limits. It is proved ((Tijms, 2003) P.6) that the combination of N_D users leads into a Poisson arrival rate of $\lambda_D = N_D \times \lambda_d$. This distribution is generally called a *merged Poisson process*. The average load of processor sharing system is $\bar{\rho} = \frac{\lambda_D}{\mu} = \frac{N_D \times \lambda_u \times c_d \times \eta \times \zeta_1}{C_X^M}$. It is clear that $\bar{\rho} < 1$ is the condition of stability. In the stationary state of the system, the data rate for a user which is located in distance r from

Table 1.1 General notation

Parameter	**Description**
$\mathbb{N}_C$	Set of voice users
$\mathbb{N}_D$	Set of data users
N_x	size of the set $\mathbb{N}_x$
Q_T	Total available bandwidth
Q_C, Q_D	Total available bandwidth to voice and data services
c_d	data volume cap
p_d, p_c	price of data and voice services
γ	general notation for data rate
γ_c,γ_d	offered data rate to voice and data users
β	Spectral efficiency
θ_d	leveling factor for user's utility
ω_d	valuation factor for the data plan
b_g	random variable of user budget for data service
η	expected size of data flow
λ_c	service request rate of each voice user
λ_u	scale factor for the rate of incoming data flows
λ_d	rate of flow requests for each user $\lambda_d = \lambda_u c_d$
λ_D	overall service request rate of all voice data
$\frac{1}{\mu_c}$	voice service time
B_c	expected blocking probability of voice calls
B_d	expected blocking probability of data flows

the macrocell is (Bonald & Proutière, 2003):

$$\gamma(r) = C_X(r)(1-\bar{\rho}), \tag{1.4}$$

hence the expected data rate is,

$$\bar{\gamma} = \int_0^{R_M} \gamma(r) f(r) dr = C_X^M \left(R_m + \frac{1}{1-\alpha} \left(R_M \left(\frac{R_m}{R_M} \right)^{\alpha} - 1 \right) \right) (1-\bar{\rho}). \tag{1.5}$$

By setting $\zeta_2 = R_m + \frac{1}{1-\alpha} \left(R_M \left(\frac{R_m}{R_M} \right)^{\alpha} - 1 \right)$,

$$\bar{\gamma} = \zeta_2 \times \left(C_X^M - N_D \times \lambda_u \times c_d \times \eta \times \zeta_1 \right), \tag{1.6}$$

which is used as the equation for the data rate for the shared single carrier approach.

1.2.1.2 Dynamic Bandwidth Allocation

In the second model, we suppose that the provider can dynamically assign a fraction of its available spectrum to each user's data flow to guarantee a minimum level of data rate during the busy hours. In this manner, the provider aggregates several sub-carriers based on channel quality to guarantee the promised quality. Let have γ_d as the offered data rate to the user and $S(r)$ as the bandwidth required to achieve γ_d in distance r from the base station. Considering the spectral efficiency factor which is technology-dependent we can map the maximum rate to the bandwidth as $\gamma_d = \beta S_d$, where S_d is the bandwidth of assigned spectrum. Using this definition in (1.2) gives,

$$\gamma(r) = \beta \times S_d \begin{cases} 1 & r \leq R_m, \\ \left(\frac{R_m}{r}\right)^{\alpha} & R_m < r \leq R_M. \end{cases} \tag{1.7}$$

To achieve a constant value of γ_d, the provider needs to allocate the following bandwidth:

$$S(r) = \frac{\gamma_d}{\beta} \begin{cases} 1 & r \leq R_m, \\ \left(\frac{r}{R_m}\right)^{\alpha} & R_m < r \leq R_M. \end{cases} \tag{1.8}$$

The above equation is the basic model for the required spectrum size (in forms of grouped sub-carriers or any dynamically allocated spectrum) for each user based on its distance. Regarding the admission policy, the provider can make two separate groups for data and cellular voices, each having different required data rate. Since the available spectrum is limited, in the busy-hour a data flow transfer can be blocked or delayed. Thus, we need to define the blocking probability of data flows and a suitable package selection mechanism based on guaranteed data rate and blocking probability of data flows. We will go deeper into this particular case in Section 1.4.

In the next two sections, we define the volume-based pricing framework. In Section IV we develop our formulation based on shared carrier scheme. Then, we expand the concepts to adopt the dynamic spectrum allocation and related blocking probabilities in Section 1.4.

1.3 Data package pricing

1.3.1 Single-package problem

The provider's profit maximization problem can be divided into two stages. In the first stage provider sets the data service parameters such as the data volume cap, its price and data rate and offers the plan to the voice users. In the second stage, users decide whether to join the data plan based on the service parameters. This model can be solved backwardly. Namely, the provider anticipates the users' subscription behavior and uses this information to set the optimum values of service parameters. In the following subsection, we describe all the necessary equations to model users' decision criterion.

1.3.1.1 Users' decision criterion

We start with the formulation of the one data package case and then extend it to the multi-package counterpart. A volume-based data package is presented by a specific amount of data c_d with price p_d that is valid for a predefined period, e.g., a month, that can be renewed at the beginning of each period. In the package information the provider announces the following network parameters:

- Average data rate that can be achieved by a user under network coverage area, which is the expected data rate from (1.6).
- The size and price of data package.

Each user applies its own evaluation criterion on service parameters to assess the package desirability. To model this decision behavior, we adopt the general practice of utility theory

(Duan *et al.*, 2013c; Gajic *et al.*, 2009; Acemoglu *et al.*, 2004; Niyato & Hossain, 2009). In our case, the utility of user j is represented by:

$$U^j = \theta_d \times \omega_d^j \times \gamma_d \times c_d - p_d. \tag{1.9}$$

where γ_d is the promised average data rate to the user and θ_d is a leveling factor which determines the combinations of [data volume size, price] that give zero utility for users with the highest data valuation, $\omega_d = 1$, and the network with normalized access data rate of 1. Data valuation is a random variable $\omega_d \sim U(0,1)$. In this decision process, if the value of U^j is bigger than a reserved utility ε_d, then the user j will subscribe to the data network. In this manner, the set of data users $\mathbb{N}_D$ can be defined as follows:

$$\mathbb{N}_D = \left\{ j \in \mathbb{N}_C \,\middle|\, \omega_d^j \in [0,1] \wedge \omega_d^j \geq \frac{\varepsilon_d + p_d}{\gamma_d \theta_d c_d} \right\}. \tag{1.10}$$

Note that the applied linear form of user utility is considered in previous literature as well (e.g. (Chen *et al.*, 2015a)); we use it due to the nature of volume-based pricing in which the bigger ratio of data to price the higher level of satisfaction.

1.3.1.2 Provider profit

The provider needs to optimize its profit function $\pi = N_D \times p_d$. The optimization function has two constraints. Having a stable system requires $\bar{\rho} < 1$ and also the average data rate should be bigger than the offered value, $\bar{\gamma} \geq \gamma_d$. We can rewrite $\bar{\gamma}$ in (1.6) based on the value of N_D:

$$\bar{\gamma} = \zeta_2 \times \left(C_X^M - \lambda_D \times \eta \times \zeta_1\right) = \zeta_2 \times \left(C_X^M - N_C \lambda_u c_d \zeta_1 \eta \left(1 - \frac{\varepsilon_d + p_d}{\gamma_d \theta_d c_d}\right)\right). \tag{1.11}$$

The profit function is strictly increasing with respect to γ_d which means if the provider can achieve a data rate value like $\bar{\gamma} > \gamma_d$, then it is better off to announce $\bar{\gamma}$. Hence we can set

$\gamma_d = \bar{\gamma}$ and solve (1.11) for it:

$$\bar{\gamma} = \frac{\zeta_2 C_X^M - \sigma c_d \pm \left((\zeta_2 C_X^M - \sigma c_d)^2 + 4\frac{\sigma}{\theta_d}(\varepsilon_d + p_d) \right)^{\frac{1}{2}}}{2}, \tag{1.12}$$

$$\sigma = N_C \zeta_1 \zeta_2 \eta \lambda_u.$$

Eq. (1.12) has two opposite sign values, where the positive value is the only acceptable solution. We can write the optimization problem of data profit as:

$$\max_{c_d, p_d} \pi(c_d, p_d) = N_C p_d \left(1 - \frac{\varepsilon_d + p_d}{\bar{\gamma}(p_d, c_d)\theta_d c_d} \right) \tag{1.13}$$

$$\bar{\gamma}(p_d, c_d) = \frac{\zeta_2 C_X^M - \sigma c_d + \left(\left(\zeta_2 C_X^M - \sigma c_d\right)^2 + 4\frac{\sigma}{\theta_d}(\varepsilon_d + p_d) \right)^{\frac{1}{2}}}{2}$$

subject to:

$$p_d \geq 0, \tag{1.14}$$

$$c_d \geq d^m, \tag{1.15}$$

$$\bar{\gamma}(p_d, c_d) \geq \gamma^m. \tag{1.16}$$

The constraint $\varepsilon_d + p_d \leq \bar{\gamma}(c_d, p_d)\theta_d c_d$, which assures the user set $\mathbb{N}_D$ has a rational size and also validates $(\max \bar{\gamma}) \leq \zeta_2 C_X^M$, is not required since otherwise the objective function obtains negative values. The maximum profit does not go to the negative side as $p_d = 0$ is giving a nonnegative maximum in the worst case. For the QoS conditions and also for regulator's minimum service obligations, we added the inequality $\bar{\gamma}(p_d, c_d) \geq \gamma^m$ as the constraint for the lower-bound of data rate and $c_d \geq d^m$ for the minimum offered data. It can be proved that the objective function is concave with respect to both p_d and c_d. We refer the readers to Appendix I-1 for the proof.

1.3.2 User monetary resources and budget

Until now we investigated a scenario in which there is no information on monetary resources of potential subscribers, nor their budget, hence, each subscribed user purchases the package according to its valuation. We distinguish between the financial capacity (or wealth) and the budget. The former shows the user ability to purchase a package but does not necessarily mean its service evaluation meets the ability to purchase. One example is a user who values the high-speed data service and large data volume cap, yet it cannot afford the cost, or on the contrary, another user who has enough funds to purchase any package but accessing the data itself is not essential for him. Mathematically speaking, the correlation between valuation and the available funds is not perfectly positive. The latter concept of budget defines the amount of money each user is willing to pay if the package meets its minimum expectations. In particular, since this type of information shows the usage willingness, data valuation ω_d can be replaced by the budget information if it is presented. We denote the wealth random variable by m_r and the budget by b_g. In the next subsection, we extend our formulation to this type of information in the multi-package market.

1.3.2.1 Multi-package data network with wealth information

In the previous subsection, the data network subscription is defined based on a uni-package provider. Even though this is a fair way to investigate user behavior under different data sizes, all users who have a data evaluation under a threshold limit are excluded from data user set and this is not optimal for the provider. Therefore, it is a common practice for providers to construct a multi-package service in which each set of [data volume size, price] addresses a unique group of users. Suppose the provider has n data packages represented by the set

$\upsilon = \{\{c_d^1, p_d^1\}, \cdots, \{c_d^n, p_d^n\}\}$ where

$$p_d^1 < p_d^2 \cdots < p_d^n, \tag{1.17}$$

$$c_d^1 < c_d^2 \cdots < c_d^n, \tag{1.18}$$

$$\frac{\varepsilon_d + p_d^1}{c_d^1} > \frac{\varepsilon_d + p_d^2}{c_d^2} > \cdots > \frac{\varepsilon_d + p_d^n}{c_d^n}, \tag{1.19}$$

are the rationality requirements for the package pricing. Comparing all packages, the highest data volume attracts the potential subscribers with lowest data valuation and above; As a result, user wealth has the main role in selecting between the preferred packages. Let define the joint-PDF $f(\omega_d, m_r)$ based on market gathered information; then if we represent the subscribers of package i with N_D^i, we have:

$$N_D^i = N_C \begin{cases} P\left(\omega_d \geq \dfrac{\varepsilon_d + p_d^i}{\bar{\gamma}\theta_d c_d^i}, p_d^i < m_r \leq p_d^{i+1}\right) & i \neq n, \\ P\left(\omega_d \geq \dfrac{\varepsilon_d + p_d^i}{\bar{\gamma}\theta_d c_d^i}, p_d^i < m_r \leq M_r^M\right) & i = n. \end{cases}$$

where M_r^M is the highest amount of money someone would pay for data access. We can represent the above equation in terms of cumulative distribution function; for $i \neq n$ we have:

$$\underset{i \neq n}{N_D^i} = N_C \left[F_{M_r}\left(p_d^{i+1}\right) + F_{\Omega_d, M_r}\left(\frac{\varepsilon_d + p_d^i}{\bar{\gamma}\theta_d c_d^i}, p_d^i\right) - F_{\Omega_d, M_r}\left(\frac{\varepsilon_d + p_d^i}{\bar{\gamma}\theta_d c_d^i}, p_d^{i+1}\right) - F_{M_r}\left(p_d^i\right)\right],$$

where the CDF parts come from the definition of joint probability as $P(X_1 < x \leq X_2, Y_1 < y \leq Y_2) = F_{XY}(X_1, Y_1) + F_{XY}(X_2, Y_2) - F_{XY}(X_2, Y_1) - F_{XY}(X_1, Y_2)$ and $P(x \leq \infty, y \leq Y_1) = F_Y(Y_1)$.

The exact same concept can be used for N_D^n. Here the profit function and $\bar{\gamma}$ are defined as:

$$\pi(\upsilon) = \sum_{i=1}^{n} N_D^i p_d^i, \tag{1.20}$$

$$\bar{\rho}^i = \frac{N_D^i \lambda_u c_d^i \eta \zeta_1}{C_X^M}, \tag{1.21}$$

$$\bar{\rho}\Big(\upsilon, \bar{\gamma}(\upsilon)\Big) = \sum_{i=1}^{n} \bar{\rho}^i, \tag{1.22}$$

$$\bar{\gamma}(\upsilon) = C_X^M \cdot \zeta_2 \left(1 - \bar{\rho}\Big(\upsilon, \bar{\gamma}(\upsilon)\Big)\right). \tag{1.23}$$

The notation of optimization problem is the same as previous sections.

Example 1 (perfect correlation between wealth and data valuation)**.** If we have a perfect correlation between data valuation ω_d and user wealth m_r, it means user with better monetary resource always pays for the bigger data volume. Therefore we can combine the concept of ω_d with m_r as a single random variable budget ($b_g \sim f(b_g)$). Here budget is defined as the maximum amount a user is willing to pay for data access. Thus, the utility for user j can be rewritten as:

$$U^j = b_g^j \times \theta_d \times \bar{\gamma} \times c_d - p_d, \tag{1.24}$$

where θ_d is having the same role as (1.9). The minimum sufficient budget level to join the network is $\frac{\varepsilon_d + p_d}{\bar{\gamma}\theta_d c_d}$ which should be equal or bigger than package price p_d, hence,

$$p_d \leq \frac{\varepsilon_d}{\bar{\gamma}\theta_d c_d - 1}.$$

Now, we can extend our formulation to a multi-package market with packages $\upsilon = \{\{c_d^1, p_d^1\}, \cdots, \{c_d^n, p_d^n\}\}$ and rationality conditions of (1.17-1.19). For every package i

with $p_d^i \leq \frac{\varepsilon_d}{\bar{\gamma}\theta_d c_d^i - 1}$, we have the following user quantities:

$$N_D^i = N_C \begin{cases} P\left(\dfrac{\varepsilon_d + p_d^i}{\bar{\gamma}\theta_d c_d^i} \leq b_g < \dfrac{\varepsilon_d + p_d^{i+1}}{\bar{\gamma}\theta_d c_d^{i+1}}\right) & i \neq n, \\ P\left(\dfrac{\varepsilon_d + p_d^i}{\bar{\gamma}\theta_d c_d^i} \leq b_g \leq b_g^M\right) & i = n, \end{cases} \tag{1.25}$$

where b_g^M is the highest available budget. Converting (1.25) to the form of (1.20) is straightforward and we exclude it. For the analytical purposes, here we represent the network parameters for uniformly distributed budget $b_g \sim U(0, b_g^M)$:

$$N_D = \frac{N_C}{b_g^M}\left(b_g^M - \frac{\varepsilon_d + p_d^1}{\bar{\gamma}(\upsilon)\theta_d c_d^1}\right), \tag{1.26}$$

$$\pi(\upsilon) = \frac{N_C}{b_g^M}\left[p_d^n b_g^M - \frac{1}{\bar{\gamma}(\upsilon)\theta_d}\left(\frac{(\varepsilon_d + p_d^n)p_d^n}{c_d^n} - \sum_{i=1}^{n-1}\left(\frac{\varepsilon_d + p_d^{i+1}}{c_d^{i+1}} - \frac{\varepsilon_d p_d^i}{c_d^i}\right)\cdot p_d^i\right)\right].$$

The traffic intensity $\bar{\rho}^i$ generated by the users who purchased the package (p_d^i, c_d^i) is:

$$\bar{\rho}^i = \frac{N_C \lambda_u \zeta_1 \eta}{C_X^M b_g^M} \begin{cases} \dfrac{c_d^i}{\bar{\gamma}(\upsilon)\theta_d}\left(\dfrac{\varepsilon_d + p_d^{i+1}}{c_d^{i+1}} - \dfrac{\varepsilon_d + p_d^i}{c_d^i}\right) & i \neq n, \\ c_d^n b_g^M - \dfrac{\varepsilon_d + p_d^n}{\bar{\gamma}(\upsilon)\theta_d} & i = n. \end{cases} \tag{1.27}$$

Total traffic intensity and expected data rate are calculated by (1.22) and (1.23) respectively. In (1.23), $\bar{\gamma}(\upsilon)$ is the expected data rate announced to the users and the following solution for it can be easily found:

$$\bar{\gamma} \;=\; \frac{1}{2}\left(\Gamma + (\Gamma^2 + 4\sigma_1\sigma_2)^{\frac{1}{2}}\right), \tag{1.28}$$

$$\Gamma = C_X^M \zeta_2 - \zeta_1 \eta c_d^n, \tag{1.29}$$

$$\sigma_1 = \zeta_1 \zeta_2 N_C \lambda_u, \tag{1.30}$$

$$\sigma_2 = \frac{\eta}{\theta_d b_g^M}\left[\varepsilon_d + p_d^n + \sum_{i=1}^{n-1} c_d^i \left(\frac{\varepsilon_d + p_d^{i+1}}{c_d^{i+1}} - \frac{\varepsilon_d + p_d^i}{c_d^i}\right)\right]. \tag{1.31}$$

Writing the objective function and KKT conditions of the new optimization problem is relatively straightforward and the concept is not much different from the previous section. Now, the optimization parameters are $2n$ values which form n data packages instead of single pair (p_d, c_d) in (1.13).

1.3.3 Package Renovation

In many cases a higher valuation of data package by a user does not always mean that this user has greater monetary resources; there would be many types of potential subscribers who appreciate a good quality data package for its price, yet they cannot afford it. In such cases the covariance of ω_d and m_r can be negative. For the uni-package offer $\upsilon = (c_d, p_d, \bar{\gamma})$, the number of subscribing users is:

$$\begin{aligned} N_D^0 = N_C P\left(\omega_d \geq \frac{\varepsilon_d + p_d}{\bar{\gamma}\cdot c_d}, m_r \geq p_d\right) &= N_C \int_{\frac{\varepsilon_d + p_d}{\bar{\gamma}\cdot c_d}}^{1} \int_{p_d}^{M_r^M} f(\omega_d, m_r)\, c_d \omega_d \, dm_r \\ = N_C\left(1 - F_{\omega_d, m_r}\left(\frac{\varepsilon_d + p_d}{\bar{\gamma}\cdot c_d}, p_d\right)\right). \end{aligned} \tag{1.32}$$

The quantity of users who can afford the package-recharge for nth time is $N_D^n = N_C\left(1 - F_{\omega_d, m_r}\left(\frac{\varepsilon_d + p_d}{\bar{\gamma}\cdot c_d}, np_d\right)\right)$. If the wireless provider sets the recharge cost for extra data as $p_r \neq p_d$, we can modify the above equation by representing:

$$N_D^n = N_C\left(1 - F_{\omega_d, m_r}\left(\frac{\varepsilon_d + p_d}{\bar{\gamma}\cdot c_d}, p_d + (n-1)p_r\right)\right). \tag{1.33}$$

Our formulation will not be completed if we do not consider a recharging incentive factor which indicates the expected probability of purchasing extra service if user has not already reached the budget limit. We define a decreasing function $\left\{G(n)\middle|G(0) = 1, G(\infty) = 0, G(a) > G(b) \iff a < b : \forall a, b \in [0,1]\right\}$ which shows this effect. This function has a major role in package definition since it connects our formulation to the reality of markets by prohibiting the provider to offer an irrationally small-sized package to utilize all the budget levels. Then, one can define the provider profit as:

$$\pi(p_d, p_r, c_d) = N_D^0 \cdot p_d + p_r \times \sum_{i=1}^{n=\lfloor \frac{M_r^M - p_d}{p_r} \rfloor} G(i) N_D^i. \tag{1.34}$$

Where $\lfloor\cdot\rfloor$ indicates the floor function. Since some users are purchasing extra bandwidth, we need to modify the expected arrival rate of flows λ_d. In the new formulation we have $G(n)N_D^n$ ($n = \lfloor \frac{M_r^M - p_d}{p_r} \rfloor$) users who generate flows with rate $(n+1)\lambda_u c_d$; $(G(n-1)N_D^{n-1} - G(n)N_D^n)$ users who produce flows with rate $n\lambda_u c_d$ and so on. If we indicate the users with i-fold package renovation as a group i, the traffic intensity, $\bar{\rho}^i$, generated by this group is:

$$\bar{\rho}^i = \frac{\lambda_u \zeta_1}{\zeta_2}\begin{cases} (G(i-1)N_D^{i-1} - G(i)N_D^i) i\eta c_d & i \neq n, \\ G(i)N_D^i (i+1)\eta c_d & i = n. \end{cases} \tag{1.35}$$

The total traffic intensity and expected data rate are calculated from (1.22) and (1.23). In the optimization problem of Subsection 1.3.1.2, the objective function along with the $\bar{\gamma}$ function can be updated with the new equations. Here, $\bar{\gamma}$ is calculated using a recursive function which appears in both the objective function and the inequality constraint. Thus, even though we reached a better framework to model volume-based data packages, the complexity of finding optimal values is increased by adding an iterative loop inside the main optimization problem to find the value of $\bar{\gamma}$. We can take a further step and combine the results of the two recent subsections to achieve a framework for the multi-package market with data recharging, however, the formulation is straightforward and to avoid redundancy in content and equations we leave it to the readers.

1.4 Dynamic Bandwidth Allocation Model

In the previous section, we investigated the volume-based data package pricing for shared-carrier networks. In this manner, provider tries to attract a subset of voice user set, $\mathbb{N}_C$, to subscribe to its data package. We assumed the provider owns bandwidth Q_D which leads to a maximum data rate $C_X^M = \beta Q_D$. In this section, we expand our framework to model the pricing with regard to another channel access method. In this approach as we described in section 1.2.1.2, the provider guarantees to grant a predefined access data rate to the connected users. Here, in contrast to the previous method, data rate is not changing based on the distance from the base station. However, the available bandwidth to the provider is limited and users may be displeased by service unavailability. This unavailability occurs when the data flow requests are blocked due to spectrum shortage. Hence one needs to add an extra variable to the utility in (1.9) to reflect the dissatisfaction by service unavailability. Henceforth instead of $\bar{\gamma}$ we use the notation γ_d to clarify the invariability of data rate in current approach. Since the utility function presents user's gain, by defining B_d^T as the expected blocking probability for data flow requests, we define the utility and user quantity as:

$$U_j = \theta_d \cdot c_d \cdot \omega_d^j \cdot \gamma_d \cdot (1 - B_d^T) - p_d, \tag{1.36}$$

$$\mathbb{N}_D = \left\{ j \in \mathbb{N}_C \,\middle|\, \omega_d^j \in [0,1] \wedge \omega_d^j \geq \frac{\varepsilon_d + p_d}{\theta_d \cdot c_d \cdot \gamma_d \cdot \left(1 - B_d^T\right)} \right\}, \tag{1.37}$$

$$N_D = N_C \int_{\frac{\varepsilon_d + p_d}{\theta_d \cdot c_d \cdot \gamma_d \cdot (1 - B_d^T)}}^{1} f_{\Omega_d}(\omega_d) c_d \omega_d. \tag{1.38}$$

Now, we need to find the expected blocking probability of data flows which is based on available bandwidth and carrier allocation in admission control process. Due to zero latency experience of 5G networks, we suppose any delay of data transfer has a similar effect to the flow blocking on user utility. Based on (1.8) we have the following relation for spectrum bandwidth and data rate:

$$S(r) = \frac{\gamma_d}{\beta} \begin{cases} 1 & r \leq R_m, \\ \left(\frac{r}{R_m}\right)^{\alpha} & R_m < r \leq R_M. \end{cases} \tag{1.39}$$

However, r which is the user distance form the macro-cell is a random variable and consequently $S(r)$ is also behaving as a random variable. With a little math work we have the CDF and PDF of $S(r)$ as:

$$F_S(s) = \begin{cases} 0 & s < \frac{\gamma_d}{\beta}, \\ \left(\frac{R_m}{R_M}\right)^2 \left(\frac{\beta s}{\gamma_d}\right)^{\frac{2}{\alpha}} & \frac{\gamma_d}{\beta} \leq s \leq \frac{\gamma_d}{\beta} \left(\frac{R_M}{R_m}\right)^{\alpha}. \end{cases} \tag{1.40}$$

$$f_S(s) = \begin{cases} 0 & s < \frac{\gamma_d}{\beta}, \\ \frac{2}{\alpha} \left(\frac{R_m}{R_M}\right)^2 \left(\frac{\beta}{\gamma_d}\right)^{\frac{2}{\alpha}} s^{\frac{2}{\alpha}-1} & \frac{C_C}{\beta} \leq s \leq \frac{\gamma_d}{\beta} \left(\frac{R_M}{R_m}\right)^{\alpha}. \end{cases} \tag{1.41}$$

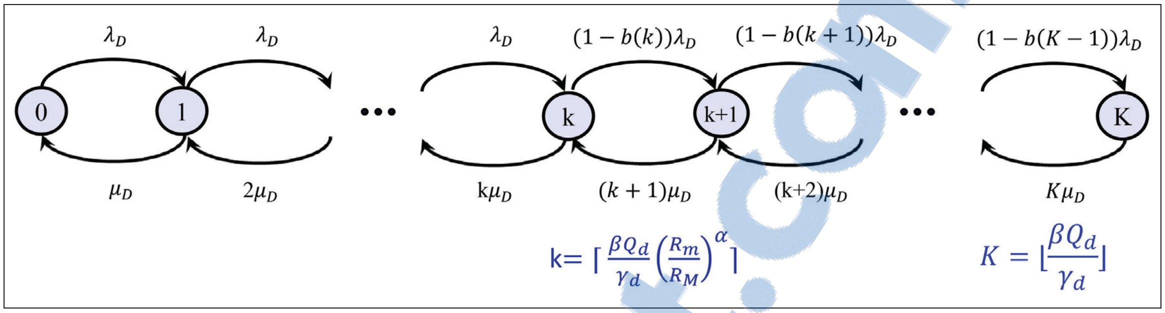

Figure 1.3 M/M/K/K birth-death model with state dependent arrival rates

$$\mu_S = E[S] = \frac{2}{\alpha+2} \cdot \frac{\gamma_d}{\beta} \cdot \left(\left(\frac{R_M}{R_m}\right)^{\alpha} - \left(\frac{R_m}{R_M}\right)^{2} \right), \tag{1.42}$$

$$E[S^2] = \frac{1}{\alpha+1} \cdot \left(\frac{\gamma_d}{\beta}\right)^{2} \cdot \left(\left(\frac{R_M}{R_m}\right)^{2\alpha} - \left(\frac{R_m}{R_M}\right)^{2} \right), \tag{1.43}$$

$$\sigma_S^2 = E[S^2] - \mu_S^2. \tag{1.44}$$

The above PDF shows the probability density of required spectrum size for each user in the coverage area. If the overall available spectrum to the provider for the data network is Q_D, then the quantity of concurrent active users A_d varies as $k = \lceil \frac{Q_D \cdot \beta}{\gamma_d} (\frac{R_m}{R_M})^{\alpha} \rceil \leq A_d \leq \lfloor \frac{Q_D \cdot \beta}{\gamma_d} \rfloor = K$, where $\lceil . \rceil$ and $\lfloor . \rfloor$ are the ceiling and floor functions respectively.

We can model the data network as a multi-server model with Poisson arrival $\lambda_D = N_D \lambda_u$ (λ_u is the arrival rate for each user's data requests) and expected service time $\frac{1}{\mu_D} = E[t_d] = \frac{\eta c_d}{\gamma_d}$. t_d is connection (data flow) duration with exponential distribution. This is a $M/M/K/K$ system with no queue and $K = \lfloor \frac{Q_C \beta}{\gamma_d} \rfloor$ servers. In standard model of queue-less multi-server system the service blocking occurs when the user request arrives in state K which means user observes K other users are being served in that time. In our model, due to randomness of bandwidth (or sub-carrier) allocation, a user service request would be blocked in a state $j < K$. Due to this fact when the system is stationary, the arrival rate observed by an observer inside the system is related to its state. This effect is depicted in Fig. 1.3 as follows: the arrival rate is λ_D in all states in which the system definitely has enough resources to guarantee the promised data rate, as soon as system reaches to the state $j > k = \lceil \frac{Q_D \cdot \beta}{\gamma_d} (\frac{R_m}{R_M})^{\alpha} \rceil$, due to request blocking, the average arrival rate observed by the viewer inside the system is $(1-b(j-1))\lambda_D$ where $b(j-1)$ is the blocking probability of flows in state $j-1$ due to resource shortage. The more precise interpretation of blocking in the state $j \geq k$, $b(j)$, is that by having j users in the system, there is not enough spectrum (in form of sub-carriers) to serve $(j+1)$th user, hence, in view of a Markov process, assuming the system is already in state j, the probability of having enough resources to go to the next state is $1-b(j) = P(Q_D - \sum_{i=1}^{j+1} S(r_i) \geq 0)$. As a result of above discussion, the instant blocking probability of flows should be calculated for each system state. With respect to this fact, the general definition for the expected blocking probability in the steady state of the system is,

$$B_d^T = \sum_{j=k=\lceil \frac{Q_D\beta}{\gamma_d}(\frac{R_m}{R_C})^{\alpha} \rceil}^{K=\lfloor \frac{Q_D\beta}{\gamma_d} \rfloor} b(j) \times P(j), \tag{1.45}$$

where j represents the system state, $P(j)$ is the probability of being in state j. In the following two subsections, we explain a method to calculate $P(j)$ and $b(j)$.

1.4.1 Finding $b(j)$

Since the bandwidth random variables for all users are i.i.d, let define a new random variable $I(j) = \sum_{i=1}^{j} S(r_i)$, and we have the following definitions for it:

$$f_{I(j)}(\iota) = f_S^{j*}(\iota), \tag{1.46}$$

$$\mu_{I(j)} = j \cdot \mu_S, \tag{1.47}$$

$$\sigma_{I(j)}^2 = j \cdot \sigma_S^2, \tag{1.48}$$

where $f^{j*}(.)$ indicates the j-fold convolution of $f(.)$. As we indicated earlier, the probability of being blocked when the system is in state j is $P(Q_D - \sum_{i=1}^{j+1} S(r_i) < 0)$, which can be represented as:

$$b(j) = 1 - F_{I(j+1)}(Q_D). \tag{1.49}$$

Since in today's cellular networks the quantity of concurrent users is usually greater than 10, with a good approximation, by using central limit theorem, a Gaussian distribution can be assumed for $f_{I(j)}(\imath)$. Hence, we can represent:

$$f_{I(j)}(\imath) \approx G(j\mu_S, j\sigma_S^2) = \frac{1}{(2\pi j\sigma_S^2)^{\frac{1}{2}}} \cdot e^{-\frac{(\imath - j\mu_S)^2}{2j\sigma_S^2}}, \tag{1.50}$$

and from (1.49):

$$b(j) = 1 - \Phi\left(\frac{Q_D - (j+1)\mu_S}{\sqrt{j+1}\sigma_S}\right), \tag{1.51}$$

where

$$\Phi(x) = (2\pi)^{\frac{-1}{2}} \int_{-\infty}^{x} e^{-x^2} dx. \tag{1.52}$$

1.4.2 Finding $P(j)$

The call admission scheme in our system is based on a queue-less multi-server model M/M/K/K which has state-dependent arrival rates for service requests due to random resource shortage. In this part, we elaborate this system which is depicted in Fig. 1.3. The stationary probabilities of states in a standard $M/M/K/K$ system is well-known: for a system with K servers and traffic intensity $\rho_d = \frac{\lambda_D}{\mu_D} = \frac{N_D\lambda_u}{\mu_D}$ we have:

$$P(\rho_d, j) = \frac{\frac{\rho_d^j}{j!}}{\sum_{i=1}^{K} \frac{\rho_d^i}{i!}}. \tag{1.53}$$

However, since the shortage of spectrum can force the admission controller to reject some data request in state $j < K$, then, from the viewpoint of an observer inside the system, the arrival rate is different from state to state. Hence, we have to write the stationary equations to obtain the probabilities of states in the system of Fig. 1.3. The stationary equations are as follows:

$$\lambda_D P(0) = \mu_D P(1), \tag{1.54a}$$

$$(\lambda_D + \mu_D) P(1) = \lambda_D P(0) + 2\mu_D P(2), \tag{1.54b}$$

$$\vdots$$

$$(\lambda_D + (k-1)\mu_D) P(k-1) = \lambda_D P(k-2) + k\mu_D P(k), \tag{1.54c}$$

$$((1-b(k))\lambda_D + k\mu_D) P(k) = \lambda_D P(k-1) + (k+1)\mu_D P(k+1), \tag{1.54d}$$

$$\vdots$$

$$((1-b(K-1))\lambda_D) P(K-1) = K\mu_D P(K). \tag{1.54e}$$

The above system of equations yields to the Markov chain stationary state probabilities as follows:

$$P(j) = \delta \times \begin{cases} \frac{\rho_d^j}{j!} & j \leq k = \lceil \frac{Q_D \beta}{\gamma_d} (\frac{R_m}{R_C})^{\alpha} \rceil, \\ (1-b(j-1))\frac{\rho_d^j}{j!} & k < j \leq K = \lfloor \frac{Q_D \beta}{\gamma_d} \rfloor, \end{cases} \tag{1.55}$$

$$\delta = \left(\sum_{i=1}^{k} \frac{\rho_d^i}{i!} + \sum_{i=k+1}^{K} \frac{(1-b(j-1))\rho_d^i}{i!} \right)^{-1}. \tag{1.56}$$

By putting (1.51) and (1.55) into (1.45) ($B_d^T = \sum_{j=k}^{K} b(j)P(j)$), we achieve the final form of expected blocking probability for data flows:

$$B_d^T = \delta \times \left(b(k)\frac{\rho_d^k}{k!} + \sum_{j=k+1}^{K-1} b(j)(1-b(j-1))\frac{\rho_d^j}{j!} + (1-b(K-1))\frac{\rho_d^K}{K!} \right). \tag{1.57}$$

1.4.3 Expected blocking with user utility applied

One of the main parameters of B_d^T in (1.57) is the traffic intensity represented by ρ_d. The traffic intensity can be represented by $\rho_d = \frac{N_D \lambda_u}{\mu_D}$; N_D directly comes from (1.36-1.38), and by using those definitions we can formulate the traffic intensity as,

$$\rho_d = \frac{N_C \lambda_u}{\mu_D} \left(1 - \frac{\varepsilon_d + p_d}{c_d \theta_d \gamma_d (1 - B_d^T)} \right). \tag{1.58}$$

In this representation, ρ_d is a function of the blocking probability itself and consequently we can conclude that (1.57) is a recursive function having the form $B_d^T(Q_D, \rho_d(B_d^T))$.

1.4.4 Optimization problem

For the case of dynamically allocated spectrum, the provider profit optimization problem is,

$$\max_{\gamma_d, c_d, p_d} N_D(\gamma_d, c_d, p_d) \times p_d, \tag{1.59}$$

subject to

$$\gamma_d \geq \gamma^m, \tag{1.60a}$$

$$c_d \geq d^m, \tag{1.60b}$$

$$p_d \geq 0, \tag{1.60c}$$

$$0 \leq B_d^T(Q_D, \rho_d) \leq B^M. \tag{1.60d}$$

We have the user quantity in data network, $N_D(\gamma_d, c_d, p_d)$, from (2.8) which includes the blocking probability as well. Here we added a constraint for maximum blocking probability B^M as a representation of regulator's obligations. The complexity of this problem arises from the fact that $B_d^T(Q_D, \rho_d)$ does not adopt a closed-form solution and the problem is needed to be solved with numerical methods.

1.4.5 Final profit

The cellular provider has two sources of profit income from data and voice users. We also know that the data users are a subset of the voice service users. The total bandwidth available to the provider is Q_T. To separately control the level of availability of each service, we can define a bandwidth splitting ratio $0 \leq \psi \leq 1$. In this manner $Q_C = \psi Q_T$ and $Q_D = (1-\psi)Q_T$. Considering this decision value and using the profit model in the previous sections, we can define a combined profit function to resolve all the network decision values:

$$\max_{\psi, p_d, c_d, \gamma_d} \pi_{c+d} = N_C \times p_c + N_D \times p_d = N_C \left[p_c + p_d \left(1 - \frac{\varepsilon_d + p_d}{\gamma_d \big(1 - B_d^T((1-\psi)Q_T, \rho_d)\big) \theta_d c_d} \right) \right], \tag{1.61}$$

subject to,

$$c_d \geq d^m, \tag{1.62a}$$

$$p_d \geq 0, \tag{1.62b}$$

$$\psi_0 \leq \psi \leq 1, \tag{1.62c}$$

$$\gamma_d \geq \gamma^m, \tag{1.62d}$$

$$B_c^T(\psi Q_T, \rho_c) \leq B_c^M, \tag{1.62e}$$

$$B_d^T((1-\psi)Q_T, \rho_d) \leq B_d^M. \tag{1.62f}$$

Where $B_c^T(\psi Q_T, \rho_c) \leq B_c^M$ defines the upper limit for the probability of call blocking in voice service. This model of profit optimization upon bandwidth splitting gives a more precise calculation since it separates the voice and data satisfaction, and gives different perceptual evaluation-tolerance of each service to the users.

1.4.6 Multi-package market for dynamic allocation method

In the pricing approaches for the shared-carrier technology, we formulated the rationality conditions for a multi-package market in (1.17)-(1.19). The same conditions are applied to the case of dynamic allocation technologies. The arrival rate of data flow requests is directly related to the data volume cap, $\lambda_d = \lambda_u c_d$. Hence, multiple plans with different cap levels, $\upsilon = \big\{\{c_d^1, p_d^1\}, \cdots, \{c_d^n, p_d^n\}\big\}$, affect the overall rate of flow requests. With the analogous notations to Subsection 1.3.2.1, we can write the following equations for the number of subscribers in each plan,

$$\begin{aligned} \underset{i\neq n}{N_D^i} = N_C \Bigg[F_{M_r}\left(p_d^{i+1}\right) + F_{\Omega_d,M_r}\left(\frac{\varepsilon_d + p_d^i}{(1-B_d^T(\bar{\rho}_d(\upsilon)))\gamma_d\theta_d c_d^i}, p_d^i\right) \\ - F_{\Omega_d,M_r}\left(\frac{\varepsilon_d + p_d^i}{(1-B_d^T(\bar{\rho}_d(\upsilon)))\gamma_d\theta_d c_d^i}, p_d^{i+1}\right) - F_{M_r}\left(p_d^i\right), \Bigg], \end{aligned} \tag{1.63}$$

$$\bar{\rho}_d(\upsilon) = \sum_{i=1}^{n} \frac{N_d^i \lambda_u c_d^i \eta}{\gamma_d}. \tag{1.64}$$

To obtain N_D^n from (1.63), M_r^M is substituted for p^{i+1} in (1.63). Suppose the values $\{c_d^i, p_d^i\}$ $\forall i \in n$ and γ_d are given. Then, (1.63) and (1.64) together define a system of nonlinear equations with n variables. Writing the constraints for the objective function, $\pi(\upsilon) = \sum_{i=1}^{n} N_D^i(\upsilon)p^i$, is not much different from the previous cases. Since finding a closed-from solution for N_D^i is hardly possible, we use numercial methods embedded in MATLAB to find the optimal values in different realistic scenarios in the next section.

1.5 Numerical Results

In this section, we analyze four realistic scenarios. Firstly, the expected blocking probability of (1.57) is analyzed. Then, we extend the analysis to the case in which the utility function of users affects the blocking probability of data flows. In the third scenario, we consider providers with the shared carrier and dynamic allocation methods and derive the optimal values of c_d, p_d, and γ_d. Finally, a provider with multiple data plans is investigated, and optimal network parameters are derived for different values of λ_u.

Scenario 1. To justify (1.57), we ran a series of numerical scenarios which are depicted in Fig. 1.4-1.9. Fig. 1.4 represents the value of B_d^T for different available spectrum sizes (Q_D) while traffic intensity ρ_d is the variable. We selected the guaranteed data rate as rational value 50 Mbps and $R_M = 4R_m$. The spectral efficiency (β) is set to 4 which represents LTE technology with SISO system (Mogensen *et al.*, 2007). As it is expected, the higher available bandwidth offers lower blocking probability even though by increasing ρ_d (simple representation of ρ_d), BP goes to 1 eventually. Fig. 1.5 is a close-up of previous setting in interval $\rho_d = [0,1]$. Network parameters in Fig. 1.6 are different from the previous group by having a lower guaranteed data rate $\gamma = 10$ Mbps which drastically decreases the blocking probability in lower values of ρ_d. Fig. 1.7 represents the probability of the system being in each state when a new request arrives. Here each curve is analogous to one value of set $[5,10,15,20]$ for ρ_d. Since the arrival rate is decreasing in state-dependent format, we observe the state probability goes to zero as the system is in higher states. Fig. 1.8 has spectral efficiency $\beta = 30$ which is similar to LTE-advanced with MIMO system. As a result of higher efficiency, the non-zero blocking probabilities spread to a broader range of system states. Finally, Fig. 1.9 depicts the curves associated with blocking probability in each system state. Each curve represents a different value of guaranteed data rate varying from 10 to 40 Mbps. Due to spectrum shortage, the blocking probability eventually goes to 1 in higher state numbers while the lower guaranteed data rate has better performance as it is expected.

Scenario 2. In the previous scenario, we justified the expected blocking probability by cases in which no user utility is involved. In this example we use the same network settings (Q=350MHz, $Rc = 4R_m$, $\alpha = 2$, $\beta = 4$) and analyze the expected blocking probability when the user utility is considered, which is formulated in (1.58). Considering (1.58), we use the notation $\rho_d^M = \frac{N_C \lambda_u c_d}{\mu_D}$ as the upper limit of traffic intensity and use the ratio $\frac{\varepsilon_d + p_d}{\theta_d \times \gamma_d \times c_d}$ (cost to gain) as the variable for the provider. Fig. 1.10 depicts the blocking probability for different

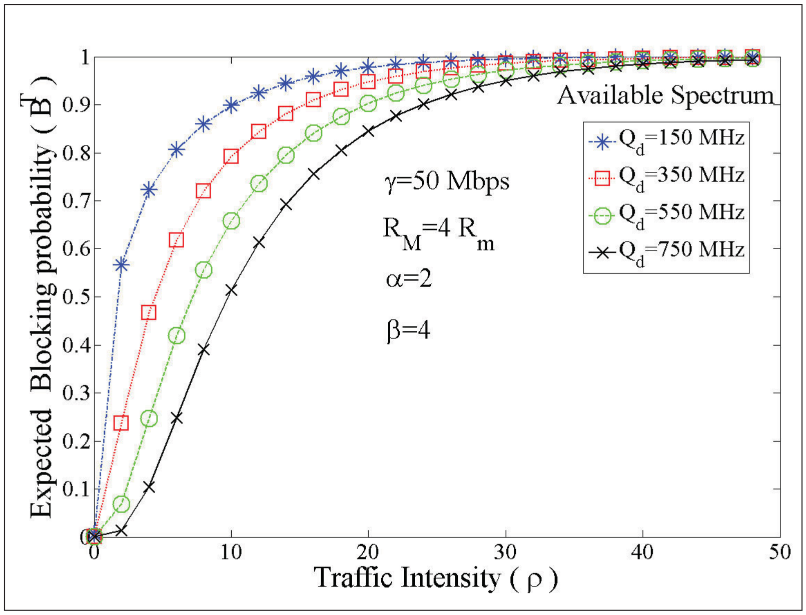

Figure 1.4 Expected blocking probability for $\rho_d = [0,100]$

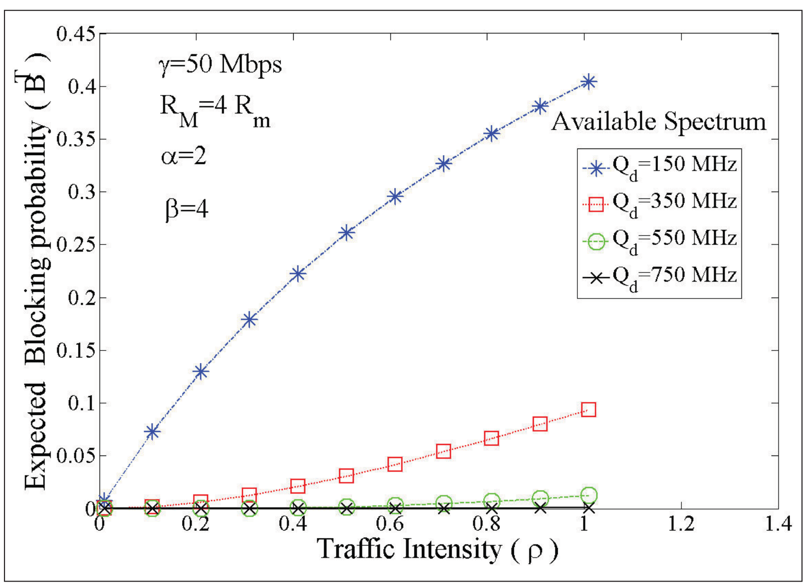

Figure 1.5 Expected blocking probability for $\rho_d = [0,1]$

values of ρ_d^M while the guaranteed data rate is 50 Mbps. As it is shown in the figure, when user's cost to gain ratio increases, the expected blocking probability decreases due to fewer network subscription. Fig. 1.11 shows the numerical results in which the guaranteed data rate γ_d is constant in $\frac{\varepsilon_d+p_d}{\theta_d\times\gamma_d\times c_d}$. Five curves represent the offered rate from 10 to 50 Mbps. In this

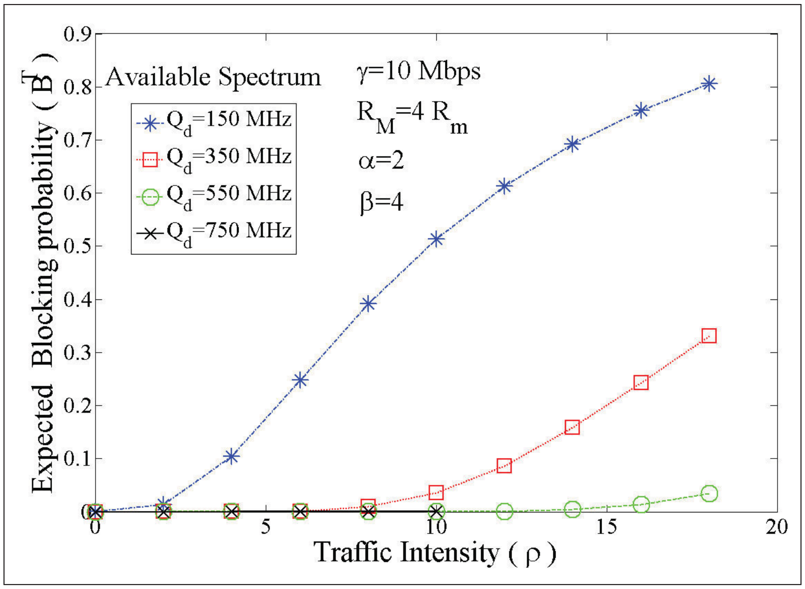

Figure 1.6 Expected blocking probability for $\gamma = 10$ Mbps

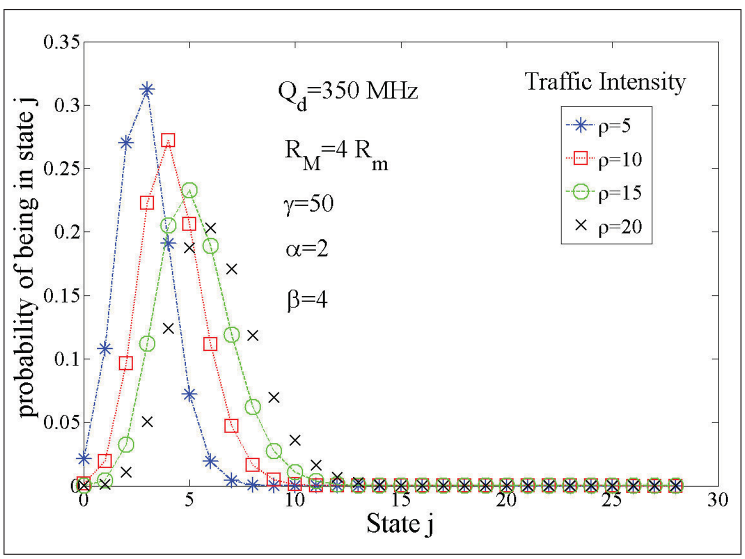

Figure 1.7 Markov chain State Probability ($\beta = 4$)

case, by increasing the cost to gain ratio, the provider needs to decrease the offered blocking probability to maintain the data rate which is assumed to have a lower bound in the busy hour indicated by γ_d.

Scenario 3. Here, we consider a realistic busy-hour scenario for each of the two bandwidth allocation methods and find the optimum values with MATLAB. Table 1.2 represents the net-

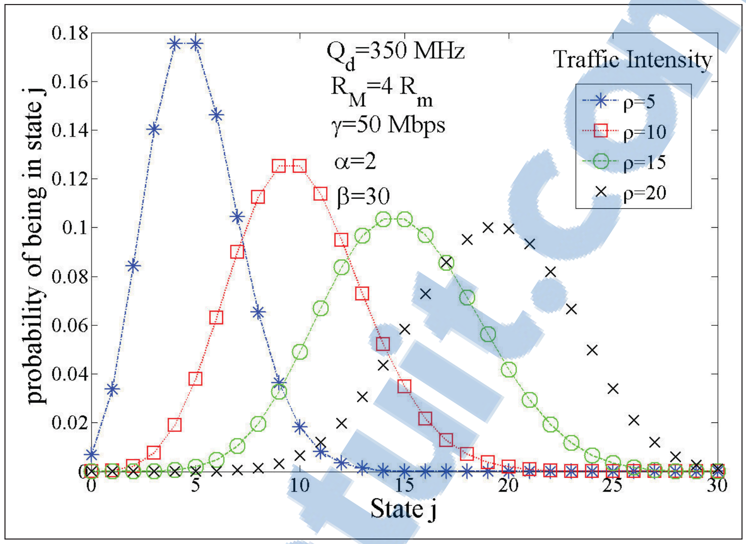

Figure 1.8 Markov chain State Probability ($\beta = 30$)

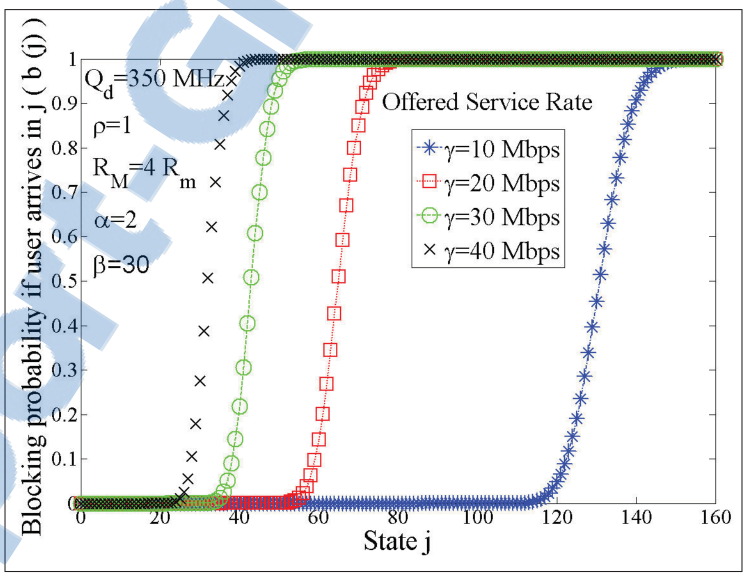

Figure 1.9 State blocking probability $b(j)$

work parameters. We assume that a macrocell has 6 sectors. Concerning the shared-carrier method, a lower total available bandwidth of $Q_T = 10$ MHz and spectral efficiency of $\beta = 0.5$ are used, which are similar to WCDMA 3G networks. The available bandwidth in dynamic allocation method is 50 MHz. We consider 150 and 1500 voice subscribers for shared-carrier and dynamic allocation schemes respectively. In both scenarios, each voice user has an arrival

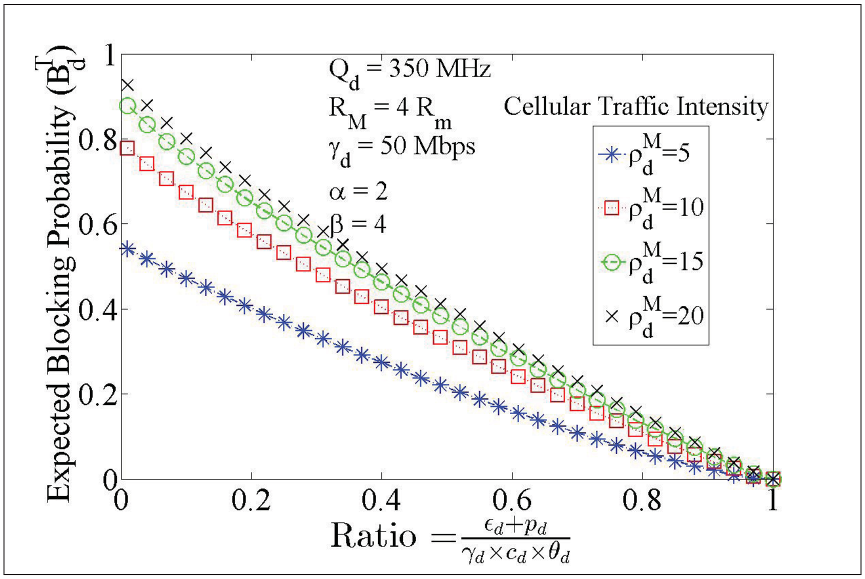

Figure 1.10 Expected Blocking Probability (B_d^T) with the ratio $\frac{\varepsilon_d+p_d}{\theta_d\times\gamma_d\gamma_d}$ as variable. Curves represent different values of ρ_C

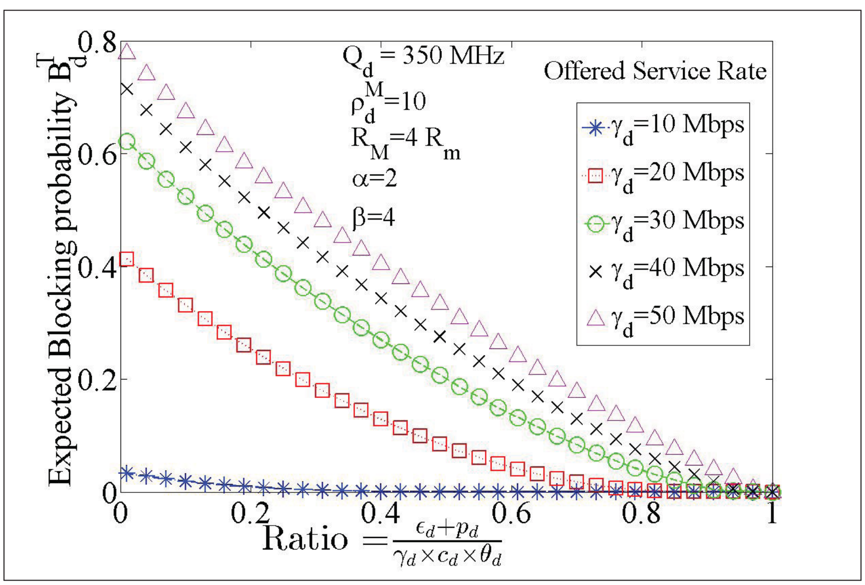

Figure 1.11 Expected Blocking Probability (B_d^T) with the ratio $\frac{\varepsilon_d+p_d}{\gamma_d\times c_d\times\theta_d}$ as variable. Curves represent different values of γ_d

rate of 10^{-4} and service time of 90 seconds with 150 Kbps of required data rate. The arrival rate of each data request is taken as an independent variable in both scenarios, and the average flow size is $\eta = 100$ KB. While data service valuation ω_d is assumed to be uniform, we consider a beta distribution for user budget, $\frac{b_g}{100} \sim \text{Beta}(2,6)$, which is depicted in Fig. 1.12. In both scenarios, θ_d is set to 50 so, with a minimum acceptable data rate of 1 Mbps and data size of 1 GB, the users with the highest service valuation are willing to pay no more than 50$ for the data service.

Figs. 1.13-1.15 show the numerical results for the shared-carrier scenario. Fig. 1.13 depicts the optimal values of data volume cap and service data rate. The scale factor for the arrival rate of data flows, λ_u, is the independent variable in all curves. Note that the arrival rate of each user's data flow is $\lambda_d = \lambda_u c_d$. Fig. 1.14 depicts the optimal price for voice and data services and finally Fig. 1.15 shows the subscribed users to each service. One important aspect of the price and data rate curves is rather small amount of fluctuations in optimal values. By increasing the data request rate, the offered data size should be decreased which yields to a drop in subscriber quantity. This is due to the dependency of the overall number of subscribers to the ratio of $\frac{p_d}{\gamma_d \times c_d}$. In this way, the provider adjusts the value of price and data rate respectively to maximize the overall profit. This causes a small variation in optimal values. To eliminate these fluctuations, one option is to fix price and data rate on their average value and having data volume cap as the only variable in service parameters. For this scenario, the optimal value of ψ is 0.1 for all values of λ_u.

For the dynamic allocation scenario, we expect a similar price and data rate adjustment as it is depicted in Figs. 1.16-1.17. We obtained the results for two values of spectral efficiency factor, $\beta \in [4,16]$. $\beta = 4$ indicates LTE SISO and $\beta = 16$ represents LTE MIMO (4×4). Here the data volume cap and subscriber quantity are also decreasing in response to the increase of λ_u. Due to the availability of broader bandwidth and higher spectral efficiency, the provider can attract more subscribers. The related curves are depicted in Fig. 1.18. The optimal value of ψ, which is the bandwidth splitting ratio, is 0.5 for $\beta = 4$ and 0.1 for $\beta = 16$, which remains constant for all values of λ_u. Fig. 1.19 shows the data rate for the case in which the provider adjusts the power of signal to achieve the equation $R_M = 2R_m$. For the both cases of $\beta = 16$ and

$\beta = 4$, the data rate dramatically increases. Particularly, for $\beta = 16$, provider can guarantee a data rate of $\gamma_d = 27$ Mbps in the busy hour.

Table 1.2 Network parameters

Parameter	**Shared carrier**	**Dynamic allocation**
N_C	150	1500
Q_T	10 MHz	50 MHz
β	0.5	$[4, 16]$
θ_d	50	50
ω_d	$U(0,1)$	$U(0,1)$
$\frac{b_g}{100}$	$\mathrm{Beta}(2,6)$	$\mathrm{Beta}(2,6)$
η	100 KB	100 KB
λ_c	10^{-4}	10^{-4}
Data λ_u	independent variable	independent variable
$\frac{1}{\mu_c}$	90 seconds	90 seconds
γ_c	150 Kbps	150 Kbps
B_d	-	0.01
B_c	-	0.001
R_M	$5R_m$	$5R_m$
α	2	2

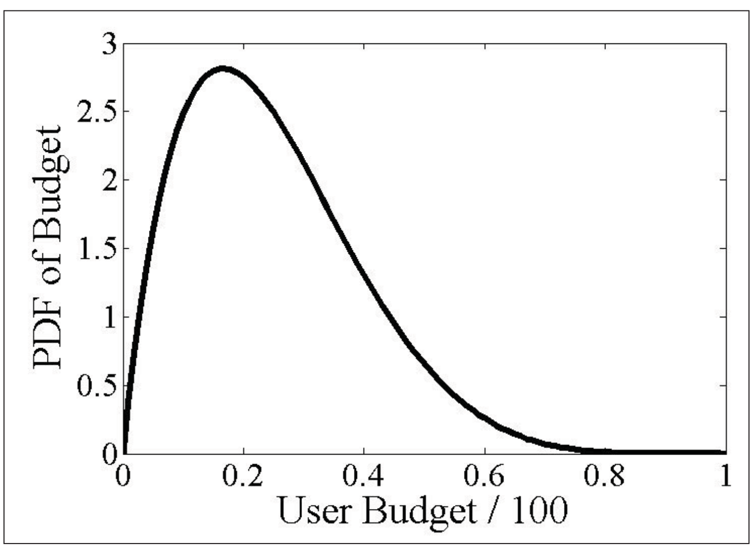

Figure 1.12 Distribution of user budget for data service

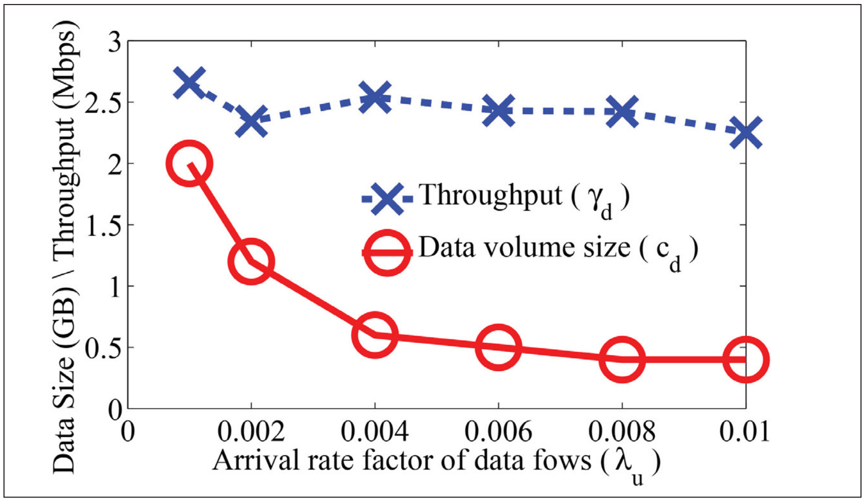

Figure 1.13 Optimal data rate and data volume cap in shared carrier network in shared carrier scenario

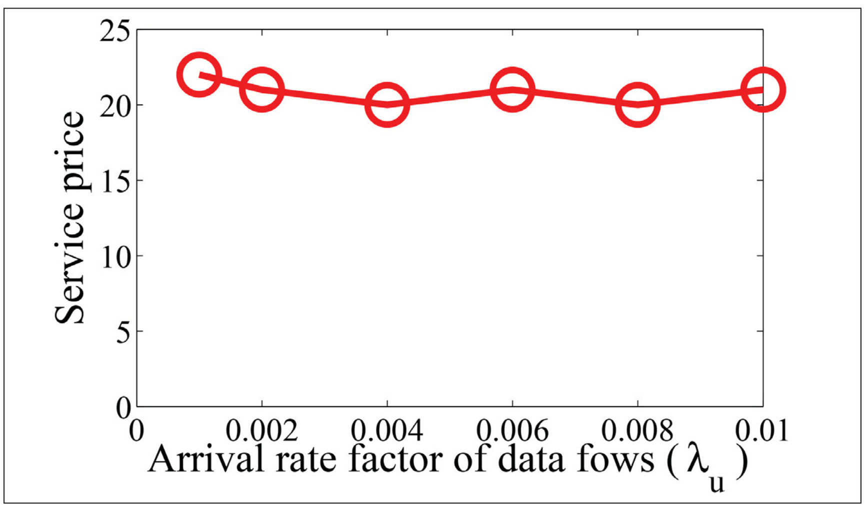

Figure 1.14 Optimal price the data service in shared carrier scenario

Scenario 4. For the case of multi-package markets, we considered a provider that uses dynamic allocation method, and its network parameters are the same as the values in Table 1.2. We used the spectral efficiency $\beta = 4$ for LTE SISO. The optimal package parameters are derived for three values of $\lambda_u \in [0.01, 0.05, 0.1]$. We assumed the price difference between each plan is a multiple of 5$ and the data volume cap of each plan should be two times bigger than its inferior plan. Table 1.3 (next page) shows the results. Similarly to the previous scenarios,

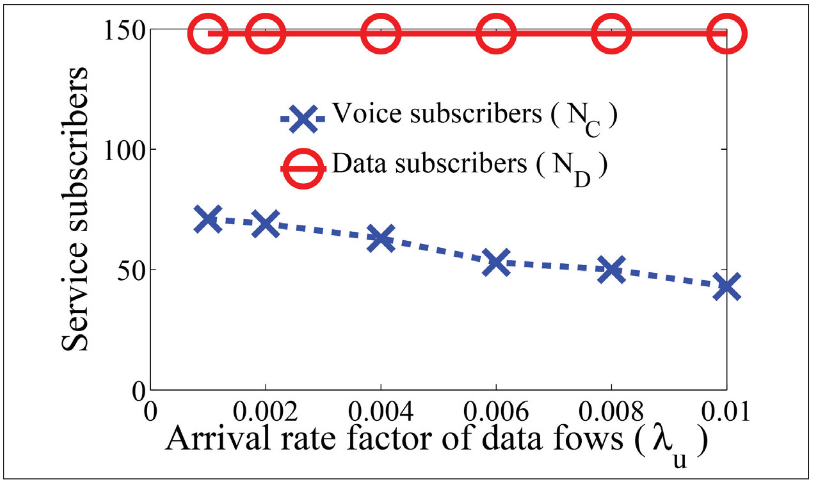

Figure 1.15 Number of subscribed users to voice and data services in shared carrier scenario

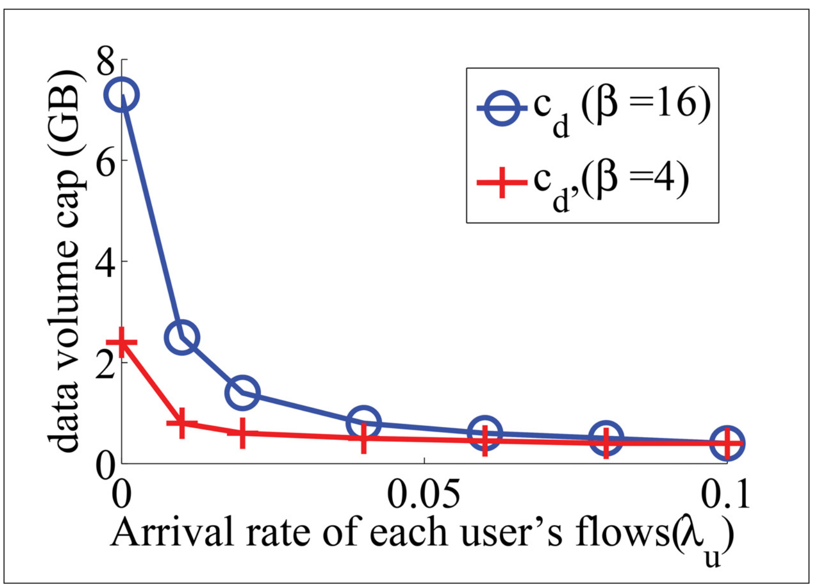

Figure 1.16 Optimal data volume cap in dynamic allocation scenario

γ_d is the minimum data rate that each user would experience during the busy hours. As we indicated before, the small variation in optimal values is due to the dependency of the overall number of subscribers to the ratio of $\frac{p_d}{\gamma_d \times c_d}$. By increasing the value of λ_u provider decreases the data volume size and increases the service price to adjust the number of subscribers and their willingness to over-utilize the network.

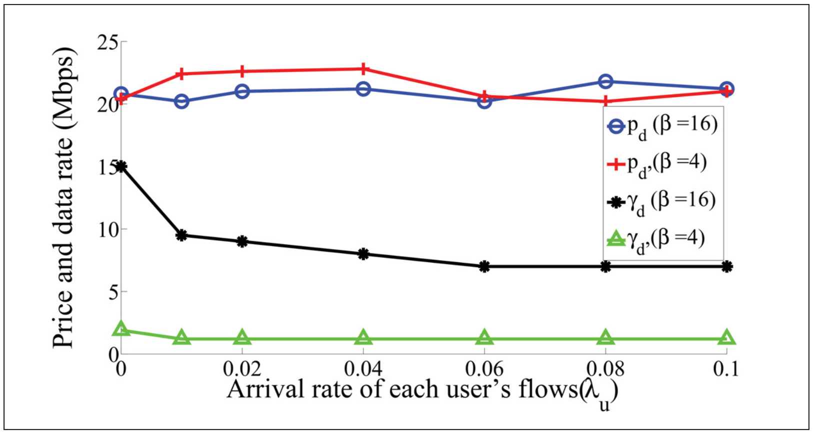

Figure 1.17 Optimal price and data rate for the data service in dynamic allocation scenario

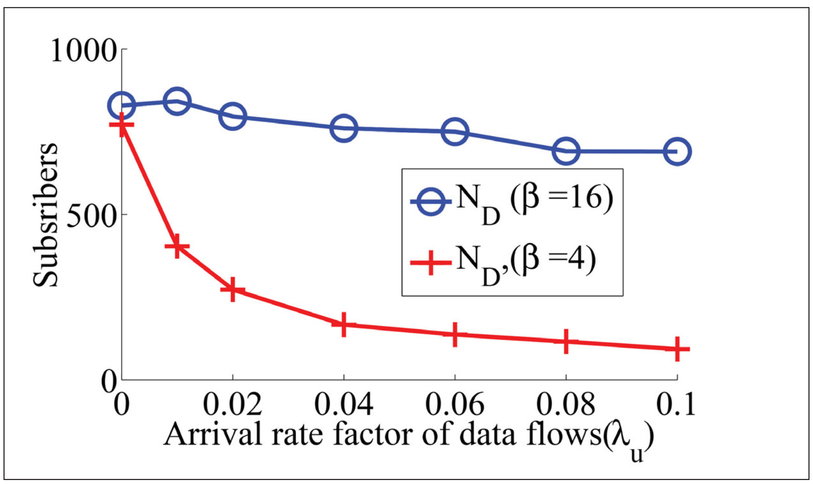

Figure 1.18 Number of subscribed users to voice and data services in dynamic allocation scenario

Table 1.3 Three-plan market with guaranteed minimum data rate γ_d

λ_u	0.01	0.05	0.1
(p_d^1, c_d^1, N_D^1)	$(20, 2, 306)$	$(25, 0.5, 198)$	$(28, 0.5, 120)$
(p_d^2, c_d^2, N_D^2)	$(30, 4, 234)$	$(35, 1, 150)$	$(38, 1, 96)$
(p_d^3, c_d^3, N_D^3)	$(40, 8, 224)$	$(45, 2, 138)$	$(48, 2, 84)$
γ_d (Mbps)	2	2.5	2

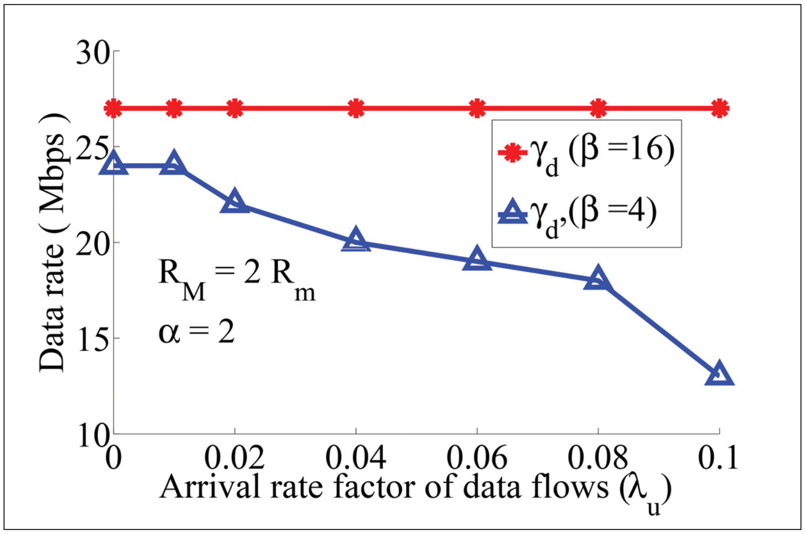

Figure 1.19 The guaranteed data rate for $R_M = 2R_m$

1.6 Conclusion

In this paper, an analytical study for pricing in cellular networks with volume-based data services is presented. Our goal was finding the optimal values of data volume cap, its price and the data rate in the busy hours. We considered two types of channel access methods, shared carrier and dynamic sub-carrier allocation. We started by defining a model for average data rate in the shared-carrier method which is linked to the offered data-volume. Then, we modeled the user utility by considering the random valuation of data and guaranteed data rate. We applied budget information to the model to reach a more realistic scenario. We expanded the framework by bringing the concept of multi-package market and its effect on profit maximization. We studied a model for data-renovation by users during a monthly period which is closely related to the budget distribution. The blocking probability of data flows due to spectrum shortage when the provider is offering consistent service quality is calculated. In this case, dynamic sub-carrier allocation is considered. Finally, a profit maximization framework is proposed in which the provider's decision values are the prices, dedicated spectrum sizes and blocking probabilities for voice and data respectively as well as the data volume cap for the data service. For the future work, we will investigate the multi-provider network and the effects of competition and cooperation on offered service quality and possible profit.

CHAPTER 2

COALITIONS IN HETEROGENEOUS WIRELESS NETWORKS: USER BEHAVIOR AND PROVIDER PROFIT

Behdad Heidarpour [1], Zbigniew Dziong [1], Wing Cheong Lau [2], François Gagnon [1]

[1] Département de Génie Électrique, École de Technologie Superieure ,
1100 Notre-Dame Ouest, Montréal, Québec, Canada H3C 1K3
[2] Department of Information Engineering, The Chinese University of Hong Kong,
Sino Building, Chung Chi Rd, Sha Tin, Hong Kong

Submitted to the journal of Computer Networks (Elsevier) in April 2017.

Abstract

Mobile users demand better speed, coverage, and reliability. To fulfill these requirements in a cost-effective manner, operators can form coalitions that can include heterogeneous technologies such as LTE and WiFi. In this paper, we propose a game-theoretic framework that can help to create stable coalitions of heterogeneous wireless operators and enable wireless regulatory bodies to determine acceptable coalitions which do not downgrade social welfare standards. In particular, we derive a simple and efficient generic model that predicts the state of the market before and after a coalition formation without focusing on short time-scale bandwidth allocation problems. The model is based on finding a pure Nash equilibrium strategy profile that defines service prices for each provider. The solution is based on experienced costs of providers and user satisfaction metrics that are related to the offered speed and coverage. The model features and usefulness are illustrated using some realistic and practical scenarios. The results show that formation of coalitions can notably increase the profits of providers and the integrated payoffs of users.

Introduction

Profit maximization is the main concern for all types of wireless service providers as their business is based on making the profit out of subscribers by providing them a rational quality of service (QoS). The better service providers offer, the more subscribers they attract. To be successful, providers need to expand their coverage areas and provide a better quality of service in terms of higher connection speed, reliability, coverage size, etc. However, the monetary resources of each provider are limited and the networks cannot be expanded or upgraded instantly. This fact motivates us to study a cost-effective approach that can help the providers in a market of heterogeneous networks (HetNets) to increase their revenue and improve the quality of service while the incurred costs are minimized. One approach that has been considered in several works, e.g., (Singh *et al.*, 2011, 2012b), is to form coalitions in homogeneous wireless markets. These works try to utilize the concepts of game theory and increase the profit of providers. This approach is also consistent with the free roaming plans that are offered by several operators. For example, since June 2015, Indian operator BSNL offers free national wide roaming plans in cooperation with MTNL (BSNL). Sprint offers a *Global Roaming* plan which includes free unlimited data and text across its worldwide participating networks including the ones from Canada, Mexico, and Latin America (Sprint, 2016). We aim to expand this idea to the field of heterogeneous operators that offer complementary strengths in terms of coverage, technologies, tiered service speeds, and pricing. Due to technological differences and implementation costs, it is critical for the operators to assess the potential gain before committing to form a coalition. Our work is based on an analytical structure that covers both competitive and cooperative states of HetNets market. Based on this structure, we derive coalition formation models for heterogeneous wireless networks that define roles and interactions of the main decision makers in the market, namely, users, providers and regulator. It should be noted that we try to avoid the formation of the grand coalition since it leads to a monopoly which is prohibited by regulatory units. Thus, our goal is to maintain competition among providers in the form of established coalitions.

Contribution

The contribution of this paper can be summarized as follows:

a. The proposed analytical framework is based on user's long-run preferences. In this manner, user's perspective of service speed and coverage size are the core decision values for selection of user's default provider.

b. We analyze provider's profit in competitive and cooperative states of the market which allows each provider to choose the best coalition unit.

c. We define a multi-provider payoff function for users which enables the wireless users to instantly switch between the networks of a specific coalition. Since the provider selection is instant and allowable inside the coalitions, it is an extra degree of freedom for the users. Hence, coverage size, service speed, and multiple network choices are three long-run QoS parameters in our work.

d. We propose an analytical framework in order to study a structure of coalitions and its impact on both user and provider sides. In this manner we define a coalition formation process and show that it leads to stable coalitions in wireless networks. While there are some works in which the Core and Shapley value for the grand coalition are studied, we avoid formation of a grand coalition since the regulatory units do not allow it (as it leads to a monopoly).

e. We define the market fairness metric for the wireless regulator. This metric helps the regulator unit to decide if forming a coalition a) improves the service experience for the users; b) maintains a certain level of competition between providers.

Analytical framework

In Sections 2.3 and 2.5, we analyze HetNet markets without coalitions and propose models that cover two types of interactions: user-provider and provider-provider. We first define metrics to quantify the effect of user-provider interaction and then we model the procedure of provider selection by users. This model also helps the providers to choose the best service parameters and pricing strategies in order to maximize their profits. The model consists of three parts which are related to the user satisfaction (a function of link speed), the network/coverage expansion costs and the pricing strategy. To characterize the behavior of users, we follow the widely adopted utility theory ((Duan *et al.*, 2013c; Gajic *et al.*, 2009; Acemoglu *et al.*, 2004; Niyato & Hossain, 2009)). The amount of data usage is the decision factor by each service subscriber. In user utility, the market-related parameters are speed satisfaction, coverage size and unit price of data. Among these values, speed satisfaction and the amount of data usage are evaluated in a fairly long period like a month. In this manner, a short service downtime or poor quality experience do not have a sudden effect on user payoff. This approach helps us to solve the higher level decision-making problem since it is separated from short-run power or bandwidth allocation problems. Accordingly, each user tries to maximize the total obtainable payoff regarding the general service parameters. Then the optimal pricing strategy of each provider is evaluated based on the actions of users. This formulation eventually leads to a two-stage Stackelberg game.

Besides characterizing the interactions between the users and the provider, we also model the provider-provider interactions as a competition among all providers in the market (which obviously includes the first type of interaction, user-provider as well). At this step, computing Nash equilibrium prices helps to understand the market state and paves the way to analyze cooperative markets. We also analyze a numerical example that shows the effect of data usage cost on the provider's profit at equilibrium. Since the model is based on general perceptions of speed satisfaction and the coverage size, one advantage of our approach is that no provider needs to know the exact coverage and bandwidth details of other providers to formulate its profit and payoff of users.

After constructing a model for the competitive state of the market, in Section 2.7, we use this model to develop an analytical framework for cooperative markets which helps us to analyze possible coalitions. Towards this end, we redefine the uni-provider utility of the users to accommodate a new concept of multi-provider network. The analytical approach that is considered in this part and the form of multi-provider payoff function is an important part of our contribution. Here, the criterion for analyzing user-provider interaction remains the same but there are two new challenges: finding stable coalitional structures and determining allowable coalitions from the regulatory viewpoint. The first problem is related to two characteristics of the wireless market: 1) negative externalities which cause the profit of any coalition to depend not only on its members but also on the reactions of other non-coalition members; 2) asymmetrical nature of HetNets, which is caused by the existence of wireless providers with a broad range of market powers. To address these issues, we modify the well-known method presented in (Bloch, 1996) to enable us to develop a process of coalition formation (PCF) for wireless markets. From the regulator's viewpoint, the main issue is to define a weighted social welfare function and to estimate its value under each coalitional structure. We present several examples that illustrate the effect of pricing strategies and welfare standards on the cooperative behavior of the providers.

Structure

The rest of this paper is organized as follows: Section 2.1 discusses the related works. Basic notation is developed in Section 2.2 while operator selection criteria are analyzed in Section 2.3. Section 2.5 describes the models for evaluating the provider profits while the analysis and modeling of coalitions are presented in Section 2.7. Finally Section 2.9 discusses the conclusions and our future work.

2.1 Related Works

There are several research efforts that are focused on market models. The majority of literature that is related to the cooperative games tries to find a solution for the grand coalition. The grand coalition is formed with the participation of all entities. An example is the bandwidth allocation mechanism by access points (APs) in 4G HetNets introduced in (Niyato & Hossain, 2006). The bandwidth allocation and admission control are analyzed based on the model of N-person cooperative game in which the allocated bandwidth to each connection is based on Shapley value. In this type of sharing, coalition members benefit from the overall payoff based on their power or importance in the coalition. Cooperation among providers in cellular networks is modeled in (Singh *et al.*, 2011) and (Singh *et al.*, 2012b). The object of (Singh *et al.*, 2011) is modeling the market when the payoff to the providers is non-transferable. In this model, providers try to coordinate their actions and each of them obtains a payoff bigger than the one achieved in the non-cooperative case. The goal is to form a grand coalition and to find the *Core* of the game. The aim of (Singh *et al.*, 2012b) is to investigate cooperations in multi-hop networks where the optimal strategies for selecting the appropriate channels and base stations are found by solving several optimization problems. Here, the payoff function is modeled by *transferable utility*. The goal of this work is also to find the stable grand coalition. Our work, on the other hand, denies the formation of grand coalition. We also propose the concept of general speed satisfaction and by using it, we develop our market framework. (Anglano *et al.*, 2014) is a very interesting work which pursues a similar goal to our work. It investigates the formation of coalition among wireless providers in green networks. Toward this end, an algorithm for coalition formation is proposed and numerically analyzed. Our work has three major differences with this work. Firstly, while (Anglano *et al.*, 2014) is focusing on the problems like resource allocation and base station assignment, we focus on long-run problems like the overall consumed data by the users in monthly periods. Secondly, we define a multi-provider payoff function which gives the choice to users to select among available providers. Finally, in our work the formed coalitions are protected by long-run contracts and the providers cannot leave their coalition during the contract period. This is based on the fact that forming coalitions is a long process which needs co-investment in many technical aspects.

From the economic perspective, (Niyato & Hossain, 2009) investigates the competition among users of wireless heterogeneous networks when the available bandwidth is limited. The competition is modeled as an *evolutionary game*. Our model, on the other hand, investigates the competitive and cooperative strategies of the providers by considering user' long-run data consumption model and their network selection behavior. The goal of (Duan *et al.*, 2013c) is to analyze the interaction between wireless providers while they aim to upgrade their networks. In this manner, the best upgrade time along with the earned profit of the providers are investigated. However, this interesting research effort does not focus on the coalition of the providers.

The optimal pricing in the WiFi market has been studied in (Duan *et al.*, 2013a) where the revenue of a WiFi provider with flat rate and usage-based pricing has been analyzed. The market model covers the interaction between users, local WiFi providers and Skype (that wants to cooperate with local WiFi providers to build a global WiFi service). A duopoly wireless market has been modeled in (Jia & Zhang, 2008) where the competition of two providers and their interaction with users form a multi-stage game. Competition between providers for users with different payoff functions is the subject of (Gajic *et al.*, 2009). Here the market is considered to be heterogeneous and the problem is solved as a two stage leader-follower game. However, none of above works analyzes the cooperation among providers in the HetNets market.

Our work is distinguished from all above works with the fact that we consider the heterogeneity of the market by modeling the behavior of the users when they have different perceptions of satisfaction. Our work also considers both competitive and cooperative states of the market based on the user satisfaction model.

2.2 Basic Notation and Assumptions

We consider an unsaturated market where there is an incoming flow of new users who need to select their default provider. In each period of time (e.g., a day or month), the expected number of incoming users is a constant represented by N_u which is the size of related user set I_u. Each provider, from the set $Pr = \{1, 2, ...n\}$, serves users by one of several access technologies, e.g., WiFi, 4G or 3G. The new users sign a data service contract with a selected provider. Provider i has a geographic coverage area (GCA) A_i with size $|A_i|$. Then we define normalized size $G_i = \frac{|A_i|}{|A_T|}$, where A_T is the GCA of the entire market. Note that we have $|\cup_1^n A_i| = |A_T|$ and $\sum_1^n |A_i| > |A_T|$ due to the overlapping areas considered in our notation. Providers charge users based on the data usage of users and provider i's data unit price is shown by p_i. The price unit is defined for an agreed unit of data usage, e.g., 1 GB. The market price vector is $\bar{P} = \{p_1, ..., p_n\}$. Provider i's cost function, C_i, has two components. The first component, α_i, is related to the cost of one data unit used by a user and the second component, $c_i(G_i)$, is the constant cost related to coverage area. If the user-set of provider i in one period of time is I_i

and the data usage of j-th user is $d_i^j, \;\; j \in I_i$, then the total cost experienced by provider i for all of its registered users in that period is:

$$C_i = c_i(G_i) + \alpha_i \sum_{j \in I_i} d_i^j. \tag{2.1}$$

Note that the ability to expand the coverage area is related to the access technology of choice. For example, a WiFi access point (AP) can serve an area of around 50-meter radius and a 3G macrocell can cover an area with a radius of several kilometers. The satisfaction factor (SF) s_i^j for user j with provider i, is a number in $[0,1]$ that reflects the user satisfaction from the service speed. *Base satisfaction factor*, S_i, for the access technology of provider i is defined as the ratio of average speed (experienced by users of that technology) to the maximum speed expected by the greediest user. For instance, if the average access speed for WiFi, 4G and 3G cellular are 150, 50 and 15 Mbps respectively and the maximum expected speed is 300 Mbps, the corresponding base satisfaction factors for these technologies are: $S_{WiFi} = 0.5$, $S_{4G} = 0.166$ and $S_{3G} = 0.05$.

We assume for each provider, the user satisfaction factor is uniformly distributed from S_i to 1 with probability density function:

$$f_i(s_i) = \begin{cases} \dfrac{1}{1 - S_i} & s_i \in [S_i, 1], \\ 0 & else. \end{cases} \tag{2.2}$$

It is highly common to use a uniform valuation in economic analysis. We refer the readers to (Duan *et al.*, 2013b) and (Chen *et al.*, 2015b) as two well-known examples. The above assumption implies that users have different perceptions of the same technology, where the greediest users are represented by S_i which is the minimum satisfaction value and the ones with least bandwidth requirements are shown by $s_i = 1$. By applying this assumption we consider the heterogeneity of network applications. In fact, the subscribers who permanently use high-demand applications such as HD-video may have less satisfaction comparing to the ones with applications such as voice or web browsing. We suppose if the speed satisfaction (SS) of a specific user under a particular access technology is known then the SS can be computed for other technologies, even though the SS is a random variable itself. A linear transformation can be applied for this case: when user j's satisfaction from the technology of provider i is known (s_i^j) then this user's satisfaction for technology of provider m can be calculated as:

$$s_m^j = a_{i-m} s_i^j + b_{i-m} = \frac{1-S_m}{1-S_i} s_i^j + \frac{S_m - S_i}{1-S_i}, \tag{2.3}$$

where a_{i-m} and b_{i-m} are the transition constants with respect to the destination network m and originating network i. If a network provider changes the quality of its service, then the speed satisfaction of each user will be affected respectively.

2.3 Operator Selection

As mentioned in the previous section, in each long period, there is a fixed expected number of users, N_u, who join the market. They select their desired provider based on the level of achievable payoff. Users sign contracts with the selected providers and are charged based on their data usage. Providers have enough long-run capacity to fulfill the total demand of their corresponding users.

2.3.1 Payoff function for the users

The payoff function is the difference between the user utility (representing satisfaction from the service) and the service cost. The utility of user j is proportional to the coverage of its provider (G_i) and the user satisfaction from service speed (s_i^j). Moreover, it should follow the *law of diminishing marginal utility*. Hence, the payoff function can be defined as:

$$U_i^j(G_i, p_i, s_i^j, d_i^j) = G_i s_i^j \ln(1 + K d_i^j) - p_i d_i^j. \tag{2.4}$$

Table 2.1 Notation

SYMBOL	DEFINITION
N_T	Number of users in the stationary market
N_u	Number of new users joining the market in each period
$[\,]_i^j$	Attribute of provider i and user j
A_i	Coverage area of provider i
G_i	Normalized size of A_i
S_i	Normalized technology speed of provider i
I_i	Set of new users joining provider i in each period
p_i	Current price of provider i
P_i	Price strategy set of provider i
s_i^j	Satisfaction factor (SF) of user j with provider i
d_i^j	Amount of data used by user j with provider i
D_i^j	Maximum amount of data used by user j with provider i
$f_i(s)$	PDF of user satisfaction for provider i
α_i	Cost of providing one unit of data (Provider i)
$c_i(G_i)$	Constant cost for provider i
V_i^j	Maximum payoff of user j in the network of provider i
$\pi_i(k,m,.)$	Profit of provider i in each period with respect to the parameters k,m, etc
K	Shape factor of payoff function
CS	Coalition Structure
$\mathscr{W}$	Social Welfare
SWF	Social Welfare Function
UWF	WiFi Usage Willingness Factor
S_{WiFi}	Base satisfaction factor for WiFi
S_{3G}	Base satisfaction factor for 3G
S_{4G}	Base satisfaction factor for 4G

This payoff function is linear with regard to s_i^j and G_i, and is concave with respect to d_i^j. Constant K is the shape factor related to the price elasticity of demand (Duan *et al.*, 2013a) (We will study the value K in Subsection 2.4). The value of payoff function defines the user gain in one period (e.g., a month) and it is used to select a provider for user j. Note that we adopt the common practice (see e.g., (Başar & Srikant, 2002), (Sengupta *et al.*, 2007) and (Duan *et al.*, 2013a)) of using a logarithmic utility function. Throughout this paper we frequently use simpler notation for user payoff such as $U_i^j(d_i^j)$. We also use the superscript j to indicate a specific user's satisfaction value and payoff.

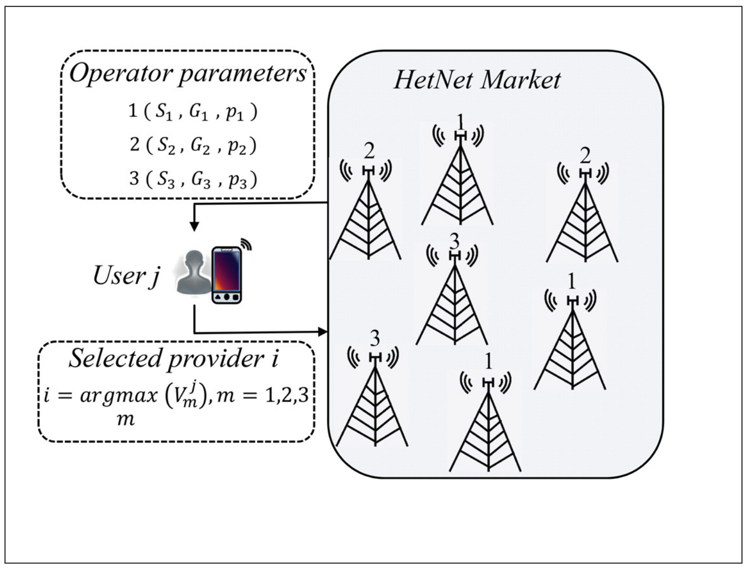

Figure 2.1 Illustration of a three-provider market where a new user selects the default provider maximizing its payoff.

2.3.2 Provider selection mechanism

When user j enters the market, it compares its maximized payoffs for all providers and selects the one maximizing its payoff, i.e.,

$$i^j = \arg\max_m V_m^j, m \in Pr = \{1, ..., n\}, \tag{2.5}$$

where

$$V_m^j = \max\{\max_{d_m^j} U_m^j(d_m^j), 0\}, \tag{2.6}$$

$$\arg\max_{d_i^j} U_i^j(d_i^j) = \frac{s_i^j G_i}{p_i} - \frac{1}{K}, \tag{2.7}$$

$$D_m^j = \max\{0, \frac{s_m^j G_m}{p_m} - \frac{1}{K}\}, \tag{2.8}$$

$$\max_{d_m^j} U_m^j(d_m^j) = s_m^j G_m \ln(\frac{s_m^j G_m K}{p_m}) - (s_m^j G_m - \frac{p_m}{K}). \tag{2.9}$$

D_m^j is the maximum data usage of user j and V_m^j is the maximum obtainable payoff by j from provider m.

2.4 Shape factor (K) and the price elasticity of demand (PED) in stationary markets

As we mentioned, K is related to the price elasticity which defines the usage response of subscribers to a price change. For user j and provider i, PED is defined as:

$$\mathcal{E}_i^j = \frac{p_i}{d_i^j} \frac{\partial d_i^j}{\partial p_i}. \tag{2.10}$$

We argue that the rational users maximize their payoff, hence, substituting D_i^j from (2.8) for d_i^j in (2.10) yields:

$$\mathcal{E}_i^j = \begin{cases} \dfrac{-s_i^j G_i K}{s_i^j G_i K - p_i} & \text{if } D_i^j > 0, \\ 0 & \text{else.} \end{cases} \tag{2.11}$$

Note that for $K \gg 1$, PED is close to one for all users while for $K = \frac{p_i}{s_i^j G_i}$, PED is equal to infinity which corresponds to the case of perfect elasticity. Since in a network, users have different perceptions of satisfaction, the average over (2.11) with respect to the range of satisfaction in user set I_i, gives the average PED in the network of provider i. To calculate $\mathcal{E}_i$, which is the true value of PED in regard to provider i, one computes the sum of all demands requested by the users $\left(\text{henceforth also called aggregate demand } D_i^T(G_i, p_i)\right)$ and substitutes it into the following equation:

$$\mathscr{E}_i = \frac{p_i}{D_i^T(G_i, p_i)} \frac{\partial D_i^T(G_i, p_i)}{\partial p_i}. \tag{2.12}$$

Let the demand of a generic user with satisfaction s be:

$$D_i(s, G_i, p_i) = \frac{sG_i}{p_i} - \frac{1}{K}, \tag{2.13}$$

then, for a given number of users N_T, the aggregate demand for users with satisfactions density $f_i(s)$ is given by:

$$D_i^T(G_i, p_i) = N_T \int_{s \in I_i(p_i)} D_i(s, G_i, p_i) f_i(s)\, ds. \tag{2.14}$$

If $I_i = [s_1, s_2]$ and $f_i(s) = \dfrac{1}{1 - S_i}$ $\forall s \in [S_i, 1]$, then:

$$D_i^T = \frac{N_T}{1 - S_i} \left(\frac{(s_2 - s_1)^2 G_i}{2\, p_i} - \frac{s_2 - s_1}{K} \right). \tag{2.15}$$

We further define the maximum and minimum usage among the user set for provider i to be:

$$Max_d = \frac{s_2\, G_i}{p_i} - \frac{1}{K}, \tag{2.16}$$

$$Min_d = \frac{s_1\, G_i}{p_i} - \frac{1}{K}. \tag{2.17}$$

Since in a stationary market, provider i knows the market related values D_i^T, Max_d and Min_d, (2.15 - 2.17) can be used to determine the values of s_1, s_2 and K. In particular:

$$K = \frac{2N_T\, p_i(Max_d - Min_d)}{2G_i D_i^T (1 - S_i) - N_T\, p_i(Max_d - Min_d)^2}. \tag{2.18}$$

If the user set $I_i = [s_1, s_2]$ does not change under a small price variation, we can derive an approximation for aggregate PED as:

$$\mathscr{E}_i \approx -\left(\frac{G_i K(s_2 - s_1)}{G_i K(s_2 - s_1) - 2p_i}\right). \tag{2.19}$$

Note that $\mathscr{E}_i$ is equal to the average PED over s_i thanks to the uniform satisfaction distribution.

2.5 Provider profit in different market forms

2.5.1 Monopoly

In a monopoly market, there is only one provider (indexed here by subscript 1) that serves the market with a single type of access technology. Users adapt their usage based on the service price and the monopolist wants to set the price that maximizes its profit. This type of market can be analyzed as a leader-follower game or a two-stage Stackelberg game (Fudenberg & Tirole, 1991). In the first stage, the provider sets its price anticipating the reaction of the rational users. This type of game can be solved by backward induction (Fudenberg & Tirole, 1991). Based on the data usage of individual users derived from (2.8), all users with $d > 0$ are in I_1 defined as:

$$I_1 = \left\{ s_1^j \in [0,1] \,\middle|\, D_1^j = \frac{s_1^j G_1}{p_1} - \frac{1}{K} > 0 \right\}, \tag{2.20}$$

$$s_1^j \in I_1 \to D_1^j = \frac{s_1^j G_1}{p_1} - \frac{1}{K} > 0 \to \frac{p_1}{G_1 K} < s_1^j \leq 1. \tag{2.21}$$

This equation implies that:

$$p_1 < G_1 K. \tag{2.22}$$

Hence, the profit of provider due to the new users in one period is given by:

$$\pi_1 = N_u \int_{s \in I_1} D_1(s) \times (p_1 - \alpha_1) \times f_1(s)\, ds - c_1(G_1) =$$

$$N_u \int_{\frac{p_1}{G_1 K}}^{1} (\frac{s\,G_1}{p_1} - \frac{1}{K}) \times (p_1 - \alpha_1) \times f_1(s)\, ds - c_1(G_1). \tag{2.23}$$

The concavity of profit in the region of $p_1 < G_1 K$ requires:

$$\frac{\partial^2 \pi_1}{\partial p_1^2} = \frac{N_u}{1 - S_1} \left(\frac{1}{G_1 K^2} - \frac{G_1\,\alpha_1}{p_1^3} \right) < 0. \tag{2.24}$$

Then, since $p_1 \geq \alpha_1$, by substituting the minimum amount $p_1 = \alpha_1$ we have:

$$\frac{1}{G_1 K^2} - \frac{G_1\,\alpha_1}{\alpha_1^3} < 0 \rightarrow \alpha_1 < G_1 K. \tag{2.25}$$

Note that Inequality (2.25) always holds as a result of Inequality (2.22). Hence, the profit function is concave and has a maximum value. Denote optimum value of p_1 by: $p_1^* = \min(\max(\max(p_1^r), 0), G_1 K)$, where:

$$p_1^r = \left\{ p_1 \in \Re \;:\; p_1^3 + (\frac{\alpha_1}{2} - \frac{G_1}{K})\, p_1^2 + \frac{\alpha_1\, G_1^2 K^2}{2} = 0 \right\}. \tag{2.26}$$

The set p_1^r contains the real roots of cubic equation and in order to find them, we need to define Δ and Δ_0 as follows:

$$\Delta = -2(\alpha_1 G_1^2 K^2) \times \left(\left(\frac{\alpha_1 K - 2G_1}{2K} \right)^3 + \frac{27\alpha_1 G_1^2 K^2}{8} \right), \tag{2.27}$$

$$\Delta_0 = \left(\frac{\alpha_1 K - 2G_1}{2K} \right)^2. \tag{2.28}$$

For $\Delta < 0$ there are three real roots. $\{\Delta = 0$ and $\Delta_0 = 0$ $\}$ leads to a multiple root. $\{\Delta = 0$ and $\Delta_0 \neq 0$ $\}$ presents a double root as well as a simple root and finally, $\Delta > 0$ causes one real and two complex roots. Since G_1 and K are both nonzero positive values, (2.27-2.28) show that Δ and Δ_0 cannot be 0 concurrently. Thus, we can define the following inequalities for root conditions:

$$\begin{cases} \alpha_1 > \frac{2G_1}{K} - 3(\alpha_1 G_1^2 K^2)^{\frac{1}{3}} & \text{3 real roots,} \\ \alpha_1 = \frac{2G_1}{K} - 3(\alpha_1 G_1^2 K^2)^{\frac{1}{3}} & \text{1 double and 1 simple root,} \\ \alpha_1 < \frac{2G_1}{K} - 3(\alpha_1 G_1^2 K^2)^{\frac{1}{3}} & \text{1 real root.} \end{cases}$$

Finding the roots of a cubic equation is well-documented and to avoid redundant content in this paper, we refer the readers to (Neumark, 2014) for a comprehensive analysis.

2.6 Duopoly

In a duopoly market, two dominant providers exist and they may use the same or different access technologies. Fig. 2.2 shows an example of the maximum-payoff curves, associated with (2.9), for each network of the two providers as a function of the user satisfaction value s_i. Note that in this case the two curves cross. Otherwise, since the maximum payoff offered by one provider is always bigger than its competitor, there would be no competition. Let s_1^* be the user satisfaction value corresponding to provider 1 at the crossing point of payoff functions. Then, supposing that provider 1 offers slower maximum speed, s_1^* can be determined by solving the following equation:

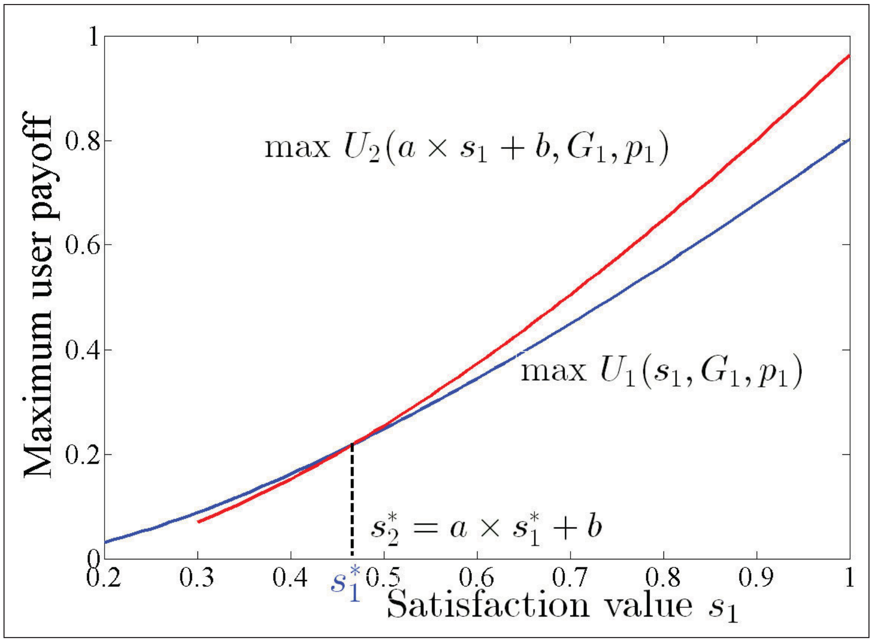

Figure 2.2 The curves of maximum user payoff and their intersection point in a duopoly. The curve related to provider 2 is mapped to the satisfaction space of provider 1 based on the Assumption 2

$$\max_d U_1(s_1, G_1, p_1) = \max_d U_2(as_1 + b, G_2, p_2), \tag{2.29}$$

$$a = \frac{1 - S_2}{1 - S_1}, b = \frac{S_2 - S_1}{1 - S_1},$$

where $\max_d U_i(s_i, G_i, p_i)$ for the provider i is:

$$s_i G_i \ln(\frac{s_i G_i K}{p_i}) - (s_i G_i - \frac{K}{p_i}). \tag{2.30}$$

Considering Assumption 2, we used $s_2 = as_1 + b$ to match the satisfaction values of each user in the two networks. Since the maximized payoff function is concave over s_i, there can be one or two intersection points in the interval $s_i = [S_1, 1]$. Suppose that there is one intersection point in the mentioned interval (the same approach can be applied for the case of two intersection points), then, the profit functions related to new users (N_u) are as follows:

$$\pi_1 = N_u \int_{\gamma}^{\zeta} (\frac{s\,G_1}{p_1} - \frac{1}{K})(p_1 - \alpha_1) f_1(s)\,ds - c_1(G_1), \tag{2.31}$$

$$\gamma = \max\left\{S_1, \frac{p_1}{G_1 K}\right\},\ \zeta = \max\left\{s_1^*, \frac{\frac{p_2}{G_2 K} - b}{a}\right\}, \tag{2.32}$$

$$\pi_2 = N_u \int_{\beta}^{1} (\frac{s\,G_2}{p_2} - \frac{1}{K})(p_2 - \alpha_2) f_2(s)\,ds - c_2(G_2), \tag{2.33}$$

$$\beta = \max\left\{a s_1^* + b, \frac{p_2}{G_2 K}\right\}. \tag{2.34}$$

In these profit formulations, we addressed the incoming users only. This is due to the fact that providers are willing to adjust the service prices for the new users; previously subscribed users are already bound by a contract. This implies that the new users can have different service prices from the old users. Obviously, the total profit in each period is related to both of these groups. If the price is changed for all users in the market, then N_u is substituted by N_T. Now it is important to find the best price and coverage size for each provider. This problem can be tackled by finding the Nash equilibrium of the system, as explained in the next subsection.

2.6.1 Oligopoly

In an oligopoly market, a small group of dominant providers exists. We assume that these providers may use different access technologies. In general, oligopoly market consists of few market leaders and several followers. The difference between the two types of providers is their market power where only the leaders have the power of price setting for the services. The followers adjust their prices based on the price of leaders. The market power of leaders comes from their monetary resources and the number of subscribed users. Hence, when we analyze the oligopoly market, we assume that the majority of users tend to subscribe to the networks of few oligopolist. Moreover, the only concern of users is maximizing their payoff and they do not prioritize the oligopolist networks based on the decisions of other users. Hence, the interaction between the users and providers is considered as a strategic game. In contrast to the case of monopoly market, the closed-form expression for providers' profits under an oligopoly market cannot be readily derived. Therefore, we resort to a numerical analysis (in *Supplementary Materials*, an analytical approach for the case of duopoly is proposed). First, we define some basic notions from game theory which are relevant to our analysis:

Definition 2.1. A provider full competition game $G(Pr,(P_i)_{i\in Pr},\pi_{i_{(S_i,G_i,(p_i,p_{-i}))}})$ consists of:

a. A set of providers, $Pr=\{1,...,n\}$.

b. An action (price) set P_i for each provider.

c. Actions $p_i \in P_i$ for provider i.

d. Profit functions $\pi_{i_{(S_i,G_i,(p_i,p_{-i}))}}$, $P \longrightarrow R$, where $P=\prod_{i=1}^{n} P_i$ is the set of all actions (possible prices) of providers and p_{-i} is the actions vector of all providers except i.

e. Action profiles (p_i, p_{-i}).

Definition 2.2. A pure strategy Nash equilibrium (one action for each player) of game $G(Pr,(P_i)_{i\in Pr},\pi_{i_{(S_i,G_i,(p_i,p_{-i}))}})$ is an action profile $p^* \in P$ such that:

$$\pi_{i_{(S_i,G_i,(p_i^*,p_{-i}^*))}} \geq \pi_{i_{(S_i,G_i,(p_i,p_{-i}^*))}} \ \forall p_i \in P_i. \tag{2.35}$$

Debreu-Gilcksberg-Fan Theorem. A strategic game with a compact and convex set of strategies, and continuous quasi-concave payoff function $\pi(i,(p_i,p_{-i}),G_i)$ over p_i which is also continuous over p_{-i}, has a pure Nash equilibrium (Fudenberg & Tirole, 1991).

Then we introduce the following theorem:

Theorem 1. The full competition game (no coalitions) always has a pure Nash equilibrium.

Proof. See Appendix II-1. □

In order to find the Nash equilibria of this game, we resort to a numerical approach by constructing an n-dimensional matrix of all possible payoffs in the game and using well-known schemes such as Lemke-Howson's algorithm or global Newton method. This approach is Algorithm 2.1 in which we determine provider profits of the user set and use Gambit (McKelvey *et al.*, 2013) (a suite of software tools for noncooperative games) to find a Nash solution. The algorithm first tries to find the pure strategies and uses discretized price and satisfaction sets. If it does not find a pure strategy equilibrium, it will take a smaller discretization value and will repeat the procedure. In our simulation experiments, with sufficiently small discretization value, the NE is always unique. Note that the result of this procedure remains valid until the general characteristics of the provider networks change. Besides prices, it is interesting to know the coverage expansion behavior of the providers when they experience different cost functions. The following two propositions consider two different forms of cost function:

Proposition 1. If the constant part of the price is of the form $c_i(G_i) = \zeta_i \times G_i$ and the provider i already has $\pi_i > 0$, then, if enough monetary resources are available, it is optimal for the provider to set $G_i = 1$.

Proof. Suppose that users in interval $[s_1(G_i), s_2(G_i)]$ are registered to provider i, then, provider profit is given by:

$$\pi_i = N_u f_i(s)\left(\left(\frac{G_i\left(\left(s_2(G_i)\right)^2 - \left(s_1(G_i)\right)^2\right)}{2p_i} - \frac{s_2(G_i) - s_1(G_i)}{k}\right) \times (p_i - \alpha_i)\right) - \zeta_i G_i, \tag{2.36}$$

that can be rewritten as:

$$\begin{aligned}\pi_i =& (G_i(\nu - \zeta_i) - (N_u f_i(s)\frac{s_2(G_i) - s_1(G_i)}{k} \times (p_i - \alpha_i)), \\ \nu =& N_u f_i(s)\frac{G_i(s_2(G_i)^2 - s_1(G_i)^2)}{2p_i}(p_i - \alpha_i)).\end{aligned} \tag{2.37}$$

Algorithm 2.1 Profit vector calculator

Input :

Provider set $Pr = 1, 2, ..., n$.
Price interval $P_i = [P^i_{\min}, P^i_{\max}]$ for each provider $i \in Pr$.
Default discretization value q.
Satisfaction pdf $f_i(s)$ for each provider $i \in Pr$.

Output:

Pure strategy equilibrium price and profit vector

for *each* $i \in Pr$ **do**
    compute P_i^q from P_i as the discretized price interval based on discretization factor q ;
    Find $S_{\min} = \min\{S_1, ..., S_n\}$;
    Discretize the interval $[S_{\min}, 1]$ by reasonable interval like 0.001;
    Compute the strategy space $P^q = P_1^q \times P_2^q ... \times P_n^q$;
end
for *each vector* $pq(1, ..., n) \in P^q$ **do**
    for *each* $s \in [S_{min}, 1]$ **do**
        $\text{selected}_{pr} \leftarrow \arg\max_i V_i \; I_{\text{selected}_{pr}} \leftarrow a_i s + b_i$
    end
    for *each* $i \in Pr$ **do**
        Compute $\pi(i, pq, G_i)$
    end
    $\Pi(pq) \leftarrow [\pi(1, pq, G_1), ..., \pi(n, pq, G_n)]$
end
Compute Nash equilibrium and corresponding strategy set;
if *There is no pure strategy set,* **then**
    $q = q/2$;
    Go to 1 ;
end
else
    Return current strategy set and equilibrium profit vector;
end

Based on the provider selection criteria, we know that by increasing the coverage size, the interval $[s_1(G_i), s_2(G_i)]$ can only grow. Hence, when G_i increases in (2.37), the profit also increases. $\square$

Note that even if the condition of this proposition was satisfied, many providers in real markets would not have enough financial resources for network expansions. This motivates our study presented in the next section where we conduct a game-theoretic analysis of the formation of provider coalitions without network infrastructure investments.

Proposition 2. Let $c_i(G_i) = e^{\beta_i G_i}$, if provider i already has a positive profit ($\pi_i > 0$), then, for known large values of β_i, the optimum coverage size can be smaller than 1 (in contrast to Proposition 1).

Proof. We prove this proposition for the case of monopoly since the providers in other market forms are under the limits of monopolist. To clarify, in all other forms of market, providers achieve lower levels of profit comparing to monopoly. Hence, if we prove that the optimum normalized coverage size for the monopolist is smaller than 1, it is valid for all other forms as well. One can write the following equation for a monopolist:

$$\pi_i = \frac{N_u}{1-S_i}\left(\frac{G_i}{2p_i} + \frac{p_i}{2G_i k^2} - \frac{1}{K}\right)(p_i - \alpha_i) - e^{\beta_i G_i} \longrightarrow \frac{\partial^2 \pi(i, G_i)}{\partial G_i^2} = \frac{p_i}{K^2 G_i^3} - \beta_i^2 e^{\beta_i G_i}. \quad (2.38)$$

Then, by using the second derivation test, it can be verified that π_i is concave over G_i for large values of β_i. Therefore, there is a $G_i < 1$ that maximizes the profit for those values of β_i. □

The immediate consequence of Proposition 2 is that when the features of access technology lead to an exponential growth of costs in case of coverage area expansion, it is not always beneficial to make investments for achieving full network coverage.

2.7 Coalition formation

In the previous sections, we have derived economic models for the competitive state of HetNets market by using user payoff functions and optimized data usage that is linear with respect to the coverage and access speed of service. Thus, provider profit can be boosted via capital investments that increase the network coverage and/or the adoption of new high-speed access technologies. Both approaches can be quite capital intensive and thus could reduce profits significantly (at least in the short term). Another way to increase the offered speed and coverage is to form coalitions among providers in a way that is beneficial to all coalition members. The coalitions can be based on 1) resource sharing or 2) serving users of one another in limited coverage areas without sharing resources. The approach presented in this section is valid for both methods.

In this section, we propose a framework to analyze coalition formation. It is based on the preferences of users and a provider selection mechanism. These methods are the extensions of models that are presented for competitive markets in the previous sections. In particular, we redefine the payoff function to take into account the existence of coalitions and extract its general properties. Then, we use the new payoff function and generated profits (via Algorithm 1) to develop a coalition formation process which leads to stable coalitions. It should be also noted that the market constraints (such as the existence of negative externalities, asymmetrical nature, and coalition size restrictions) prevent providers from forming a grand or very powerful coalition. The proposed coalition formation process considers these constraints. In the remainder of this section, we first define the user-payoffs under coalition followed by the presentation of the coalition formation process .

2.7.1 Multi-provider payoff for single price networks

Since the utility model of users under coalitions is a foundation to construct a model for the provider profits, we need to redefine the uni-provider payoff function defined in (2.4). The multi-provider payoff function should have the following properties:

a. When users switch between different networks, MPP function must be compatible with the *law of diminishing marginal utility*. Hence, consuming each extra unit of bandwidth gives a lower level of satisfaction compared to the beginning of the usage. In other words, the slope of the payoff decreases with respect to the usage and should be considered when the provider is changed.

b. The user payoff for a two-provider coalition having identical prices $p_1 = p_2$, and technology speed satisfaction, $s_1 = s_2$, should be equal to the uni-provider payoff function of a user with provider j that has coverage size $G_j = \frac{|A_1 \cup A_2|}{|A_T|}$ and the same price and speed. In other words, the coverage overlap should not be assumed twice when we compute users' usage and costs.

c. For each user, the MPP function is based on the coverage sizes of coalition providers. This feature is consistent with the form of uni-provider utility that is linear with respect to coverage size.

Theorem 2. For a coalition $C = \{1,2,...,n\}$, where each $i \in C$ corresponds to a provider with price p_i, coverage size $G_i = \frac{|A_i|}{|A_T|}$ and technology speed S_i, if $A_i \cap A_k = \emptyset\ \forall i,k \in C$, then the three properties of a multi-provider payoff function holds if and only if the payoff function of a user j follows the form:

$$U_C^j(d) = (\sum_{i=1}^{|C|} G_i s_i^j) \ln(1+K\,d) - \frac{\sum_{i=1}^{|C|} p_i G_i}{\sum_{i=1}^{|C|} G_i} d. \tag{2.39}$$

Proof. See Appendix II-2. □

Corollary 1. If coalition C satisfies the conditions of Theorem 2, then the coalition can be represented as a single provider with the following price and cost equations:

$$p_C = \frac{\sum_{i \in C} G_i p_i}{\sum_{i \in C} G_i}, \tag{2.40}$$

$$\alpha_C = \frac{\sum_{i \in C} G_i \alpha_i}{\sum_{i \in C} G_i}, \tag{2.41}$$

$$f_C(s_C) = \begin{cases} \dfrac{1}{1-S_C} & s_C \in [S_C, 1], \\ 0 & else\,. \end{cases} \tag{2.42}$$

$$S_C = \left\{ S_i \ , \ i \in C \mid S_i \leq S_j \forall j \in C \text{ and } \ j \neq i \right\}. \tag{2.43}$$

Proof. Firstly, since the unique price and costs do not change the form of payoff function for each user, neither would the overall profit output for the provider, the only part that needs an explanation is the distribution function of satisfaction in the coalition. Since in the considered coalition we can have users with different satisfaction levels, the distribution function of satisfaction should cover all satisfaction levels as it is given by $f_C(s)$ in (2.42). Note that for the uniform distribution, a wider range of satisfaction yields a lower level of the probability density $\frac{1}{1-S_i}$. □

Lemma 1. Usage Rationality Lemma: Suppose that some providers of a coalition $C = \{1,\dots,n\}$ have coverage overlap and each user j has possibly different speed satisfactions denoted by $\{s_1^j, s_2^j, \dots, s_n^j\}$ for each of the n providers in C. Then even if the user is served with expected speed by the default provider, there are some price conditions which drive a user to switch between operators in the overlapping area to maximize its payoff.

This lemma states that if a user has a better marginal payoff in a specific network at the beginning of the usage period, it could change the default network afterward to increase the user payoff. In what follows, we describe and prove the case of a two-provider coalition mathematically.

Let the coverage area, price and user's speed-satisfaction of provider i be $G_i = \frac{|A_i|}{|A_T|}$, p_i and s_i^j respectively. We have the following conditions:

C(1). The first condition states that at the beginning, user j gets a better payoff from one of the networks, say 1. Hence, in the overlapping area where the total coverage difference is not

$$U_C^j(d) = \begin{cases} (G_1 s_1^j + G_{l2} s_2^j)\ln(1+Kd) - \dfrac{G_1 p_1 + G_{l2} p_2}{G_1 + G_2 - G_i} d & d < d_T \\ (G_1 s_1^j + G_{l2} s_2^j)\ln(1+Kd_T) + (G_2 s_2^j + G_{l1} s_1^j)(\ln(1+Kd) - \ln(1+Kd_T)) & d > d_T \\ -\dfrac{G_1 p_1 + G_{l2} p_2}{G_1 + G_2 - G_i} d_T - \dfrac{G_{l1} p_1 + G_2 p_2}{G_1 + G_2 - G_i} (d - d_T), & \end{cases} \tag{2.47}$$

where $G_i = |A_1 \cap A_2|$, $G_{l1} = G_1 - G_i$ and $G_{l2} = G_2 - G_i$.

$$U_C^j(d) = (G_1 s_1^j + G_{l2} s_2^j)\ln(1+Kd) - \frac{G_1 p_1 + G_{l2} p_2}{G_1 + G_2 - G_i} d. \tag{2.48}$$

considered [1], we have:

$$\lim_{d\to 0} \frac{\partial U_1^j(d)}{\partial d} > \lim_{d\to 0} \frac{\partial U_2^j(d)}{\partial d} \Longrightarrow \lim_{d\to 0} \frac{s_1 K}{1+Kd} - p_1 > \lim_{d\to 0} \frac{s_2 K}{1+Kd} - p_2 \Longrightarrow$$
$$K(s_1 - s_2) > p_1 - p_2. \tag{2.44}$$

C(2). The second condition is that, at a specific level of usage, user j has the same preference for using either one of the provider networks. At this usage level, denoted by d_T, the slopes of payoff functions, regarding the network of both providers, are the same. Note that d_T is smaller than $d_{\max}(2) = \arg\max_d U_2^j(d)$:

$$d_T < d_{\max}(2) \rightarrow \frac{s_1 - s_2}{p_1 - p_2} < \frac{s_2}{p_2}. \tag{2.45}$$

C(3). Switching between providers means that user obtains a larger payoff:

$$\max_d U_2^j(d) > \max_d U_1^j(d) \cdot \tag{2.46}$$

[1] Here the coverage factor is eliminated because the user is already in the overlap area and wants to choose one of the networks to use and hence, it senses the same coverage at this point

Proof. Based on three above conditions, Lemma 1 can be proven by constructing an example which fulfills all three conditions and in which the user payoff is greater when the user uses a mix of two networks compared to the payoff from a single network. Fig. 2.3 illustrates such an example with $p_1 = 0.08$, $p_2 = 0.025$, $s_1 = 0.76$, $s_2 = 0.53$ and $K = 0.7$. Observe from the figure that this setting satisfies the 3 conditions and the mixed payoff of the user is given by:

$$\begin{aligned}\max U_{\{1,2\}}(d) &= U_1(d_T) + U_2(d_{\max}(2)) - U_2(d_T) = 1.0475 \\ &> U_2(d_{\max}(2)) = 0.9624 \\ &> U_1(d_{\max}(1)) = 0.7942.\end{aligned}$$

□

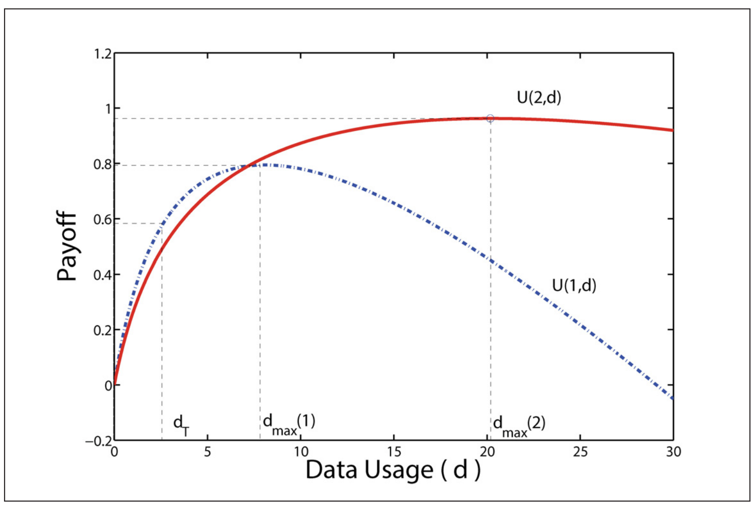

Figure 2.3 The payoff experienced by the user under two different networks

Lemma 1 shows an important characteristic of logarithmic utility functions: users at the beginning of their usage cycle are more sensitive to the speed and at the end they are concerned more about the cost than speed.

Theorem 3. The payoff for user j, with a 2-provider coalition $C = \{1,2\}$ that has coverage overlap, is as follows:

a. If conditions C(1), C(2) and C(3) all hold for the single provider payoff experienced by user j, then the payoff function is given by (2.47).

b. If any of C(1) to C(3) does not hold, then the payoff function for the user j is given by (2.48).

Proof. For the case where at least one of C(1) to C(3) does not hold, the marginal payoff obtained from Provider 1 is always better and the user prefers to utilize the first provider solely. However if all the three conditions hold, then the user begins to utilize the first provider (even in the overlap areas where both providers have coverage) until the usage reaches d_T. From this usage level, the user prefers to utilize the second provider. The proof for the form of the payoff is the same as Theorem 2 which states the function for the multi-provider payoff function (MPP) should follow the three main properties of the uni-provider payoff function. □

2.7.2 The multi-provider payoff function for dual price scheme networks

Note that the *flat rate* pricing scheme is commonly adopted by most WiFi providers. In this scheme, users pay a fixed price for a specific usage duration, e.g., an hour or a day. Since the wireless providers incur a cost for the amount of bandwidth that they provide, this method of pricing that offers an unlimited amount of data usage is useful for the cases where the provider has limited coverage (e.g., inside the airports and hotels) so that the registered users are not utilizing the bandwidth all the time.

To define the MPP function under a dual-pricing network, e.g., one which offers both 4G and WiFi services but charging differently for each technology (4G and WiFi), we need to define the average utilization level by the users when they have free access to the WiFi provider. Based on our assumptions, the network has stable speed and mobile users appear in all points of the coverage area in a long-run analysis. Thus, for the greediest users, the WiFi usage is given by:

$$D_W = \text{Period time (Sec)} \times \text{Usage per second } \times \text{Coverage size of WiFi.} \tag{2.49}$$

We can also extend our definition to the average data usage of each user by multiplying D_W with a usage willingness factor (θ) between 0 and 1 which indicates the expectation level of willingness (considering all the users) to fully utilize the WiFi network when users have free access to it, thus we have average data usage of: $\delta = D_W \times \theta$.

Example 2. By setting the WiFi speed at 150 Mb/s, the usage unit as one GB and the usage period as a month (which is a billing cycle), we have:

$$\begin{aligned} \delta &= 24 \times 30 \times 3600 \times (\frac{150}{8 \times 1024}) \times G_{WiFi} \times \theta \\ &= 47461 \times G_{WiFi} \times \theta. \end{aligned} \tag{2.50}$$

The MPP function for a dual-price network, defined by coalition $C = \{1,2\}$ where Provider 1 is a 4G and 2 is WiFi and the coverage area of WiFi is a subset of $4G$ area, is given by:

$$\begin{aligned} U_C^j(d_1, \delta) = &\left(\left((G_1 - G_2)s_1^j + G_2 s_2^j \right) \times \ln \left(1 + K(d_1 + \delta) \right) \right) \\ &- p_2 - p_1 d_1. \end{aligned} \tag{2.51}$$

This MPP is based on the idea that the WiFi coverage area is usually a subset of the coverage area of 4G cellular services. We also know that users switch from 4G to WiFi connections as soon as they are under the coverage of the latter due to its flat-rate pricing. Under this information, if subscript “1" and “2" refers to the 4G and WiFi provider respectively, the maximum cellular usage can be expressed as:

$$D_1^j = \max\left\{\frac{(G_1 - G_2)s_1^j + G_2 s_2^j}{p_1} - \delta - \frac{1}{K}, 0\right\}. \tag{2.52}$$

This equation defines the amount of cellular data usage in a WiFi-cellular utility function. Since we already defined a metric for the WiFi data usage, the cellular data usage is the amount of data beside the WiFi usage that can be consumed to maximize the utility. The reason behind fixing the WiFi usage (in the utility) is due to the nature of package pricing in the WiFi networks that is time-based. Hence, the user tends to utilize the WiFi network as soon as it is under the WiFi network coverage. However, the amount of WiFi data consumption is related to the WiFi coverage size and user's data consumption greediness which is considered in $UWF\ (\theta)$ of WiFi. To sum up, the value of marginal utility is getting lower by consuming more data which means that if a user has better access to the cheap WiFi data, it is less motivated to use the cellular data service.

2.7.3 Coalition formation

There are several models for coalition formation, each of them covering a specific type of game. (Hart & Kurz, 1983) proposed the γ-stable and δ-stable models where players announce their desired coalitions to join. In ((Yi, 1997)), a coalitional formation process has been studied that considers the externalities, but the model only applies to symmetric games. One of the most relevant models, that can be fitted to our framework, has been developed in (Bloch, 1996) where the externalities are considered and coalitions are constructed in a sequential manner. To consider this model, our definitions in the wireless market and the formation process are set as follows:

- A random provider starts the game and announces a desired coalition.

$$
\begin{aligned}
\pi_{C_{\{I_C,p_1,p_2\}}} =&\pi_{C_{\{I_C^1,p_1,p_2\}}} + \pi_{C_{\{I_C^2,p_1,p_2\}}} = \\
& N_u \Bigg(\int_{s_1 \in I_C^1} D_1(s_1) \times (p_1 - \alpha_1) + (p_2 - \alpha_2 \times \delta) \times f_1(s)\, ds_1 \\
& + \int_{s_1 \in I_C^2} \Big(\frac{s_1 G_1}{p_1} - \frac{1}{K}\Big)(p_1 - \alpha_1) f_1(s)\, ds_1 \Bigg) - c_1(G_1) - c_2(G_2),
\end{aligned} \tag{2.53}
$$

where:

$$
I_C = I_C^1 \cup I_C^2 \quad \text{and} \quad
\begin{cases}
I_C^1 = & \text{All } s_1^j \in [\max\{\dfrac{(\delta + \frac{1}{K})P_1 - b G_2}{G_1 + (a-1)G_2}, S_1\}, 1] \\
& \text{that } \max\limits_{d_1} U_C^j(d_1, \delta) \geq \max\{\max\limits_{d_1} U_1^j(d_1), 0\}, \\
I_C^2 = & \text{All } s_1^j \in [\max\{\frac{p_1}{G_1 K}, S_1\}, 1] \\
& \text{that } \max\limits_{d_1} U_1^j(d_1) > \max\{\max\limits_{d_1} U_C^j(d_1, \delta), 0\}.
\end{cases} \tag{2.54}
$$

$$
\begin{aligned}
& s_2^j = a s_1^j + b \\
& [.]_1 = \text{4G parameter}[.]_2 = \text{WiFi parameter}C = \text{coalition of } \{1,2\}
\end{aligned} \tag{2.55}
$$

- The expected profit of provider $i \in C$ is calculated based on the following equation:

$$
\begin{aligned}
& \pi_i = \phi_i \times \pi_C i \in C \quad , \quad C \in CS, \\
& \sum_{i \in C} \phi_i = 1.
\end{aligned} \tag{2.56}
$$

  where CS is a coalitional structure of which C is a member. The profit of provider i is a portion of the expected payoff, gained by the coalition that includes provider i. ϕ_i is a division factor agreed between coalition members (e.g., the Shapley value). This profit is calculated under the equilibrium price P^* (computed by Algorithm 1) while providers compete under the coalition structure CS.

- All providers who are included in the proposed coalition can agree or disagree. If all members agree, a coalition C forms and the game is continued with $Pr \backslash C$ players (Pr is the set of all providers). If a member disagrees, then it proposes its desired coalition in the next step. The size of coalitions is restricted by the maximum *Size-Limit*, defined by the regulator.

- Once a coalition with size *Size-Limit* has been formed its members cannot deviate and leave it. This is reasonable in today's market where implementing the infrastructure of cooperation requires a joint investment and is supported by long time contracts between providers.

Fig. 2.4 (page 14) depicts the coalition formation process which is proposed in Section VI-B of main paper.

Remark. The proposed coalition formation process leads to a coalition structure which does not include the grand coalition. Hence, the solution concepts for a grand coalition like Core and Shapley value are not applicable to all members of the structure concurrently.

Proposition 3. For the wireless markets in which with a known churn rate users leave the current providers for the newer technologies, the sub-game perfect equilibrium (SPE) of the coalition game always can be found.

Proof. See Appendix II-3. □

Example 3. In this example, we show the coalition structures and their existence conditions in a market of 3 providers ($1,2$ and 3). The same process can be applied to bigger networks. In the following we assume that the formation of grand-coalition, $\{1,2,3\}$, is not allowed due to monopoly avoidance rules. Then the set of all allowable coalition structures is:

$$\Big\{\pmb{\{}\{1\},\{2\},\{3\}\pmb{\}},\pmb{\{}\{1,2\},\{3\}\pmb{\}},\pmb{\{}\{1\},\{2,3\}\pmb{\}},\pmb{\{}\{2\},\{1,3\}\pmb{\}}\Big\}.$$

Let us start with $\{\{1\},\{2\},\{3\}\}$ case that represents the full competition. This structure is preferable if the following condition is true:

$$\underset{\substack{i\in\{k,m\}\\ \forall k,m\in\{k,m,l\}\\ CS=\{\{k\},\{m\},\{l\}\}}}{\pi_i} \geq \underset{\substack{i\in\{k,m\}\\ \forall k,m\in\{k,m,l\}\\ CS=\{\{k,m\},\{l\}\}}}{\pi_i}, \tag{2.57}$$

that means there is no coalition in which both providers can get bigger profit than their full competition state. This occurs when the providers cannot provide a higher payoff to the users, hence, all providers have the same technology and 100% coverage overlap. The second case is when a coalition of two providers can be formed and it gives a higher profit to both of them. Here, the worst-case scenario is that provider 1 prefers a coalition with 2, 2 prefers 3 and 3's choice is 1, which leads to a loop. In our analysis, the coalition profits are the outcome of extra data consumption and the maximum data consumption is related to the maximum payoff offered to the users. Thus, in regard to (2.6-2.9) one can write the maximum data and payoff preferences as:

$$\begin{aligned}
D^j_{\{1,2\}} > D^j_{\{1,3\}} \rightarrow V^j_{\{1,2\}} > V^j_{\{1,3\}} \forall j \in I_u,\\
D^j_{\{2,3\}} > D^j_{\{1,2\}} \rightarrow V^j_{\{2,3\}} > V^j_{\{1,2\}} \forall j \in I_u,\\
D^j_{\{1,3\}} > D^j_{\{2,3\}} \rightarrow V^j_{\{1,3\}} > V^j_{\{2,3\}} \forall j \in I_u,
\end{aligned}$$

which leads to a paradox:

$$\begin{aligned}
V^j_{\{1,2\}} > V^j_{\{1,3\}} \forall j \in I_u,\\
V^j_{\{1,3\}} > V^j_{\{1,2\}} \forall j \in I_u.
\end{aligned}$$

Hence, no loop exists in providers' preferences and one dominant coalition will be formed eventually. The last possibility is $\{V^j_{\{1,2\}} = V^j_{\{1,3\}} = V^j_{\{2,3\}}, p_1 = p_2 = p_3, \forall j \in N_u\}$. This condition is due to the existence of identical providers (in terms of size and technology) which have 0% coverage overlap. In this case, the first proposer forms the coalition with its preferred provider.

2.7.3.1 Regulatory unit policies on coalition formation

From the viewpoint of a wireless market regulator, the market should meet certain levels of fairness or at least wealth. To evaluate the market fairness, several concepts and corresponding metrics are developed in economics, each of them addressing a specific aspect of the market. For example, by one definition, the *social welfare (SW) function* measures the cumulative payoffs experienced by all entities in the market (Nisan *et al.*, 2007) and the *Gini coefficient* (Garetto *et al.*, 2008) shows the fairness level of wealth distribution which can be called justice. Note that a lower value of the *Gini coefficient* yields more uniformity of income (payoff and profit) distribution. In this paper, we use the social welfare function as our metric although the same approach could be applied to the Gini coefficient. We indicate the SW function by $\mathscr{W}$ and define it as follows:

$$\mathscr{W}(CS,\mathbb{W}) = \mathscr{W}^U(CS,\mathbb{W}) + \mathscr{W}^P(CS,\mathbb{W}) = \sum_{i\in CS}\left(\overbrace{w_u N_u \int_{s\in I_i} \max_{d_i(s)} U_i(d(s))\mathrm{d}s}^{\text{User part}} + \overbrace{w_p\,\pi_i}^{\text{provider}}\right), \tag{2.58}$$

$$\mathbb{W} = \{w_p, w_u\}, \tag{2.59}$$

$$w_p + w_u = 1. \tag{2.60}$$

Members of CS are coalitions and $\mathbb{W}$ is a weighting set.

Definition 2.3. *Suppose that the current coalition structure (status quo) is CS. If some providers decide to join other coalitions or form a new one (we call it a move), then under the new coalition structure CS^N, the move social efficiency factor, $\mathscr{M}$, for the new price equilibrium is* *defined as:*

$$\mathscr{M}(CS^N, \mathbb{W}^N, CS, \mathbb{W}) = \frac{\Big(\mathscr{W}^U(CS^N, \mathbb{W}^N) - \mathscr{W}^U(CS, \mathbb{W})\Big)\mathscr{W}^P(CS, \mathbb{W})}{\Big(\mathscr{W}^P(CS^N, \mathbb{W}^N) - \mathscr{W}^P(CS, \mathbb{W})\Big)\mathscr{W}^U(CS, \mathbb{W})}. \tag{2.61}$$

Definition 2.4. *From regulatory unit's viewpoint, a move is feasible if the conditions* $\mathscr{M}(CS^N, \mathbb{W}^N, CS, \mathbb{W}) \geq 0$ *and* $\mathscr{W}(CS^N, \mathbb{W}^N) - \mathscr{W}(CS, \mathbb{W}) > 0$ *are both true.*

The above two definitions specify the evaluation criteria for the possible moves in each coalition structure. Being aware of this information can be timesaving for the providers when they engage in negotiations. In particular, it shows the maximum coalition size and the possible coalitions that obey the regulator's criteria. Moreover, it indicates which coalition can maximize the profit.

2.8 Numerical Analysis

Scenario 1. In this example, we study the coalition of one cellular and one WiFi provider. As defined before, when a cellular service provider forms a coalition with a WiFi provider, the user payoff function can be represented by (2.51). By integrating the format of cellular maximum data usage in (2.52) and WiFi flat pricing model, we obtain Equation (2.53) as the profit function of the coalition $C = \{4G, \text{WiFi}\}$. For the sake of simplicity, we indicate all the parameters related to $4G$ with subscript 1 and for WiFi we use parameters with subscript 2. In this equation, the profit is separated into two parts associated with two different user sets, denoted by I_C^1 and I_C^2. These two sets are defined in (2.54). I_C^1 indicates the set of the users who utilize both WiFi and 4G network. The MPP of these users has a higher maximum level than their single provider payoff function. This set of users still pay for both WiFi and $4G$, however, in the next example we show that under a coalition the cost incurred is less. I_C^2 represents the set of the users who do not pay for coalition service and for which the single provider payoff function has a higher maximum level than MPP.

Table 2.2 Network settings for "one WiFi-one 4G" coalition

Property	**Value**
G_{4G}	0.5
G_{WiFi}	Variable
α_{4G}	$0.2\,G_{4G}$
α_{WiFi}	$[0.001, 0.01, 0.1]$
$UWF(\theta)$	$[0.01, 0.05, 0.1, 1]$
S_{4G}	$50/300$
S_{WiFi}	$150/300$
$K_i \quad \forall i \in C = \{4G, WiFi\}$	10

Scenario 2. To analyze a coalition of “WiFi-4G", let us consider the network settings presented in Table 2.2. Figures 2.5 and 2.6 show the coalition profit and aggregate payoff of the users as a function of the size of WiFi coverage area for three different WiFi bandwidth costs (per unit of consumed bandwidth) $\alpha_{WiFi} = [0.001, 0.01, 0.1]$. As seen in these figures, by increasing WiFi coverage size the profit of the coalition and the integrated user payoff initially increase and then coalition profit smoothly decreases to the no coalition level. At this point, the coalition is not profitable and should be terminated so that prices are reset to their non-coalition values (as shown in the no coalition zone in Fig. 2.6). This leads to an immediate drop in the user aggregate payoffs as shown in Fig. 2.5 (dropping points of two upper curves). At the maximum level, the coalition profit is 12% higher than the no-coalition profit and the aggregate user payoff improves by 10%. Also, in Fig. 2.6 we can observe that the pricing under the coalition can be a two-part tariff with equilibrium prices. For example, the cost of coalition C is given by:

$$\text{Cost for user j in Coalition C} = p_{C2} + p_{C1} \times d_{C1}^{j}.$$

Where p_{C1} is the price of $4G$ service under the coalition and p_{C2} is the price of WiFi under the coalition. Hence, there are three types of users in the market: $4G$ only, WiFi only and coalition users. However, as a result of the coalition, the price is dropped for all three groups (Fig. 2.6). Thus, two levels of long-run QoS are improved in this coalition; firstly, the coalition users have access to high-speed WiFi with cheaper price and secondly by transferring a part of traffic to the WiFi network, the price and network utilization of $4G$ provider are decreased which means a better $4G$ throughput overall.

Fig. 2.7 shows the coalition profit as a function of the size of WiFi coverage area for three levels of the usage willingness factor (UWF), $\theta = [0.01, 0.05, 0.1]$. As we discussed earlier, θ is an indicator for the usage greediness when users have free access to the network. The results indicate that with lower levels of θ, a WiFi provider with a larger coverage area can be accepted as a coalition member and with higher levels of θ, users have more incentive to utilize the WiFi network. Therefore, with bigger WiFi coverage sizes, the usage and profit of $4G$ service dramatically decreases since WiFi service is charged based on time and not data consumption. Hence, as θ increases the coalition is unprofitable under bigger sizes of WiFi coverage.

Table 2.3 Asymmetric 3-provider network settings

Property	**Value**
G_{3G}	0.5
G_{4G^1}(Bigger provider)	0.3
G_{4G^2} (Smaller provider)	0.1
S_{3G}	$\frac{15}{300}$
S_{4G} for both	$\frac{50}{300}$
$\alpha_i \quad \forall i \in \{4G^1, 4G^2, 3G\}$	$0.2\,G_i$
K	10

Scenario 3. This example considers an asymmetric market where two 4G operators are dominated by a 3G monopolist (4G operators gain much less profit than 3G provider). The context is realistic in wireless markets where it takes time for newer technologies to expand their cover-

age and at the same time their service area is only a proper subset of the older technology. The settings are given in Table 2.3. Concerning the total cost C_i (Equation (2.1)) of the providers, we used the linear form of $\alpha_i(G_i) = 0.2\,G_i$ in this numerical model and eliminated $c_i(G_i)$ for simplicity. Hence, the cost model is given by: $C_i = \sum_{j \in I_i} d_i^j \times 0.2\,G_i$.

Fig. 2.8 shows the simulation results as a function of the coverage overlap of 4G providers. It contains three sets of data. The first set of curves shows the profit of each provider in full competition state. The second set of curves gives the profits of the 4G-4G coalition and the 3G provider. The last curve represents the accumulated payoff of all users in the market.

The results in Fig. 2.8 show that the profits of smaller providers are negligible in full competition mode. This is due to the fact that they are overpowered by the bigger 3G provider. When 4G operators form a coalition, their profit increases while, at the same time, the 3G monopolist experiences profit loss. It can be also noted that when the coverage overlap of 4G operators is small, the coalition enjoys its maximum profit and 3G profit is at its minimum level. By increasing the overlap size, the profit of coalition decreases until it reaches the non-coalition level at 100% overlap. Thanks to the formation of the coalition, the competitiveness of 4G-4G coalition is increased, the market is more balanced and the users are enjoying an extra payoff. Also, it can be noticed that the user accumulated payoff is at its maximum level when the coverage overlap is small. This is due to the extra competitiveness of 4G-4G coalition powered by small coverage overlap. One hidden advantage of this type of coalition is the creation of an incentive for the market to upgrade to new technologies. This cooperation is also allowable by the regulatory unit since it gives better SW value. In contrast, the coalition of 3G and either one of the 4G providers would lead to a strong monopolist that blocks the profit of other provider. Therefore, such coalition would not be allowed by the regulator.

Scenario 4. To analyze a simple coalition formation game, we consider the settings presented in Table 2.4 where there are four symmetric providers that have coverage overlap. In this example, we assume 80% coverage overlap that is rational in the case of cellular providers. This is justified by the fact that cellular providers usually start to develop their new network in ur-

Table 2.4 Symmetric 4-provider network settings

Property	Value
$G_i \quad \forall i \in C = \{1,2,3,4\}$	0.5
$\frac{\lvert A_i \cap A_j \rvert}{\lvert A_T \rvert} \quad \forall i \in C = \{1,2,3,4\}$	0.4
$\frac{\lvert A_j \cap A_i \cap A_k \rvert}{\lvert A_T \rvert} \quad \forall i,j,k \in C = \{1,2,3,4\}$	0.35
$\frac{\lvert \cap_{i=1}^{4} A_i \rvert}{\lvert A_T \rvert}$	0.3
$S_i \quad \forall i \in C = \{1,2,3,4\}$	$50/300$
$\alpha_i \quad \forall i \in C = \{1,2,3,4\}$	$0.2G_i$
$K_i \quad \forall i \in C = \{1,2,3,4\}$	10

Table 2.5 Coalition structures and associated profits of Scenario 4

Coalition structure (CS)	**Provider profit /N_u**	**Aggregate user payoff /N_u**	$\mathscr{W}/N_u$ ($w_p = 0$, $w_u = 1$)	$\mathscr{W}/N_u$ ($w_u = 5w_p$)
{1,2,3,4}	{0.3079}	0.4464	0.4464	0.4233
{{i,j,k},m}	{{0.077},0.0052}	0.75	0.75	0.6387
{{i,j},{k,m}}	{{0},{0}}	0.76	0.76	0.6333
{{i,j},{k},{m}}	{{0.022},0.0025,0.0025}	0.72	0.72	0.6045
{{1},{2},{3},{4}}	{0,0,0,0}	0.70	0.70	0.5833

ban areas and then they expand the coverage area in the remaining years of that technology life-cycle. Hence, most of their coverage area is the same during the early stages of network build-up. Table 2.5 shows the market equilibrium profits for different coalition structures along with cumulated user payoff in the market. The symmetric structures like $\{\{i,j\}\{k,m\}\}$ lead to zero profit for the providers due to price wars. It is interesting that even in unbalanced structures, the integrated user payoff can increase due to coverage expansion. As it is depicted in Fig. 2.9, different weighting values lead to divergent feasible coalition structures. User friendly values like $\{w_p = 0, w_u = 1\}$ can cause more balanced coalitions like $\{\{i,j\},\{k,m\}\}$. By increasing the weight of providers in social welfare function i.e. in this example $\{w_p = 5w_u\}$, an imbalanced coalition structure like $\{\{i,j,k\},\{m\}\}$ can form. Note that parts a and b of Fig. 2.9 show the feasible transitions between different coalitional structures from regulator's viewpoint. These transitional diagrams are the tools used by the regulatory unit to define the market state based on different social welfare standards.

2.9 Conclusion and future work

In this paper, we proposed a game-theoretic framework that can help to set up stable coalitions of heterogeneous wireless operators and enable wireless regulatory bodies to determine acceptable coalitions which do not downgrade social welfare standards. We derived a simple and efficient generic model that predicts the state of the market before and after coalition formation without focusing on short time-scale bandwidth allocation problems. The model is based on finding a pure Nash equilibrium strategy profile that defines service prices for each provider. Our method is also based on the behavior of users and their satisfaction perceptions which are represented by the random utility. We showed that with specific types of cost functions, it is not beneficial for providers to expand their coverage above a certain size. In some other cases, if the financial resources are available, the provider is better off with full coverage on the market. We proved the form of multi-provider payoff (MPP) function for the coalitions. Based on such MPP functions, we constructed a modified version of the coalition formation process. Based on the proposed Lemma, we showed that in some cases it is beneficial for the users to switch between different networks to maximize their payoffs. Finally, the cost and coalition models are illustrated with several numerical examples. The results show that formation of coalition can notably increase the profits of providers while increasing the integrated payoffs of users. In the tested scenarios the profits and the integrated payoffs were increased by up to 12% and 10%, respectively. As future work, we plan to investigate different pricing schemes and provider selection mechanisms to further enhance and generalize our coalition formation process.

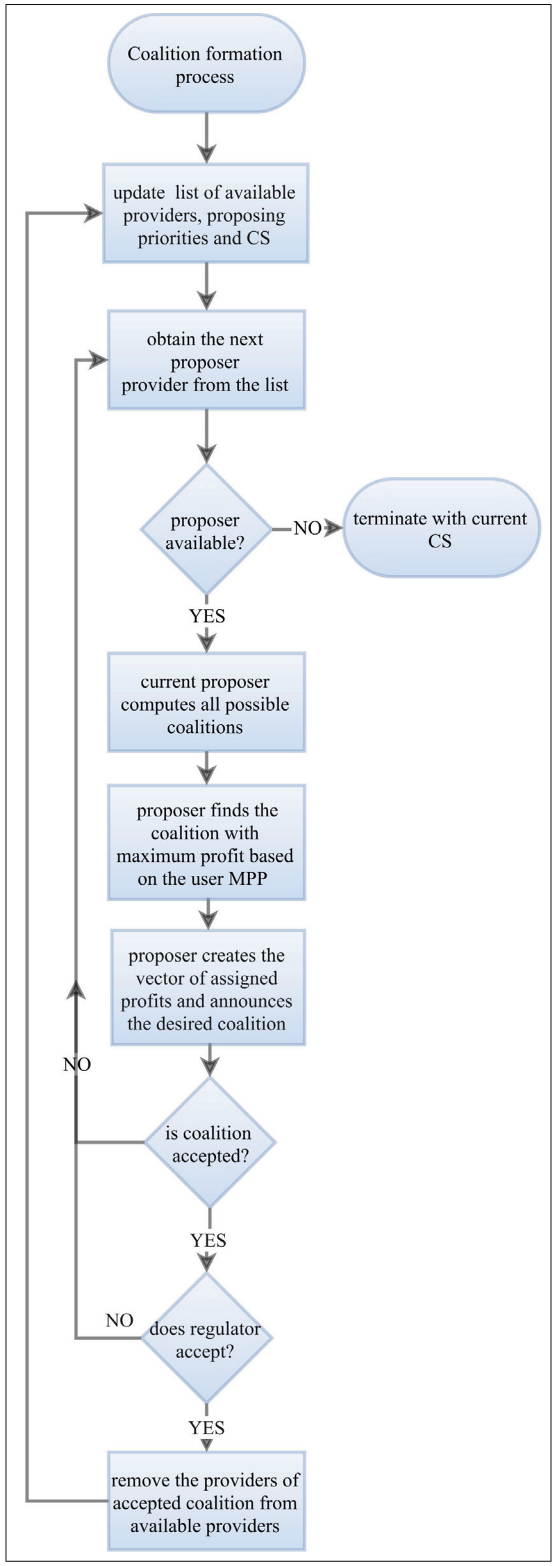

Figure 2.4 The coalition process

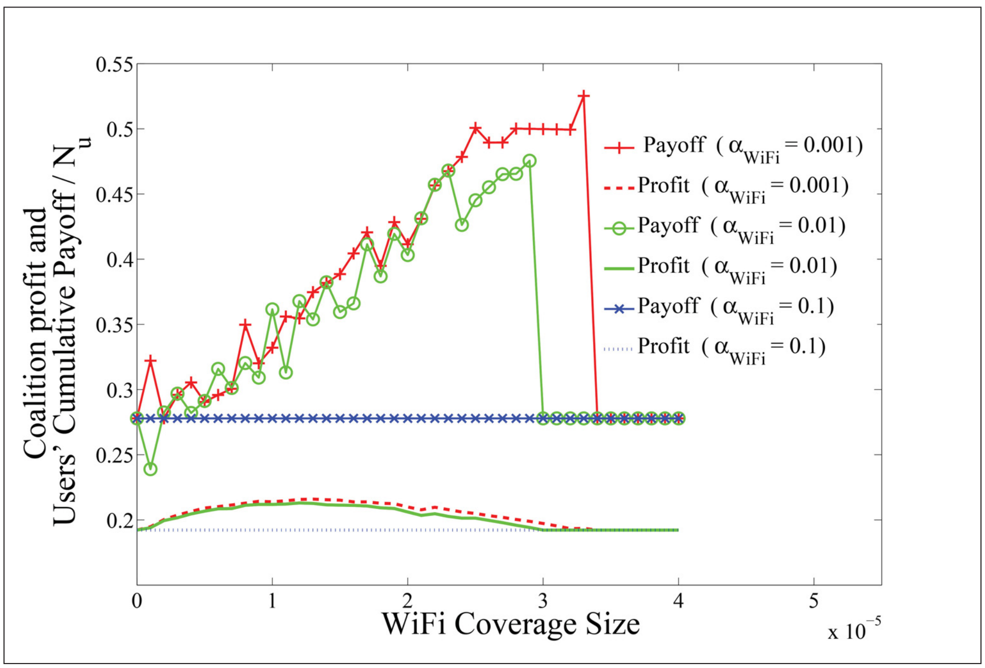

Figure 2.5 Coalition profit and cumulative payoff of all users for $\alpha_{WiFi} = [0.001, 0.01, 0.1]$. Note that the three lower curves are coalition profit and the other three are their associated user payoffs

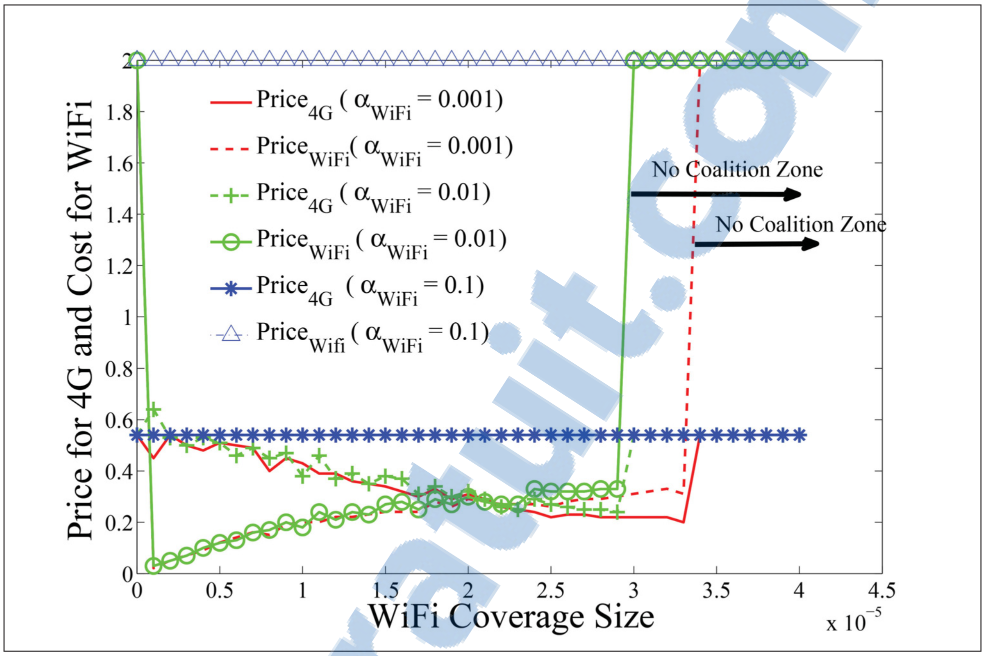

Figure 2.6 Optimum price for WiFi and 4G in the coalition for $\alpha_{WiFi} = [0.001, 0.01, 0.1]$. As WiFi coverage size increases, the WiFi price goes higher and the 4G price decreases. At the no-coalition zone, the coalition takes apart and prices stand on their default values

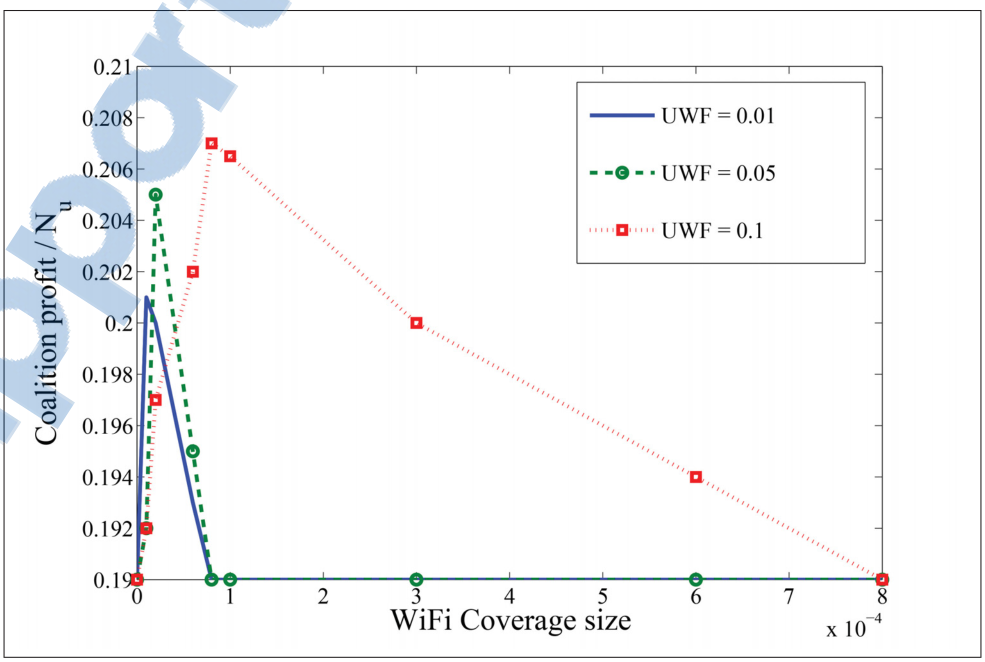

Figure 2.7 Profit of 4G-WiFi coalition when the WiFi coverage size varies. Three levels of UWF (θ) are considered

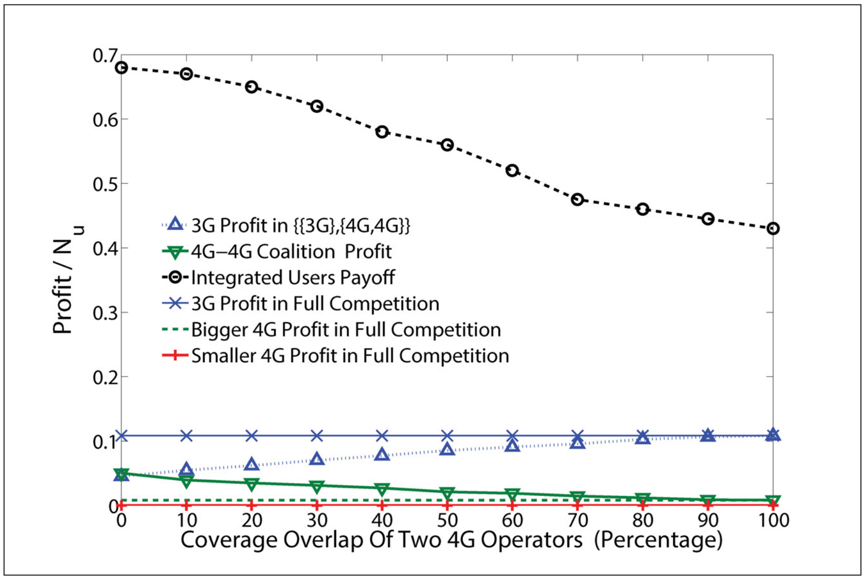

Figure 2.8 Profit of providers and integrated payoffs of all users in 3-provider structure (Scenario 3)

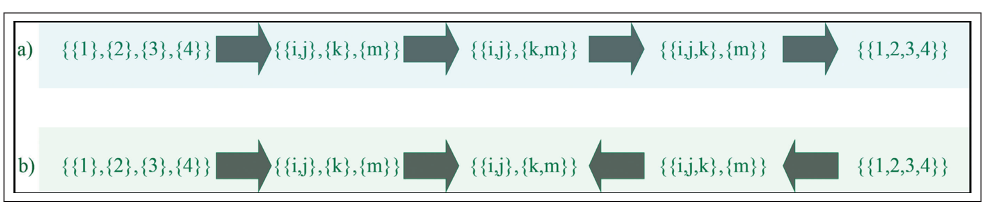

Figure 2.9 Feasible moves allowed by regulator for two weighting rules a) $w_p = 0, w_u = 1$ and b) $5\,w_p = w_u$. Each of these diagrams shows the possible and allowed moves from any status quo coalition structure

CHAPTER 3

SELECTIVE FREE CONTENT IN CELLULAR NETWORKS

Behdad Heidarpour [1], Zbigniew Dziong [1], Wing Cheong Lau [2], Shahin Vakilinia [3]

[1] Département de Génie Électrique, École de Technologie Superieure ,
1100 Notre-Dame Ouest, Montréal, Québec, Canada H3C 1K3
[2] Department of Information Engineering, The Chinese University of Hong Kong,
Sino Building, Chung Chi Rd, Sha Tin, Hong Kong
[3] Laboratory for multimedia communication in telepresence, École de Technologie
Superieure, 1100 Notre-Dame Ouest, Montréal, Québec, Canada H3C 1K3

Submitted to IEEE Transactions on Mobile Computing in April 2017.

Abstract

Over the past several years, sponsored mobile data and the payment directions on the Internet have been two major subjects in network economics. Several tier-1 service providers (SP) such as AT&T and Verizon created their frameworks for sponsored mobile data by cooperating with content providers (CP). Based on these frameworks, users can have free data transfer if they accomplish a predefined task such as buying an offered product, watching advertised videos or completing a survey. In this paper, we investigate particular types of data content which could be delivered to all cellular users free, even those without a data plan. Our approach does not force users to click on advertised content to obtain free data access. These applications if offered free of charge, can naturally generate a level of profit for the CP that make it able to compensate the profit loss of SP by using side-payments. We call this approach a selective free content (SFC) program that defines the specific types of contents eligible for such treatment. We consider a multi-stage game consisting of cellular users, SP, and CP. We solve this game by backward induction. In this way, we define the thresholds of price and data usage and maximum preferred usage that makes an application suitable for an SFC program. Finally, we verify our method by several numerical examples.

Introduction

Since the establishment of first telephone companies, network connectivity has been sold as a product for over a century. The circuit-switched networks have no knowledge about information or its worth. Hence, the dominant type of pricing for these networks is based on the duration of each connection. With the introduction of packet-switched networks in the 60s, providers were able to resolve the second type of network products which is the size of transferable data. This service is defined mainly by data amount, transfer rate, and quality. Still, with pricing based on this definition, there is no resolution among different types of data and their value to the end-users during the billing process. Providers do not set the price based on the content itself but its volume only.

To customize and to improve the current pricing policies, there are several issues to be addressed. Firstly, putting information-awareness aside, today's access networks are not content-aware in the first place. Secondly, in each wireless market, the worth of different content types to the end users are not statistically and economically defined. Concerning the first issue, the content-aware networks (CAN) have been the subject of many recent studies e.g., (Yin *et al.*, 2013) and (Subbiah & Uzmi, 2001). It is expected to have a wide implementation of such networks in near future Internet. The vast implementation of content delivery networks (CDN) (Spagna *et al.*, 2013) and edge-computing is the preliminary step toward the future CAN. Hence, by recent advancement in this field, the economic aspects of CAN and the appropriate process for revenue making and service billing should be studied. This new opportunity that comes with a better understanding of the value of each data flow motivates us to study new types of pricing schemes that have an eye on fairness and user satisfaction. In particular, we focus on providing mechanisms in which specific types of data transfer have no cost to the end users. In this methods, SPs and CPs cooperate to leverage the natural behavior of users such as on-line shopping to generate profit. Also, providers do not force users to click on advertised contents to obtain free data transfer. Apparently, such mechanisms do not work for all content types, and only special kinds of contents can be treated in such way. For example, video contents with high demand cannot be offered free for all users without experiencing profit loss. The applications such as mapping services with embedded advertising capability for local businesses are the best candidates for our method. Hence, we call our method a selective free content (SFC) program.

To have a deeper understanding about an SFC program, we first explain several industrial and academic endeavors toward partially free data access. One of the first introduced mechanisms is sponsored data by (AT&T, 2016). In this method, users can have free data access beside their regular data plan for sponsored content. One example is sponsored videos that are provided by AT&T approved CPs. If users watch such videos, there is no impact on the usage of their regular data plan. T-Mobile and Verizon also introduced Binge On (T-Mobile, 2016) and FreeBee (Verizon, 2016) respectively. They both follow a similar philosophy to AT&T's plan with several differences in detail. The key to all of these plans is the presence of CP's who are eager to sponsor free data transfer. Due to this reason, the offered free content is restricted to specific CPs and moreover, to the selected content that CP sponsors. Also, these plans are offered to the users who already have a data plan which is a major drawback regarding fairness and social welfare. Another concern about sponsored data program is the violation of network neutrality. Since the major SPs can attract powerful CPs by charging them for their access to the end users, the smaller CP's and SP's cannot compete in this field; that is in contradiction to widely accepted practice which suggests an equal and neutral policy on all data over Internet. For further explanation, we consider the concepts of one- and two-sided payments which are proposed in (Musacchio *et al.*, 2011). As it is depicted in Fig. 3.1-a, in a neutral network, the payment for data transfer is from the user side. However, in a non-neutral network, CP should pay for the contents being transferred to the users as well (Fig. 3.1-b). It is clear that weaker CP that cannot pay such fees to SP are vulnerable in non-neutral networks.

Challenges mentioned above motivate us to study alternative pricing methods for specific high-value/low-usage contents that shift the burden of all the data transfer costs from the end users to the related CPs. Since we focus on the type of content and not the CPs itself, any application that meets the eligibility condition can be offered in an SFC program. Hence, it does not affect the CPs that are competing and provide similar content types. This approach mitigates the adverse effects of an entirely non-neutral network. One example of eligible applications is mapping services that have small data usages and generate their profit from local businesses. These businesses can be hotels, shopping centers and any market relying on the on-line advertisement. The second example of eligible applications is real-time IoT services like health monitoring wearable devices connected to cloud-based applications. These applications usually use small amounts of data transfer, yet carry highly valuable information which is processed and billed by third-party cloud-based services. We show the major difference between different types of eligible contents. In all of them, the payment direction of SFC program is similar to Fig. 3.2, in which user do not pay for the data transfer associated with the eligible content.

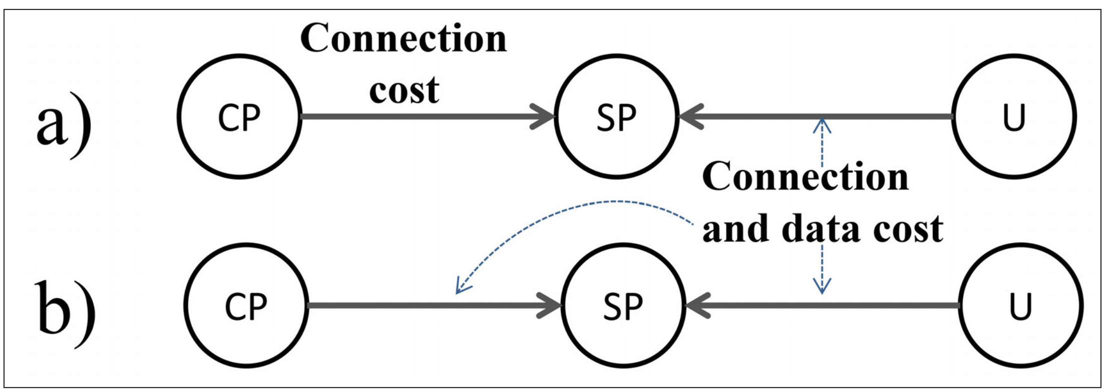

Figure 3.1 The payment directions for a) neutral networks b) non-neutral networks. *U* represents users

There are several academic endeavors to study the economics of sponsored data, e.g., (Joe-Wong *et al.*, 2015) and (Lotfi *et al.*, 2015) that analyze the optimal values of sponsored data for CPs, (ElDelgawy & La, 2015) which analyses the interaction between CP and SP for improving the delivered QoS, (Andrews *et al.*, 2014) that considers selecting the best CP locations for offering sponsored data promotions, and (Andrews *et al.*, 2016) which studies the optimal profit of SP in a sponsored content program. These works consider the same philosophy introduced by service providers, such as AT&T, for sponsored data and try to optimize several network and market parameters. They do not discriminate different types of contents based on their importance and user traffic pattern. Also, no work considers an entirely free data access for vital cellular applications like mapping services and this is one of the issues that we aim to address in this paper.

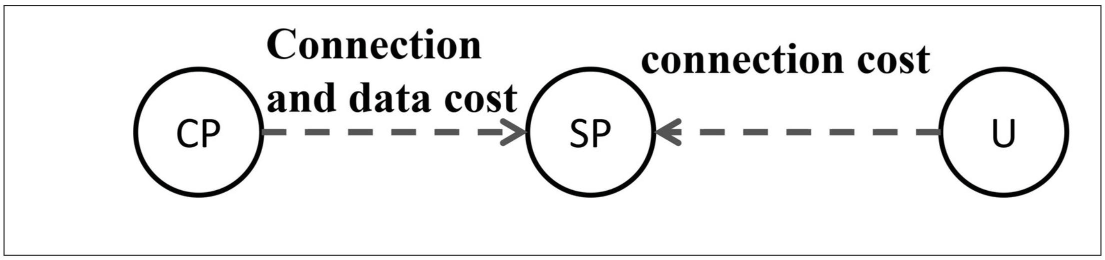

Figure 3.2 The payment directions for SFC program

The rest of this paper is organized as follows. In Section 3.1 we introduce three categories of applications that are candidates for the SFC program. These applications are mapping services, IoT related services and finally, smart city and e-governance applications. We demonstrate the eligibility of such applications by analyzing several statistical reports from Ericsson (Ericsson, 2016) and comScore (ComScore, 2016). In this part, the current traffic trends and behavior of cellular users are also provided. In Section 3.2, the sequential game for the first category of applications is developed and analyzed (we selected for analyzing the first category due to its higher complexity). Section 3.3 includes numerical examples. Finally Section 3.4 concludes the paper.

3.1 On the possibility of selective free access

In this section we focus on those types of applications that can be offered free of charge to the end users. For such applications, SP can allow a free of charge data transfer without being concerned about its profit loss or the extra load of traffic that would be generated. This comes from the fact that while, under an SFC program, SP loses a part of its profit from the end users, it can be compensated by CP. To find content types that work well with the SFC program we analyze characteristics of types of content to select the ones that can add value under the SFC program. First we define three categories of the eligible applications. Then, the characteristics of such applications are extracted from Internet statistics.

3.1.1 Category 1: mapping and other business related applications

First we consider the low data usage applications that are highly valuable to the end users. To be more precise, consider Fig. 3.3 that shows the most reached mobile applications in U.S. The data is acquired from comScore's 2016 report (ComScore, 2016). On top of the list is Facebook having 80% of the audience. The Facebook application is known for its moderate to high data usage. In fact, it can consume up to three Megabytes per minute even if the user does not play any video in it. The second rank is Facebook Messenger which is less traffic greedy. However, its overall consumption can be very high since it can be used repeatedly during a short time period as a messaging service. The third most reached application is YouTube which generates the most traffic when compared with other services in the list. In fact, based on the YouTube's statistics, the average viewing session for mobiles is 40 minutes as of 2016 (Youtube, 2016). This means for 480P videos, having 2.5 Mbps data rate, YouTube consumes 750 MB per average session. For 1080P, the required bandwidth is 2.4 GB. With the current trend in video sharing and the new capability of $4K$ video recording on smartphones, one can expect an exponential growth of traffic in this section in the next coming years.

The fourth most reached application is Google Maps with around 55% of reachability in the U.S market. From SFC viewpoint, Google Map has three interesting features comparing to the top three services. First, it's not a social media application or entertainment service. Hence, every time a user opens this application, it is due to the importance of information that is required. Second, while the first three applications in the list have moderate to high traffic demand, the amount of required data transfer for Google Maps is negligible per request; as of today, based on our measurements, it uses 300-500 KB to process each location request. The final aspect is the new feature of Promoted Pins that lets local businesses to offer different kinds of promotions to their customers. The advertisements appear as pins on the map when a user searches for a related location. For example, when a user requests for nearby restaurants, the special offers would appear. Google Maps also supports the bidding mechanisms for hotels. In all of these cases, Google highly relies on its reachability to the users which is directly related to the quantity of data subscribers in local cellular networks. However, as the data acquired from

Table 3.1 Subscriber and traffic shares in advance mobile markets
Taken from Ericsson mobility report

	$<$ 100 MB	100 MB – 1 GB	1 – 10 GB	10 – 100 GB	$>$ 100 GB
Subscriber share	$\approx 35\%$	28.8%	32%	$\approx 3.5\%$	$\approx 0.7\%$
Traffic share	$\approx 0.7\%$	$\approx 11.5\%$	$\approx 48.2\%$	$\approx 35.2\%$	$\approx 3.5\%$

Ericsson Mobility report (Ericsson, 2016) in Table 3.1 shows, over 35% of wireless users in advance markets have a data cap of less than 100 MB. The total traffic generated by this group is 0.7% of total traffic. The traffic share for the group of 100 MB- 1 GB plans is about 11.5% while this group includes 29% of all subscribers. Thus, while Google requires high connectivity of users for its business model, near 64% of subscribers do not have the necessary data connection to use Google Maps freely. The features mentioned above indicate that the mapping applications such as Google Maps have the potential to be offered under SFC program. The traditional payment directions for Google Maps are depicted in Fig. 3.4. Where here the end users pay for their data connectivity, local businesses pay Google for advertisement, and finally the end users may pay local businesses for their offers on the mapping application. Under the SFC program, the payment direction would be defined as in Fig. 3.5. In this case, the end users do not pay for their usage of Google Maps. Instead, the content-aware cellular network allows them to use this application free of charge. To compensate for the SP's lost revenue from the end users, Google would share part of its extra profit with SP. The extra profit comes from the increased advertisement clicks which is due to the higher service access by SP's users. Note that this alternative scenario is feasible due to unique characteristics of Google Maps. One could argue that the free access may overload the cellular network. However, the Ericsson's data in Table. 3.2 shows that users are not becoming greedy for certain kinds of applications when the available data volume is increasing. To be more precise, their volume share is diminishing unlike, for example, the share of video services that is growing with the amount of accessible data. The greediness and other relevant characteristics of potential free services are defined in more details in Subsection 3.1.4.

3.1.1.1 Other candidates in this category

While Google Maps is one of the best candidates for the SFC program in this category, there are several other potential services as well. One is Apple Maps which is #12 in comScore's list (ComScore, 2016). Aside from mapping applications, two popular intelligent personal assistants, Siri by Apple and Microsoft's Cortana are other candidates. They can distribute offers from local businesses and add value to the cellular operator's network without putting the burden on end users; similar to mapping services, the information delivered by these services has a high value to the users as well.

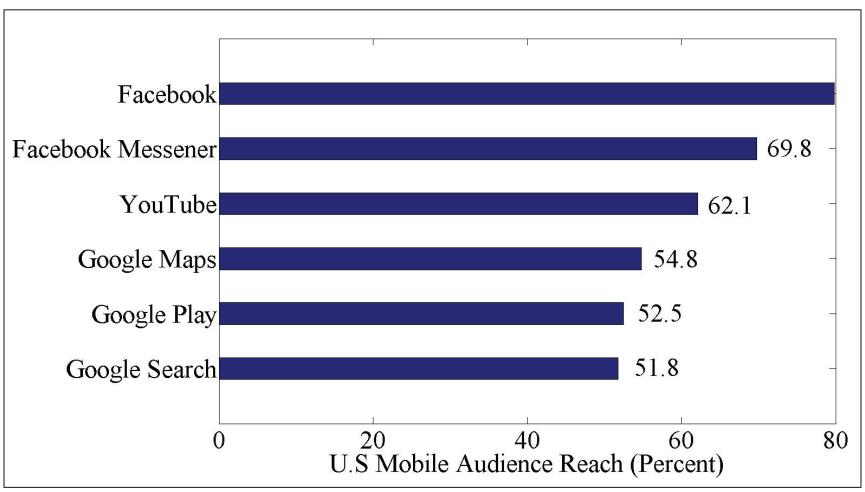

Figure 3.3 U.S audience reach for mobile applications based on comScore's statistics
Taken from comScore (2016)

3.1.2 Category 2: real-time cloud-based IoT services

The second category of applications eligible for SFC program is related to rapid development of wearable devices, IoT services, and edge-computing. In contrast to the first category in which

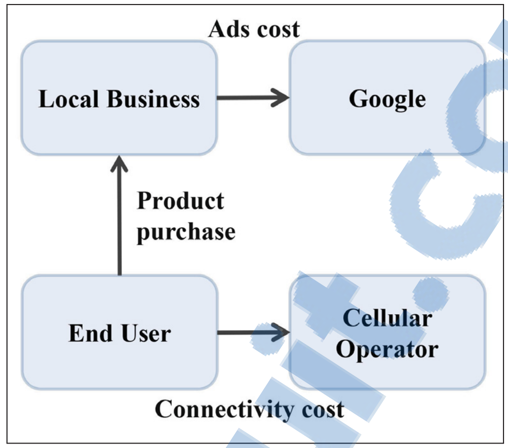

Figure 3.4 Traditional payment direction for Google Maps

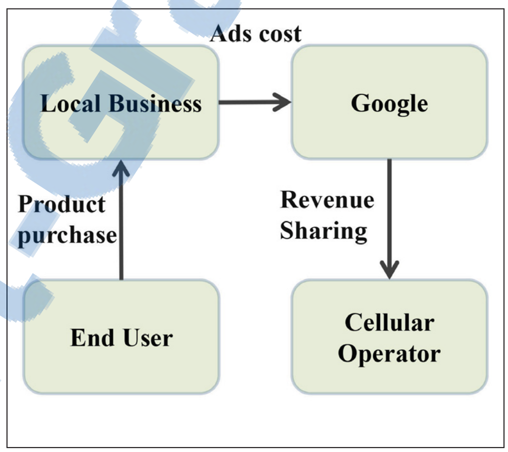

Figure 3.5 Alternative payment direction model for Google Maps based on free content-aware connectivity for users

the end users would not directly pay for the services of Google or Apple, in the second category users pay for the cloud-based services. In the current market model, end user pays for both data connectivity and cloud-based services that collect event-triggers from sensors and react. However, there are some scenarios in which the current market model can be inefficient or even dangerous. For example, consider the health monitoring system which loses its connection to cloud-based service since the data plan reached its cap. In such scenario, while the user already

Table 3.2 Application volume shares of different subscriber groups adopted. Taken from Ericsson Mobility report

Application	< 0.1 GB	$0.1 - 1$ GB	$1 - 10$ GB	$10 - 100$ GB	> 100 GB	All users
Video	4%	16%	39.5%	67%	67.7%	46.7%
Social Networking	13.7%	17.7%	17.7%	6.5%	1.2%	13.7%
Web Browsing	20%	18.5%	12%	5.6%	2.4%	10.4%
Comm. Services	12%	8.8%	4%	2.4%	0.8%	3.2%
Software Download	16.1%	15.3%	6.4%	2.4%	1.6%	6.4%
Audio	0.8%	3.2%	3.2%	1.6%	0.8%	3.2%
System	0.2%	1.6%	$\approx 0\%$	$\approx 0\%$	$\approx 0\%$	$\approx 0\%$
File Sharing	$\approx 0\%$	$\approx 0\%$	$\approx 0.5\%$	$\approx 1.6\%$	$\approx 16\%$	1.6%
Other	$\approx 33.2\%$	$\approx 18.9\%$	$\approx 16.7\%$	$\approx 15.3\%$	$\approx 9.5\%$	$\approx 14.8\%$

paid for a critical service, the service cannot save its life. Thus this service could benefit greatly from the SFC program as well as a broad range of IoT services based on low date usage sensors that provide valuable information.

The traditional payment model requires the end users to pay both network provider and cloud-based services located on the edge of provider's network. This model of payment directions is depicted in Fig. 3.6-a. The alternative SFC model removes the data transfer and connectivity cost from the end user. In this model, the cost of data transfer is being paid by cloud-based service owner. In many markets, similar scenarios have been implemented. For example, the majority of laptops in the today's market come with a version of MS Windows and a free subscription to MS Office for a defined period. For several years Samsung offered free 48 GB Dropbox storage with their smartphones. Wacom tablets come with a free license for Photoshop or Corel applications. The examples in this category are countless and it proves the effectiveness of such business model. Hence, the SFC program for the second category of applications can be implemented using experience from the already existing models in different markets.

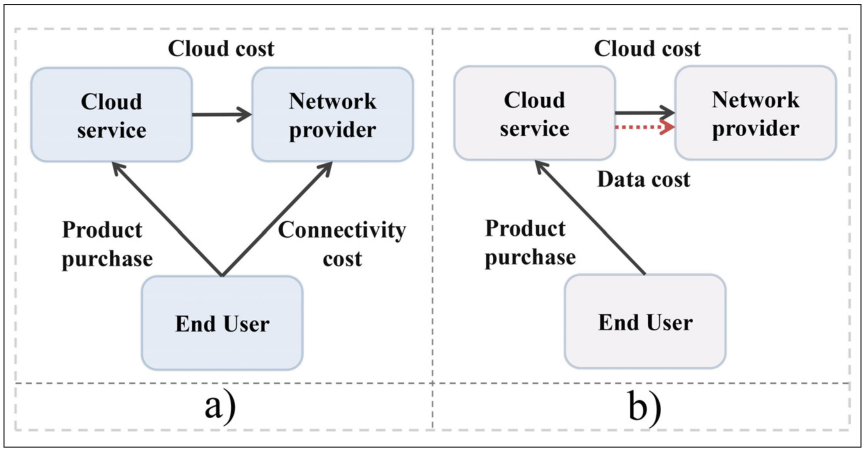

Figure 3.6 a) Current payment directions for cloud-cellular IoT services b) alternative SFC model

3.1.3 Category 3: Smart cities and the social right to access the Internet

In the previous two categories, the final goal of content providers, whether they are single purpose cloud-based companies or giant multi-role corporations like Google, is generating profit out of users. The third category, on the other hand, is dedicated to those entities whose aim is leveraging the quality of life by providing ICT services to a broad range of people. These entities would be governmental or municipal bodies that intend to implement and develop the concept of smart cities. To achieve a densely interconnected structure, such as smart cities, it is required to implement several related technologies such as IoT for transport and traffic monitoring, energy consumption metering and health care systems. This type of services can also cover electronic governance (e-governance) and on-line voting concepts in which the interaction between governments and citizens takes place via ICT services. Since having access to the Internet is playing a primary role in future smart cities, public networks are being implemented around the globe. These networks are primarily based on WiFi technology. The service access is given via paid plans or in some cases freely. The free WiFi access is mainly provided in governmental facilities or touristic sites. For example, the Old Port neighborhood of Montreal is offering free Internet access to the public; this is a part of a long-run plan for city-wide free WiFi access (Montreal, 2016). There are numerous examples of free and paid urban WiFi. We refer the reader to (Kong, 2016; York, 2016; Toronto, 2016; Liverpool, 2016) for several examples. The WiFi service has the advantage of high-speed access to the network via free license spectrum. However, implementing a city-wide network requires a vast amount of financial resources; this is due to small coverage range of WiFi access points (AP) comparing to the cellular counterparts. Deploying hundreds or thousands of APs also increases the maintenance and managing cost. As of today, the cost of such networks is being paid by municipal entities or via community support in free service methods or by private companies for paid or advertised services.

Due to the limited coverage size of current municipal networks based on WiFi, one can consider an alternative or additional option of selective free cellular access in the urban area. This option would be mainly useful for the case of free or advertised services that are funded via predefined budgets dedicated to municipal network plans. For the previous two application categories proposed for the SFC program, the primary question was how to make selective free access profitable; in the third category, the major challenge is how to develop the ICT applications to make them accessible free of charge on cellular networks. This challenge is mainly related to service type definition and traffic shaping. In particular, since the cellular data is traditionally more expensive than the land-line services, the smart city application should transfer the minimum traffic with the most valuable data. In the next subsection, we define general characteristics of all eligible applications for the SFC program.

3.1.4 Characteristics of eligible applications for the SFC program

Until now we defined three categories of applications, shown in Fig. 3.7, that are good candidates for an SFC model. In this paper, our goal is building a mathematical framework in which the estimated profit values of associated entities are resolved. Hence, to have a precise formulation, we need to find the common characteristics of mentioned application. In this section we define some general characteristics of the eligible applications while the mathematical definition of each characteristic is presented in the next section. As discussed until now, a candidate application for the SFC program has the following characteristics:

a. Let define the content unit as the result of a predefined information request such as obtaining a map location. Then, in the eligible applications, the required data transfer of a content unit is relatively small, and its perceptual value to the user is high. On the other hand, for the content types such as video, the expected size of each video is respectively high and the data does not have the same importance or time criticality. In other words, in most cases when a user requires a map location data or health-care service, the request cannot be postponed till another time. Let us represent the content size with θ and the perceptual importance to a user with random variable α. Then, the importance to size ratio is $\rho = \frac{\alpha}{\theta}$. Since the two variables are generally independent, the average ratio is $E[\rho] = \frac{E[\alpha]}{E[\theta]}$. We expect this ratio to be highest for the eligible contents among all content types in the network. This definition lacks two pieces of important information. First, there is no metric for the perceptual importance. Thus, we need to use a utility function to model the user behavior. Second, ρ does not carry any information about the user greediness for the application usage which forces us to define the second property.

b. The second characteristic of eligible applications is that the user should not be greedy for the application usage. We define the overall size of content transferred in time t by application a and user j as $\Theta^a_{j,t}(d)$. Where d is the cap of user's data plan. Then, the Application Usage Index (AUI) among all users can be defined as,

$$I^a_j(d) = \lim_{t \to \infty} \frac{1}{t} \frac{\Theta^a_{j,t}(d)}{\sum_{a=1}^{A} \Theta^a_{j,t}(d)}. \tag{3.1}$$

For the eligible applications, the global application usage index should decrease with increasing d, that is:

$$\text{Greediness condition: } \frac{1}{N_T} \sum_{j=1}^{N_T} \frac{\partial I^a_j(d)}{\partial d} \leq 0, \tag{3.2}$$

where N_T is the total number of users in the market. This condition is supported by the data provided by Ericsson in (Ericsson, 2016). Namely, based on Table 3.2, the mobile application can be categorized in three groups regarding $I_j(d)$. Fig. 3.7 depicts the general shape of the usage index for each category as a function the cap of user's data plan. Fig. 3.7-a shows the usage index shape for Type I applications for which users have the highest usage greediness; this type includes the video applications. Fig. 3.7-b illustrates the usage index shape for Type II applications. A user considers utilizing these applications if it has enough bandwidth available. However, these applications are not important enough to be used in plans with a small data cap. Audio services belong to this category. Finally, 3.7-c depicts the usage index shape for the critical applications that user requires under any data plan. A user may utilize only these applications when the data cap is limited to a small value, e.g., one- or two-gigabytes. Also, users are not greedy for these applications so condition (3.2) is satisfied in this case. Web browsing and mapping services belong to this application type. Being a Type III application is a necessity to be eligible for SFC program. However, it is not sufficient; the business model should also support the SFC program. Hence, a third characteristic should be defined to resolve necessity and sufficiency conditions for eligible applications.

c. Until now we considered the usage characteristics of eligible applications. The third characteristic is related to the market condition. For any service to be considered as SFC eligible, there should be a business or social entity that can compensate the profit loss of cellular provider. This characteristic may look trivial, but when we compare a mapping service with web browsing applications, one can notice a structural difference in the business model. Namely, for the mapping applications such as Google Map, there is an explicit financial loop from local businesses to Google to SP to users and again local businesses. On the other hand, there is no such loop for the browsing applications since the potential gainers are distributed throughout the Internet. The only exception would be injecting direct advertisement from cellular provider to the web browsing data and making a payment loop similar to Google Map's business model in 3.4.

Among the three categories of eligible application for SFC program, the first category has the most complicated structure. It includes users, SP and CP that directly affect each other's behavior. The SFC feasibility models for category 2 and 3 applications are simpler and can

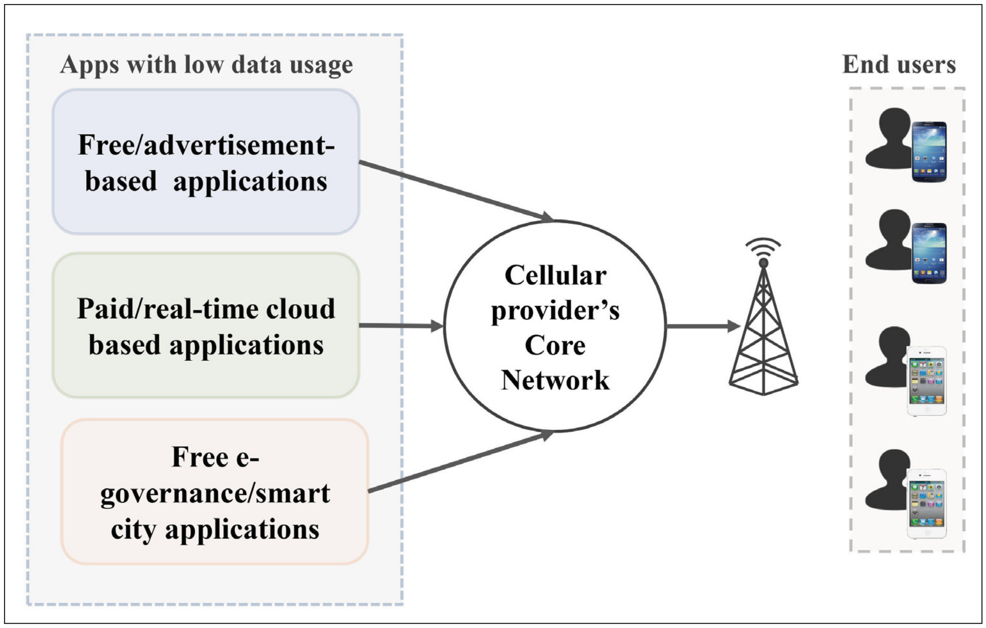

Figure 3.7 Three categories of candidate applications for selective free access

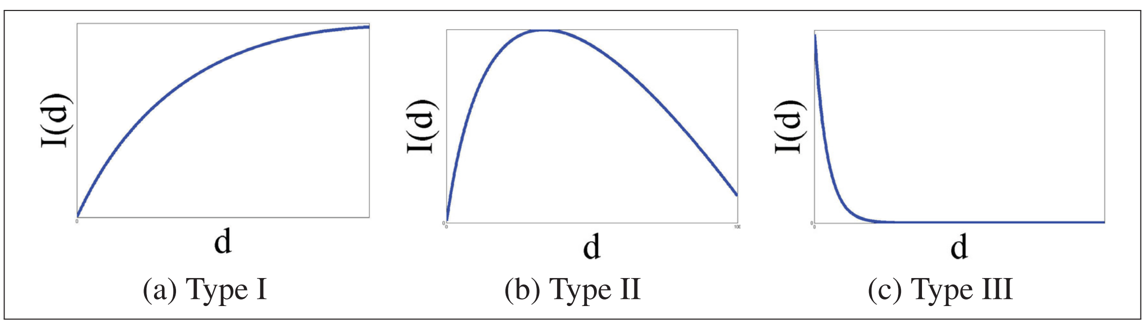

Figure 3.8 Three types of mobile applications based on subscriber's usage behavior defined by $I_u(d)$

be derived by some modification of the first category model. Hence, in the reminder of the paper we propose and analyze a detailed analytical model for SFC feasibility of the Category 1 applications.

3.2 The Game for Category 1 applications

In this section, we consider a three-stage game which defines the best strategies of SP and CP to whether join to or refuse an SFC program for an eligible Category 1 application. The game consists of three entities; namely, cellular users, SP and CP. Users adjust their subscription and data usage behavior based on the offered price from SP. CP generates its profit based on the number of subscribed users and the amount of content requests they generate. Similar to any market, since the volume of content requests is in close relation to the unit data price offered by SP, CP earns an extra profit if SP applies the minimum possible price. Our goal is to investigate and define the conditions in which an entirely free access gives the sufficient amount of extra revenue to CP to compensate the profit loss of SP. We model this scenario as a Stackelberg game and solve it by backward induction. The three stages of our game are defined as follows:

a. Each user observes the offered data unit price from SP and decides the amount of data consumption for each application. In this stage, the amount of content requests for the target Category 1 application is resolved.

b. In the next stage, SP calculates its profit for two scenarios. The first scenario is a standard pricing strategy and the second scenario considers the presence of an SFC program for specific content. In this part, an SP-CP cooperation means SP provide SFC to all users and demands from CP the compensation from its profit loss. If CP cannot compensate the profit loss of SP, then SP does not participate in the SFC program.

c. In the last stage, CP should decide whether to take part in the SFC program or not and if it participates, what share of profit should be proposed and transferred to SP.

In what follows, we demonstrate the details of each stage of the game.

3.2.1 Stage III: user's utility and best response

Since the user behavior analysis is the foundation of this framework, we need to define the proper metric to find the amount of data consumption. Similar to many related works, we use the concepts from utility theory to formulate this part. With the help of the data provided by Ericsson (Ericsson, 2016) in Tables 3.1 and 3.2, we know that 70% of current users have a limited data plan with less than 2 GB; the primary concern of these users is choosing between high priority services like mapping applications and other less critical application such as video. For these users, the utilization percentage of high demanding applications such as video is negligible. For the rest of users who share almost 80% of overall traffic, the decision concern is mainly about the amount of traffic they need to buy for their video streaming applications. For these users, the traffic ratio of high-value applications to the rest of services is under 10% (Type-III applications in Fig. 3.7). Hence, we can define a two-part utility function which considers the importance of eligible application in one part and the high demand services in the other part. For each part, we use the familiar form of logarithmic utility function due to its conformity to the *law of diminishing marginal utility* (Hall & Lieberman, 2012). The adaptation of this law is essential in studying the cases of data consumption. Also, the logarithmic utility is a common practice in related works e.g. (Başar & Srikant, 2002), (Sengupta *et al.*, 2007), (Duan *et al.*, 2013a) and (Lotfi *et al.*, 2015). The utility for a specific user j has the form of:

$$u^j(p) = \frac{\alpha_e^j \beta_e log(1+d_e^j) - pd_e^j}{U_e} + \frac{\alpha_r^j \beta_r log(1+d_r^j) - pd_r^j}{U_r}. \tag{3.3}$$

The first part of the above function defines the gained normalized payoff from using an SFC eligible application. This application is indicated by subscript e. The second part belongs to the rest of applications with lower importance and higher traffic demand indicated by index r. α_i^j $i \in \{e,r\}$ is a random variable which shows the importance of the application i to user j. This importance is coupled with the amount of money that user is willing to pay for a specific type of content. For the sake of simplicity in analysis, we assume that α_e^j and α_r^j are i.i.d having a uniform PDF of $U(0,1)$. β_i is a user-independent variable which controls the amount of data consumption for a given price. d_i^j is the amount of preferred data usage for each content type. We also define constant D_i which indicates the maximum amount of data consumption users tend to achieve. Based on the definition, we expect D_e to be negligible comparing to D_r. $p \in \Re^+$ is the unit price for data implied by SP. Finally, U_e and U_r are the normalizing factors which control the peak of utility for each content. These two constants are essential since the two parts of utility have different peaks, yet they may represent the same amount of satisfaction to each user. By this definition, $U_i = \beta_i log(1+D_i)$ and the maximum of $u^j(p)$ for the most demanding user can be 2. For the rest of users the maximum utility is $u^j_M(p) = \alpha^j_e + \alpha^j_r < 2$ which shows that the maximum value of satisfaction is related to the perceptual importance of the services to the user. it is clear that $u^j(p)$ is concave with respect to d^j_e and d^j_r. The first derivative of $u^j(p)$ with respect to d^j_i is:

$$\frac{\partial u^j(p)}{\partial d^j_i} = \frac{1}{U_i}\left(\frac{\alpha^j_i \beta_i}{1+d^j_i} - p\right), \tag{3.4}$$

$$d^{j\prime}_i = \frac{\alpha^j_i \beta_i}{p} - 1, \tag{3.5}$$

where $d^{j\prime}_i$ is the global maximum of $u^j(p)$. By considering the positivity and the maximum level of usage, we have the optimum value as:

$$d^{j^o}_i = \min\left(\max\left(\frac{\alpha^j_i \beta_i}{p} - 1, 0\right), D_i\right). \tag{3.6}$$

The above equation indicates that $p \geq \alpha^j_i \beta_i$ leads to zero usage for the content of type i, and $p \leq \frac{\alpha^j_i \beta_i}{1+D_i}$ gives the user the opportunity to reach the maximum demand for the content of type i. To have the analysis of user's best responses, we need to categorize the users based on usage threshold orders. These orders can be represented by two main sets:

$$\text{Order set I-}\begin{cases} 1)\alpha^j_e \beta_e \geq \alpha^j_r \beta_r > \frac{\alpha^j_e \beta_e}{1+D_e} > \frac{\alpha^j_r \beta_r}{1+D_r}, \\ 2)\alpha^j_e \beta_e > \frac{\alpha^j_e \beta_e}{1+D_e} > \alpha^j_r \beta_r > \frac{\alpha^j_r \beta_r}{1+D_r}. \end{cases} \tag{3.7}$$

$$\text{Order set II-}\begin{cases} 3)\alpha^j_r \beta_r > \alpha^j_e \beta_e > \frac{\alpha^j_e \beta_e}{1+D_e} > \frac{\alpha^j_r \beta_r}{1+D_r}, \\ 4)\alpha^j_r \beta_r > \alpha^j_e \beta_e > \frac{\alpha^j_r \beta_r}{1+D_r} > \frac{\alpha^j_e \beta_e}{1+D_e}, \\ 5)\alpha^j_r \beta_r > \frac{\alpha^j_r \beta_r}{1+D_r} > \alpha^j_e \beta_e > \frac{\alpha^j_e \beta_e}{1+D_e}. \end{cases} \tag{3.8}$$

The main difference between the two sets is the user's content prioritizing behavior. The first set represents the users who prioritize the Type e contents and the second set is for those who favor the Type r applications. To have a better understanding of the user behavior, let us define the best response function as follows:

Proposition 4. The best response data values for the users in the first order (set I-1) are as follows:

$$\text{BR/I-1} \begin{cases} \boldsymbol{d_e^{jo}} = 0,\ \boldsymbol{d_r^{jo}} = 0 & p > \alpha_e^j \beta_e, \\ \boldsymbol{d_e^{jo}} = \frac{\alpha_e^j \beta_e}{p} - 1,\ \boldsymbol{d_r^{jo}} = 0 & \alpha_r^j \beta_r < p \leq \alpha_e^j \beta_e, \\ \boldsymbol{d_e^{jo}} = \frac{\alpha_e^j \beta_e}{p} - 1,\ \boldsymbol{d_r^{jo}} = \frac{\alpha_r^j \beta_r}{p} - 1 & \frac{\alpha_e^j \beta_e}{1+D_e} < p \leq \alpha_r^j \beta_r, \\ \boldsymbol{d_e^{jo}} = D_e,\ \boldsymbol{d_r^{jo}} = \frac{\alpha_r^j \beta_r}{p} - 1 & \frac{\alpha_r^j \beta_r}{1+D_r} < p \leq \frac{\alpha_e^j \beta_e}{1+D_e}, \\ \boldsymbol{d_e^{jo}} = D_e,\ \boldsymbol{d_r^{jo}} = D_r & p \leq \frac{\alpha_r^j \beta_r}{1+D_r}. \end{cases} \tag{3.9}$$

Proof. The thresholds come directly from (3.7) and the optimum values follow (3.6). □

Table 3.3 General notation

Parameter	Description
$\mathbb{N}_T$	Set of all users of SP
N_T	Size of $\mathbb{N}_T$
$I(d)$	Application Usage Index (AUI)
$u^j(p)$	utility of user j when the data unit price is p
α_e^j, α_r^j	Random variables indicating the perceptual importance of service e and r respectively for user j
α_e, α_r	The general form of above.
β_e, β_r	Traffic shaping factors
D_e, D_r	maximum preferred usage for content types e and r
d_e^j, d_r^j	The overall usage of content Type e and r for user j
p	price of data service
p^o	optimal price strategy of SP
p^{CP}	Side-payment unit price, from CP to SP
π^{CP}, π^{SP}	profit of CP and SP respectively
η	profit factor of CP for content utilization
ζ	Bargaining power of SP over CP

The best response for the rest of the threshold orders can be easily defined based on the above definition. We omit their presentation to simplify the presentation. Instead, we show the typical curves of best responses for the threshold orders in Fig. 3.9. As depicted in Fig. 3.9(a)-(e), the main difference between the best response curves is the usage behavior when the price is high. Sub-figures 3.9-(a) and (b) represent the users who prioritize the eligible contents over the rest of applications. Hence, when the price is high, they use only the eligible application. This makes a significant difference in the AUI curve. The single-user AUI of the eligible application, $I_e(p) = \frac{d_e(p)}{d_e(p)+d_r(p)}$, for the first two orders is similar to the one of the Type-III applications (a horizontally flipped version of the curve in Fig. 3.7, having d inversely related to p). Order II-1 shows a pattern similar to the Type-II applications for the presumably eligible applications. Orders II-2 and II-3 represent our eligible applications similar to Type-I applications. Based on the three characteristics of the eligible applications for the SFC program, we know that only Orders I-1 and I-2 are a realistic representation. This assertion does not imply that all users act based on the first two orders. However, since the marketwide AUI (Eq. (3.1)) represents the aggregated usage of an application in the entire market, when it comes to an eligible application the majority of users behave based on Orders I-1 and I-2. Hence we can propose the following proposition:

Proposition 5. For an eligible Category 1 application, $\beta_e > \beta_r$ always holds.

Proof. See Appendix III-1. □

3.2.2 Stage II: The best strategy for SP

In Stage II, after the analysis of users' best responses, SP should resolve its best strategy. As discussed earlier, SP decides whether it wants to participate in the SFC program or not and also, sets the data price that maximizes its profit. Thus, the SP's strategy is defined by triple (p, γ^{SP}, p^{CP}) where $\gamma^{SP} \in \{0,1\}$ defines the participation strategy and p^{CP} is the data unit price for the Type e content when SP participates in the SFC program, $\gamma^{SP} = 1$. p^{CP} is the base for

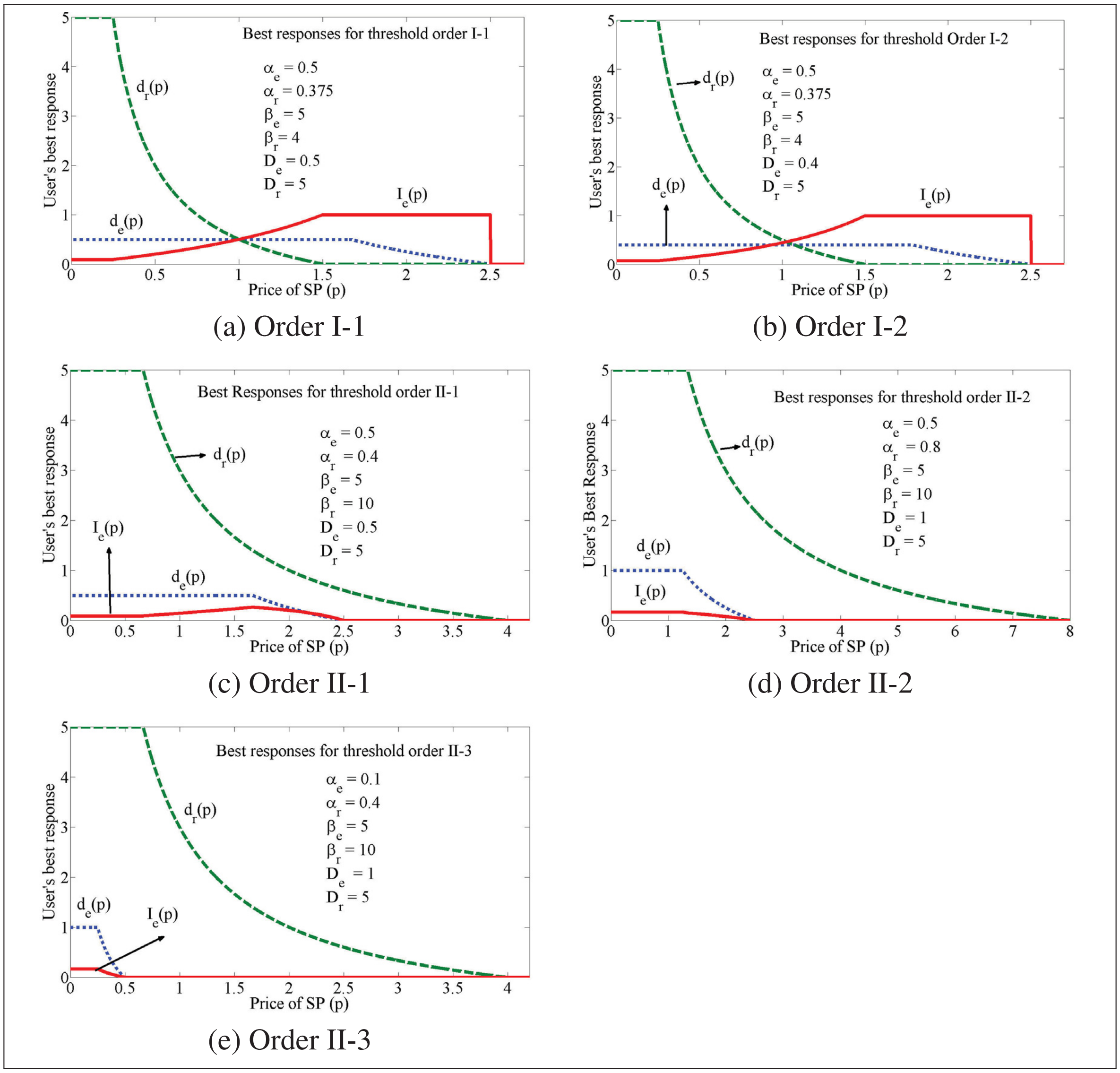

Figure 3.9 Best response of users base on threshold orders

any payment from CP to SP to compensate the SP's profit loss. In the following, we derive the optimum profit values for each strategy triple.

The profit of SP when it does not participate in SFC program is directly related to the overall data consumed by the subscribed users. When SP agrees to join the SFC program, it loses a part of its profit which comes from the eligible application's traffic. However, in this scenario, CP may compensate the profit loss of SP by making a side-payment. Thus, having N_T as the total number of users in the SP's network, we can define the profit function of SP as:

$$\pi^{SP}(\gamma^{SP}=0,p) = N_T \times \left[\int_{\alpha_e=1}^{1} d_e(\alpha_e,p)d\alpha_e + \int_{\alpha_r=1}^{1} d_r(\alpha_r,p)d\alpha_r \right], \tag{3.10}$$

$$\pi^{SP}(\gamma^{SP}=1,p,p^{CP}) = N_T \times \left[\int_{\alpha_r=1}^{1} d_r(\alpha_r,p)d\alpha_r + D_e * p^{CP} \right]. \tag{3.11}$$

Eq. (3.10) represents the non-SFC profit and (3.11) is the profit of SP under the SFC program. In (3.11) the side-payment from CP to SP it is defined as $N_T D_e p^{CP}$ that implies that under the SFC program, in which users are not charged for transferring Type e contents, all users reach maximum usage D_e. Based on the above profit equations, we define a detailed profit structure of SP based on its pricing and participation strategies in the following two subsections.

3.2.3 The profit of SP in non-cooperative strategy ($\gamma^{SP} = 0$)

When SP is not engaged in the SFC program, its only source of profit are the direct payments from the users for their data usage. In this case, SP should set the price value that maximizes its profit. Based on (3.6), the price threshold above which user j does not demand any content i is $p = \alpha_i^j \beta_i$. Hence, if SP sets the price $p > \beta_i$, no user would demand content type i. We have two thresholds $p = \beta_e$ and $p = \beta_r$ representing the upper limit of price for each content type. Also, for the same user j, $p <= \frac{\alpha_j^i \beta_i}{1+D_i}$ leads to maximum data usage. We can take the thresholds $p = \frac{\beta_i}{1+D_i}$ and $p = \frac{\beta_r}{1+D_r}$ as the price values for which greediest users start to enjoy full data usage for the respective content. Based on above definitions and Proposition 5, there are two orders of thresholds:

$$\text{SP's price threshold orders:} \begin{cases} 1)\ \beta_e > \beta_r \geq \frac{\beta_e}{1+De} > \frac{\beta_r}{1+Dr}, \\ 2)\ \beta_e > \frac{\beta_e}{1+De} \geq \beta_r > \frac{\beta_r}{1+Dr}. \end{cases} \tag{3.12}$$

The above definitions can also be derived from order set I in (3.7). Since the above orders are related to the nature of applications and general user behavior, we select the first order as the base for the further analysis. The same framework can be applied to the wireless markets with the second order. We define the SP's best response price and the associated profit under each threshold regime as follows:

3.2.3.1 Ultra-high price regime: $\beta_r < p < \beta_e$

When SP applies an ultra-high price regime, no user reach its maximum usage regarding content Type e. However, as it is depicted in Fig. 3.10, all the users with $\alpha_e \geq \frac{p}{\beta_e}$ can enjoy a partial usage of $d_e = \frac{\alpha_e \beta_e}{p} - 1$. Considering the Type r applications, since p is above the minimal usage threshold, no user will utilize these applications and hence $d_r = 0$ for all the users. The overall profit of SP is:

$$\pi_{uh}^{SP}(\gamma^{SP} = 0, p) = N_T p \int_{\alpha_e = \frac{p}{\beta_e}}^{1} \frac{\alpha_e \beta_e}{p} - 1 \, d\alpha_e = N_T \left(\frac{p^2}{2\beta_e} - p + \frac{\beta_e}{2} \right).$$

The first derivative of above profit function is $N_T(\frac{p}{\beta_e} - 1)$ and the second derivative is $\frac{N_T}{\beta_e}$. Hence, the profit function in ultra-high price regime is convex and its maximum occurs at the boundary price $p = \beta_r$:

$$\max_p \pi_{uh}^{SP}(\gamma^{SP} = 0, p) = N_T \left(\frac{\beta_r^2}{2\beta_e} - \beta_r + \frac{\beta_e}{2} \right). \tag{3.13}$$

3.2.3.2 High price regime: $\frac{\beta_e}{1+D_e} < p < \beta_r$

Considering the user's best responses, the difference between the ultra-high and high price regimes is that in the latter, a part of users with $\alpha_r \geq \frac{p}{\beta_r}$ utilize the r type applications. The best

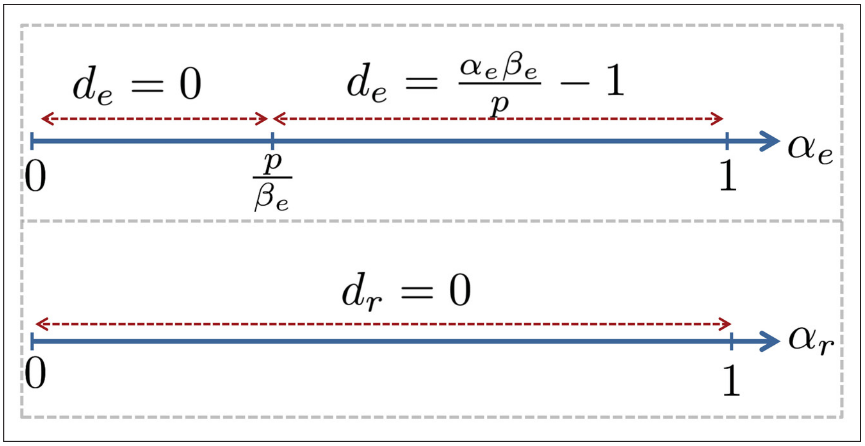

Figure 3.10 Optimal content usage with respect to α_e and α_r in ultra-high price regime $\beta_r < p < \beta_e$

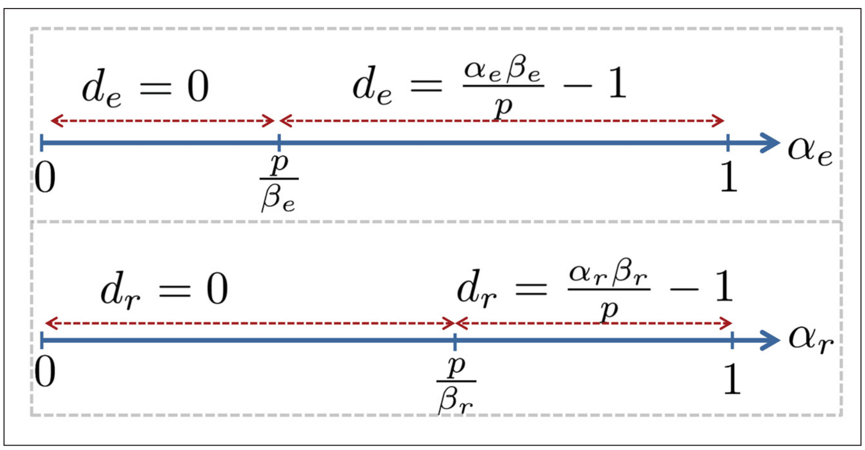

Figure 3.11 Optimal content usage with respect to α_e and α_r in high price regime $\frac{\beta_e}{1+D_e} < p < \beta_r$

response for e type application remains the same. This behavior is depicted in Fig. 3.11.

$$\pi_h^{SP}(\gamma^{SP} = 0, p) = N_T p \left[\int_{\alpha_e = \frac{p}{\beta_e}}^{1} \frac{\alpha_e \beta_e}{p} - 1 \, d\alpha_e + \int_{\alpha_r = \frac{p}{\beta_r}}^{1} \frac{\alpha_r \beta_r}{p} - 1 \, d\alpha_r \right]$$
$$= N_T \left(\frac{p^2(\beta_e + \beta_r)}{2\beta_e \beta_r} - 2p + \frac{\beta_e + \beta_r}{2} \right). \tag{3.14}$$

The profit function of (3.14) is convex and similar to (3.13) the maximum value is in lower boundary of price $p = \frac{\beta_e}{1+D_e}$:

$$\max_{p} \pi_h^{SP}(\gamma^{SP} = 0, p) = N_T \left(\frac{\beta_e + \beta_r}{2} \left(\frac{\beta_e}{\beta_r (1+D_e)^2} + 1 \right) - 2 \frac{\beta_e}{1+D_e} \right). \tag{3.15}$$

3.2.3.3 Moderate price regime: $\frac{\beta_r}{1+D_r} < p \leq \frac{\beta_e}{1+D_e}$

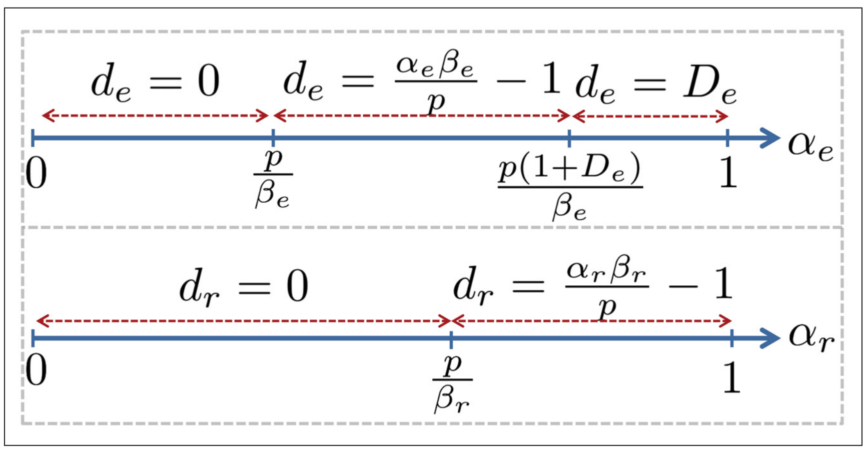

Figure 3.12 Optimal content usage with respect to α_e and α_r in moderate price regime $\frac{\beta_r}{1+D_r} < p \leq \frac{\beta_e}{1+D_e}$

In the moderate price regime, SP allows a part of users with $\alpha_e \geq \frac{p(1+D_e)}{\beta_e}$ reach their maximal usage for content e. However, with such price regime no user is willing to achieve a maximum usage for content r. These conditions are shown in Fig. 3.12. The profit of SP in moderate price regime is defined as shown in Eq. (3.16).

Proposition 6. The profit function in moderate price regime has a maximum at $p = \frac{\beta_e \beta_r (D_e - 1)}{\beta_r((1+D_e)^2 - 1) - \beta_e}$, if $D_e > 1$ and $D_r > \frac{D_e(\beta_r(D_e+2) - \beta_e)}{\beta_e(D_e - 1)}$, otherwise, the maximum occurs at lower boundary price $p = \frac{\beta_r}{1+D_r}$.

Proof. See Appendix III-2. □

$$\pi_m^{SP}(\gamma^{SP}=0,p) = N_T\, p\left[\int_{\alpha_e=\frac{p}{\beta_e}}^{\frac{p(1+D_e)}{\beta_e}} \frac{\alpha_e\beta_e}{p} - 1\, d\alpha_e + \int_{\alpha_e=\frac{p(1+D_e)}{\beta_e}}^{1} D_e\, d\alpha_e + \int_{\alpha_r=\frac{p}{\beta_r}}^{1} \frac{\alpha_r\beta_r}{p} - 1\, d\alpha_r\right]$$
$$= N_T\left(\frac{p^2}{2}\left(\frac{1}{\beta_r} - \frac{1}{\beta_e}((1+D_e)^2-1)\right) + (D_e-1)p + \frac{\beta_r}{2}\right) \tag{3.16}$$

3.2.3.4 low price regime: $p \le \frac{\beta_r}{1+D_r}$

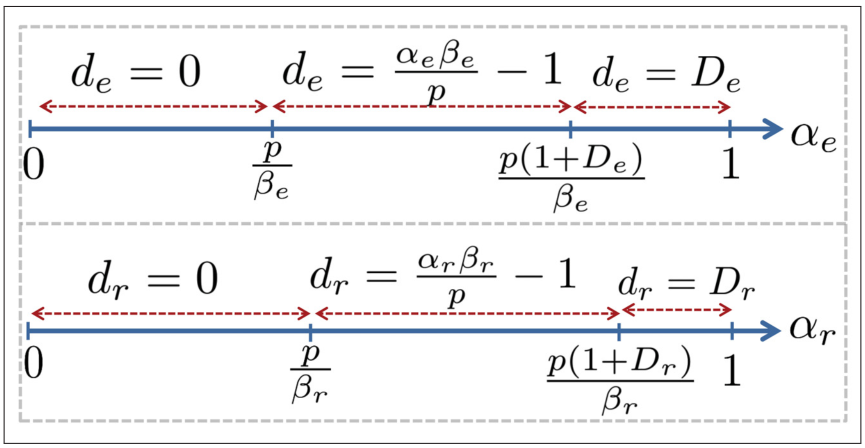

Figure 3.13 Optimal content usage with respect to α_e and α_r in low price regime $p \le \frac{\beta_r}{1+D_r}$

When the low price regime is applied, a part of users achieve the maximum usage for content types e or r or both, as depicted in Fig. 3.13. The profit of SP is given by (3.18) that is placed on the next page. The quadratic profit function of (3.18) is concave and the optimal price is given by:

$$p_l = \frac{\beta_e\beta_r(D_e+D_r)}{\beta_e\left((1+D_r)^2-1\right)+\beta_r\left((1+D_e)^2-1\right)}. \tag{3.17}$$

Since $D_r >> 1$, it can be easily proved that $p_l < \frac{\beta_r}{1+D_r}$ and the maximum value of concave profit function is the optimal value within the price boundary.

$$\pi_l^{SP}(\gamma^{SP}=0,p) = N_T \times p \times \Bigg[\int_{\alpha_e=\frac{p}{\beta_e}}^{\frac{p(1+D_e)}{\beta_e}} \frac{\alpha_e\beta_e}{p} - 1 \, d\alpha_e + \int_{\alpha_e=\frac{p(1+D_e)}{\beta_e}}^{1} D_e \, d\alpha_e$$
$$+ \int_{\alpha_r=\frac{p}{\beta_r}}^{\frac{p(1+D_r)}{\beta_r}} \frac{\alpha_r\beta_r}{p} - 1 \, d\alpha_r + \int_{\alpha_r=\frac{p(1+D_r)}{\beta_r}}^{1} D_r \, d\alpha_r \Bigg]$$
$$= N_T \left(-\frac{p^2}{2} \left(\frac{1}{\beta_r}\big((1+D_r)^2-1\big) + \frac{1}{\beta_e}\big((1+D_e)^2-1\big) \right) + (D_e+D_r)p \right) \tag{3.18}$$

The optimal value of p is the one which maximizes the profit of SP. Since we derived the optimal value for each pricing regime, the final value can be defined as:

$$\pi_o^{SP}(\gamma^{SP}=0,p) = \max \Bigg(\max_p \pi_{uh}^{SP}(\gamma^{SP}=0,p), \max_p \pi_h^{SP}(\gamma^{SP}=0,p),$$
$$\max_p \pi_m^{SP}(\gamma^{SP}=0,p), \max_p \pi_l^{SP}(\gamma^{SP}=0,p) \Bigg), \tag{3.19}$$

$$p^o = \underset{p}{\operatorname{argmax}} \, \pi_o^{SP}(\gamma^{SP}=0,p). \tag{3.20}$$

3.2.4 The profit of SP in cooperative strategy ($\gamma^{SP}=1$)

In the previous subsection, we analyzed the profit of SP under non-cooperative strategy, $\gamma^{SP} = 0$. We categorized the best price responses of SP into four regimes which yield different usage patterns for both content-types e and r. Consequently, the profit values for these regimes vary. If SP decides not to cooperate, then it selects the price regime that maximizes its profit. Since the profit in all price regimes is related to four market parameters β_e, β_r, D_e and D_r, we must adopt a parametric solution for the cooperative strategy of SP as well. In this manner, by assuming that each price regime is applied to the SP's network, one can derive the cooperative profit counterpart. We start our analysis by defining user behavior when SP participates in the SFC program.

When SP aims to implement the SFC program, users are not charged for demanding content Type e. The worst scenario for SP is that all users utilize content e to its maximum level D_e and, simultaneously, no user is willing to raise its content r demand. Since the very first condition in the SFC program is the price invariance, SP loses all the profit from the content e without obtaining extra value transfer of content e. This condition is previously formulated in (3.11). One can apply this equation to different price regimes to obtain the profit of SP in the cooperative state. For example:

3.2.4.1 Ultra high price regime: $\beta_r < p < \beta_e$

In the ultra-high price regime, the entire data traffic belongs to content Type e. Hence, by participating in the SFC program, the profit of SP solely comes from CP as follows:

$$\pi^{SP}(p, \gamma^{SP} = 1, p^{CP}) = N_T \times D_e \times p^{CP}. \tag{3.21}$$

It is clear that SP agrees to participate in the SFC program if and only if $\pi^{SP}(p, \gamma^{SP} = 1, p^{CP}) \geq \pi^{SP}(\gamma^{SP} = 0, p)$. Based on (3.13) and (3.21), $p^{CP} > \frac{1}{D_e}\left(\frac{\beta_r^2}{2\beta_e} - \beta_r + \frac{\beta_e}{2}\right)$ is the sufficient condition for this case. Deriving the profit function for the other three price regimes is straightforward so it is omitted to simplify the presentation.

3.2.5 Stage I: The strategies of CP

In Stage I, CP decides if the SFC program is profitable to itself and if yes, which data unit price should be offered to SP for its profit loss. Hence, one can define the strategy pair (γ^{CP}, p^{CP}) for CP in which $\gamma^{CP} \in \{0, 1\}$ represents the SFC participation of CP and $p^{CP} \in \Re^+$ is the data unit price as a base for payment to SP. As we discussed in the previous section, a Category 1 application is offered free of charge to the users and the central part of CP's profit comes from advertisements. The ad price is related to the number of clicks, and it is accepted in related studies to connect the click frequency to the number of content requests from users. While it is common to consider a logarithmic payoff function for CP (e.g., see (Lotfi *et al.*, 2015)), we aim to consider a worst-case scenario in which the profit of CP is linearly related to the content requests. The benefit of such consideration is that by proving the possibility of SFC program under a linear profit model of CP, the logarithmic profit model also holds valid. The reason for the validity is the direction of payment which is from CP to SP. Thus, the more profit CP makes, the bigger chance of SFC possibility. Since the type of profit for the CP and SP is defined based on actually gained money, their utility is transferable by a side-payment. We define the profit function of CP as follows:

$$\pi^{CP}(\gamma^{CP} = 0) = N_T \eta \int_{\alpha_e=0}^{1} d_e(\alpha_e, p)\, d\alpha_e, \tag{3.22}$$

$$\pi^{CP}(\gamma^{CP} = 1, p^{CP}) = N_T D_e \left(\eta - p^{CP}\right), \tag{3.23}$$

where η is the CP's profit ratio for overall usage of content Type e. When $\gamma^{CP} = 0$, CP does not make a side-payment to SP and hence, $p^{CP} = 0$. The overall data usage for $\gamma^{CP} = 1$ is $N_T D_e$ which is considered along with a side-payment to SP in (3.23). To make a cooperation feasible, CP's profit after cooperation should be greater than the sum of its profit before implementing SFC program and the profit loss of SP, that is:

$$\begin{aligned} & N_T D_e \eta > N_T(\eta + p^o) \int_{\alpha_e=0}^{1} d_e(\alpha_e, p^o)\, d\alpha_e \rightarrow \\ & \eta > \frac{p^o \int_{\alpha_e=0}^{1} d_e(\alpha_e, p^o)\, d\alpha_e}{D_e - \int_{\alpha_e=0}^{1} d_e(\alpha_e, p^o)\, d\alpha_e}, \end{aligned} \tag{3.24}$$

where p^o is the optimal price of SP in Stage II. $\int_{\alpha_e=0}^{1} d_e(\alpha_e, p^o)\, d\alpha_e$ is the overall usage of content e in the non-cooperative form of the game. If the above feasibility condition holds, CP can consider the SFC program. Otherwise, the best response of CP is $\gamma^{CP} = 0$. In the case of possible cooperation, the only remaining decision value for CP is p^{CP} or in general, the amount of side-payment to SP. Several options can be considered in such case. One can find this game as a bargaining game and compute p^{CP} as the solution of a Nash bargaining game

(Rubinstein & Osborne, 1990). Another option is considering the game as a cooperative type. In this case, the solution concepts such as Core and Shapley value (Fudenberg & Tirole, 1991) can be applied. In this paper, we consider both the bargaining solution and Shapley value.

3.2.6 Nash bargaining solution (NBS)

In this part, we find p^{CP} as a solution to the Nash bargaining game (NBS). Nash firstly introduced the NBS in (Nash Jr, 1950) and described a bargaining situation in which players try to reach an agreement. The agreement can be a price definition or a contract between bargainers. Nash built his solution based on four axioms. Namely, Invariance to Equivalent Utility Representations, Symmetry, Independence of Irrelevant Alternatives, and Pareto efficiency. We refer the reader to (Rubinstein & Osborne, 1990) for more information on these axioms. In what follows we give a general definition of two-player NBS.

Definition 3.1. Consider two players 1 and 2 who try to reach an agreement in a bargaining game. Set A contains the agreement alternatives. If they cannot reach the agreement, a disagreement event D occurs. Players have a preference ordering on set $A \cup D$. We define $U^i : A \cup D \to \Re$ as the utility of player i. The union of all payoff pairs $(U^1(a), U^2(a))$ $a \in A$ is indicated by S. The disagreement utility point is defined by the pair $d = (U^1(D), U^2(D))$.

Definition 3.2 ((Rubinstein & Osborne, 1990))**.** The unique solution to Nash's four axioms of bargaining in a two player game is a pair $f^2 \in \Re^2$ given by:

$$f^2(S,d) = \underset{(d_1,d_2)<(s_1,s_2)\in S)}{\arg\max} (s_1 - d_1)(s_2 - d_2). \tag{3.25}$$

If player 1 has a relative bargaining power $\zeta \in [0,1]$ over its opponent, NBS is given by:

$$f^2(S,d) = \underset{(d_1,d_2)<(s_1,s_2)\in S)}{\arg\max} (s_1 - d_1)^{\zeta}(s_2 - d_2)^{1-\zeta}. \tag{3.26}$$

Based on the above definition, we can define the following solution for our problem:

Proposition 7. In a CP-SP game in which SP has a bargaining power $\zeta \in [0,1]$ over CP, if $\zeta \geq \frac{p^o(D_e - \int_{\alpha_e=0}^{1} d_e(\alpha_e, p^o)\, d\alpha_e)}{\eta D_e - (\eta + p^o)\int_{\alpha_e=0}^{1} d_e(\alpha_e, p^o)\, d\alpha_e}$ and feasibility condition in (3.24) is satisfied, then the NBS price p_b^{CP} is given by:

$$p_b^{CP} = \zeta\eta - \frac{(\zeta(\eta + p^o) - p^o)\int_{\alpha_e=0}^{1} d_e(\alpha_e, p^o)\, d\alpha_e}{D_e}, \tag{3.27}$$

otherwise, a disagreement occurs.

A proof is given in Appendix III-3. Nonnegative NBS price in (3.27) is supported by the SFC feasibility condition in (3.24). In other words, NBS price (3.27) is not a solution for a SFC program if feasibility condition of (3.24) does not hold. NBS price p^{CP} should be calculated for each pricing regime of SP and its associated overall usage of content e. In Stage II of the game, since we already derived a closed-form representation of the parameters mentioned above under each pricing regime in Stage II of the game, we omit redundant equations that are created by straightforward parameter substitution.

3.2.7 Shapley value

The multi-stage game is considered as a strategic type and should be solved by the related solution concepts as we did in the previous subsection. However, in the game that we consider, increasing the profit of CP does not decrease the profit of SP. To be more precise, CP and SP are not direct competitors. Thus, one can consider the CP-SP game as a cooperative form. There are several options to solve a coalitional game. As an option, we consider Shapley value which defines the profit of each player by its relative power in the market. As previously mentioned, we know that the direction of payment is from CP to SP. Also, the utility of providers is represented by a monetary unit which is transferable. For such case, the definition of Shapley value is as follows:

Definition 3.3. Consider an n-player game which the set of players N. The function $v(S)$ defines the utility of coalition $S \subset N$. The Shapley value to player $i \in N$ is defined by a unique function Φ that satisfies Shapley's three main axioms. Namely, Symmetry, Carrier and Linearity (see (Myerson, 1991)) and is given by:

$$\Phi^i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N|-|S|-1)!}{|N|!} \left(v(S \cup i) - v(S)\right). \tag{3.28}$$

For a two person game, the above equation gives:

$$\Phi^1 = \frac{1}{2}\left(v(12) + v(1) - v(2)\right), \tag{3.29}$$

$$\Phi^2 = \frac{1}{2}\left(v(12) + v(2) - v(1)\right), \tag{3.30}$$

where $v(12)$ is the profit of cooperation.

Proposition 8. In the CP-SP game, the Shapley value of SP, Φ^{SP}, is given by $D_e \times p^{CP}_{b|\zeta=\frac{1}{2}}$, where $p^{CP}_{b|\zeta=\frac{1}{2}}$ is the NBS price with $\zeta = \frac{1}{2}$. Hence,

$$\Phi^{SP} = \frac{N_T}{2}\left(\eta D_e - (p^o + \eta)\int_{\alpha_e=0}^{1} d_e(\alpha_e, p^o)\, d\alpha_e\right). \tag{3.31}$$

For the proof, see Appendix III-4. In the next section, we show the feasibility of SFC program and the value of shared profit for several numerical scenarios.

3.3 numerical results

To have a visual representation of the feasibility of SFC program, we consider several examples which differ in user and provider parameters such as D_e, β_e, η and bargaining power ζ. Similar to real markets and characteristics of Type e and r applications, we set $\beta_e > \beta_r$ and $D_r >> 10D_e$. These settings assure us that the numerical examples exactly follow the real behavior of cellular users covered in Ericsson's statistics. Hence, throughout this analysis the only constants are $D_r = 100$, $\beta_r = 5$. The main parameters that we like to analyze are the profit values of SP and CP, side-payment price p^{CP}, and the minimum required bargaining power of SP, ζ, that makes the SFC program feasible. We take D_e which is the maximum desired usage of the eligible application as the primary independent variable in the x-axis. However, in each example, there is an additional variable whose effect is shown by introducing several curves in each figure. For example, Figs. 3.14 and 3.15 represent the profit of SP and CP for $2 \leq D_e \leq 5$ and $\beta_e \in \{6, 10\}$. As indicated in Fig. 3.14, when SP has the equal bargaining power, $\zeta = 0.5$, the desired SFC area starts from $D_e \approxeq 2.1$ when $\beta_e = 6$. Increasing β_e to 10, leads to higher profit for non-SFC program for SP and it requires a value of $D_e > 3.9$ to make SFC feasible. The same analysis can be applied to the profit of CP in Fig. 3.15. In Figs. 3.16 and 3.17 we decided to freeze β_e at 10 and change η as the profit factor of CP. As it is expected, increasing the value of η, decreases the required value of D_e for SFC feasibility, that is, for $\eta = 2$ the minimum value for D_e should be 3.9 while for $\eta = 4$, D_e can be 1.5 or higher. Fig. 3.18 shows the unit price for side-payment and the minimum bargaining power, ζ_{Min}, for the feasibility of SFC. For $\beta_e > \beta_r$, ζ_{Min} acquires a lower value comparing to $\beta_e = 10$. The main reason can be found in Fig. 3.14 where a lower β_e gives a higher profit value to SP which contrasts with what happens to CP in Fig. 3.15. Hence, SP needs less bargaining power to dictate the SFC program.

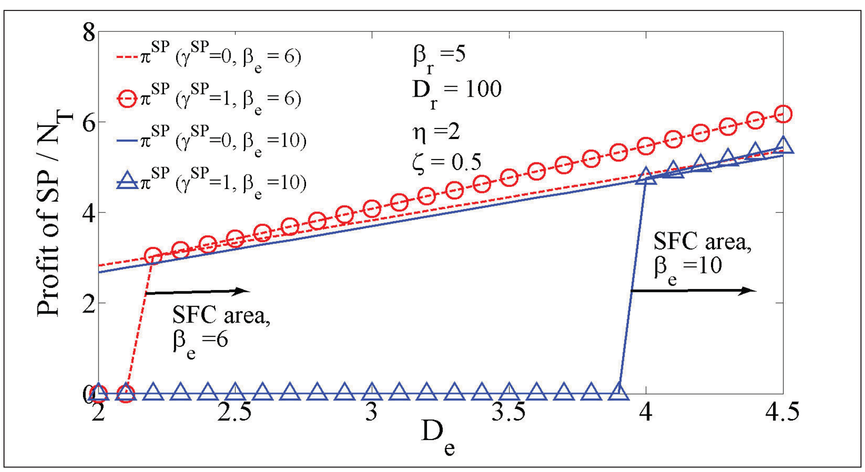

Figure 3.14 Profit of SP with and without SFC program for different values of D_e and β_e

Finally, in Fig. 3.19 we take the bargaining power of SP, ζ, as the independent variable in the x-axis. Here we can observe two effects considering D_e and ζ. Firstly, by having a higher

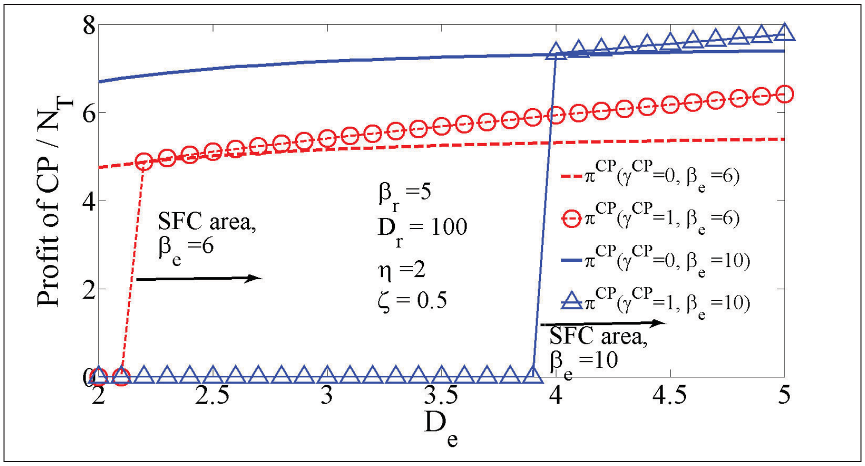

Figure 3.15 Profit of CP with and without SFC program for different values of D_e and β_e

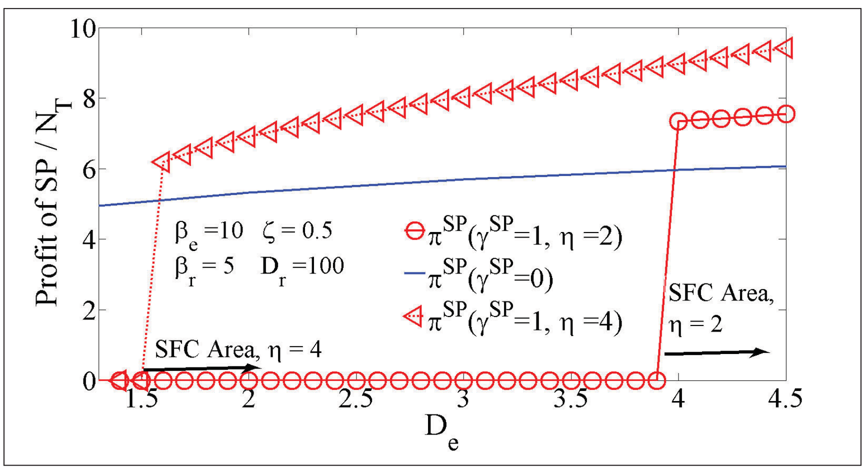

Figure 3.16 Profit of SP with and without SFC program for different values of D_e and η

value of D_e, the overall profit of both CP and SP increases. Secondly, by increasing ζ, SP can force CP to pay it a bigger chunk of profit in SFC program. Also, a bargaining power of 1 leaves no additional profit for CP in SFC program. In summary, the presented results show the possibility of SFC program for the eligible applications, even if CP's profit is linearly related to the size of transferred content.

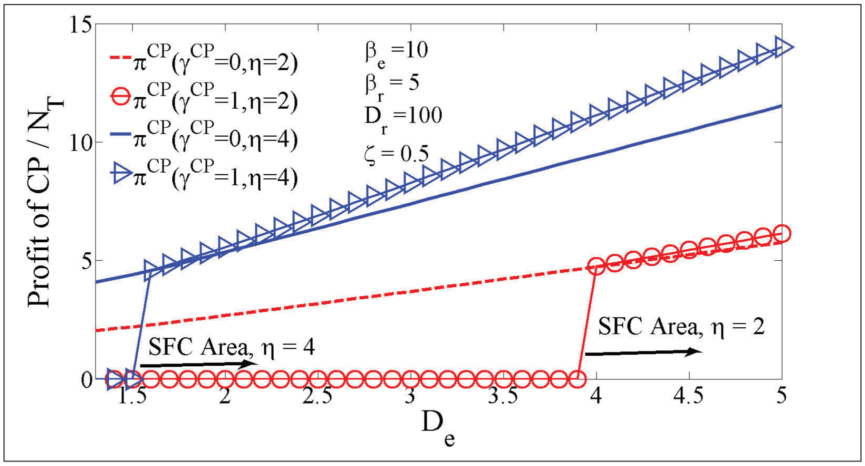

Figure 3.17 Profit of CP with and without SFC program for different values of D_e and η

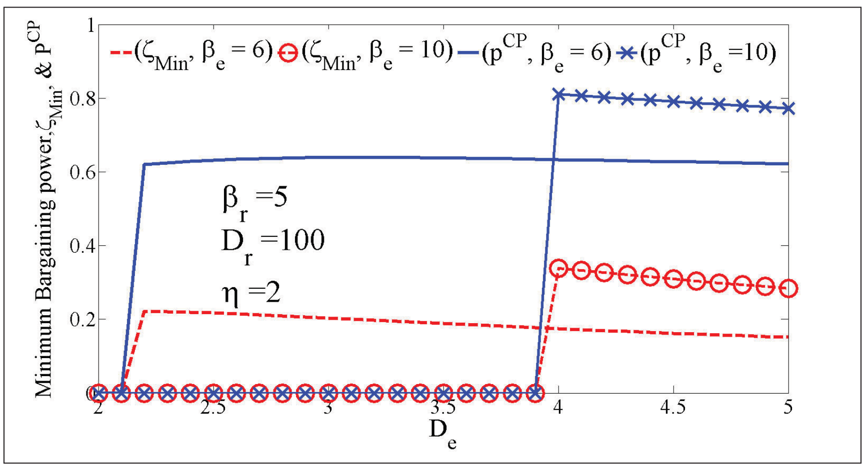

Figure 3.18 Minimum bargaining power, ζ, for SFC program and related p^{CP}

3.4 Conclusion

In this paper, we analyzed the recent statistics of user behavior in cellular markets and identified three types of services. Type I that requires a vast amount of data transfer, yet has low priority to the users. Type II where the services require high data transfer, but users are not greedy to utilize them constantly. Type III contains important applications that require low bandwidth but carry sensitive information for the users. We showed that there are several examples of

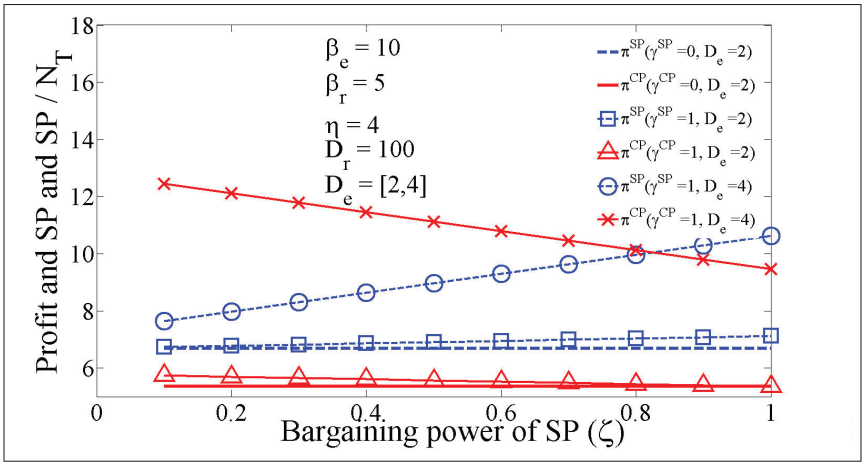

Figure 3.19 Profit of SP and CP with and without SFC program for different values of D_e and ζ

the Type III services that can be candidates for a selective free content (SFC) program for the users. This program should be implemented by cooperation between SP and CP. Three categories of such services are introduced. Namely, mapping and personal assistant applications, real-time cloud-based services and non-profit e-governance applications. Then, we analyzed the Category 1 services which are the most complicated group for analysis among the three categories. A mathematical framework for the feasibility of SFC program is introduced. We built the framework by modeling the game as a Stackelberg game with three stages. In each stage, one group of market entities is involved; namely, users, SP and CP. The game is solved by backward induction. Finally, several numerical examples are demonstrated. These cases are constructed based on the real behavior of users that were acquired from recent Internet statistics. With these examples, we visually explained the conditions in which an SFC program is feasible.

CONCLUSION AND RECOMMENDATIONS

In this thesis, we investigated several types of pricing schemes in the heterogeneous wireless networks and used them to build a framework for coalition analysis. In the first part of this thesis, we started by analyzing the volume-based pricing plan which is the dominant type of pricing in today's market. We used a utility function to model the plan selection process of users. The main parameters in the user's utility are the price and cap on data plan along with the data rate and service availability. To model the access methods and subsequently the guaranteed data rate we considered two spectrum allocation methods in the access networks: the shared carrier scheme and the dynamic spectrum allocation. The service availability is modeled based on the number of users who subscribed to the plans and their usage pattern based on the size of data plan. A multi-package market in which user's budget defines its plan selection mechanism is studied in the next stage. The final objective was to maximize the profit of the provider based on the price of data plan and the number of users. The analysis is supported by several numerical examples which simulate the real world market and network conditions. The outcome of numerical scenarios showed that our analytical framework is a suitable benchmark to build volume-based plans with SLA included.

In the second part of the thesis, we analyzed the flat pricing method in which users pay for their data usage based on a fixed data unit price. Concerning the potential users entering the market, we used a form of the utility function that has the parameters of price, data, and speed. We used a speed satisfaction random variable in the utility function which shows the perceptual importance of service speed to each user. Several market forms including monopoly, duopoly, and oligopoly are studied. The analysis of oligopoly market is the foundation to define the competitive state of a real wireless market. We proved that when the providers experience a linear cost increment with respect to their coverage size, they are better off to expand their network without forming any coalition. However, in certain forms of exponential costs, providers would gain higher profit under cooperation. Then, we defined a multi-provider payoff function for the users which is used to analyze their usage behavior when they subscribed to a multi-provider coalition. We also defined a utility function designed specifically for the coalition of cellular and WiFi providers. In this function, the fixed-price fixed-time pricing scheme is considered which is popular in WiFi markets such as the ones in hotels and airports. Finally, we used a coalition formation process and modified it based on the profit of heterogeneous wireless providers. We proved that in wireless markets, in which there is a churn rate for the newer technologies, the coalition formation process is always led to stable coalitions. We investigated several scenarios such as WiFi-Cellular, WiFi-3G-4G, and multi-cellular markets and showed that forming coalitions can increase the profit of the providers.

In the third part of this thesis, we focused on service-oriented HetNets in which content providers, service providers, and users are the main entities. We studied the usage patterns on cellular subscribers based on the type of content such as video, voice, web browsing, etc. We showed that for certain types of data content such as the ones associated with mapping applications, users are not greedy to utilize them regularly, yet the information delivered by such applications is highly valuable to the users. We used this fact, derived from real market information, to build a type of coalition between CPs and SPs in which the candidate content types are delivered to the end users free of charge. In this case, the SP profit loss is compensated from the CP increased profit associated with advertisement sources. In such way, users pay only for their data connectivity and not the data transfer. We showed that such treatment is not applicable to the video contents since users are extremely motivated to utilize such applications continuously.

For the future work, we see many research opportunities in this field. For example, the multi-provider market for the volume-based pricing is one of the potential areas. Since the majority of wireless providers set their primary data plans based on the data volume cap, a comprehensive analysis of their competition strategies is required. Another research opportunity is analyzing the usage pattern of the subscribers based on a given data rate and price. Since the utility function is playing the fundamental role in game theory based analysis, we need to define the value of main network parameters in a way that represents the real user's behavior. This analysis can be based on regional interests and can cover multiple wireless technologies as well. Finally, we believe that a pilot implementation of the SFC program can be a perfect complement to the results of this thesis and can help us to estimate the required network parameters for the further analysis.

APPENDIX I

PROOFS OF THE PROPOSITIONS IN CHAPTER 1

1. Concavity of objective function in Eq. (1.13)

The objective function in Eq. (1.13) has the form:

$$\max_{c_d, p_d} \pi(c_d, p_d) = N_C \times p_d \left(1 - \frac{\varepsilon_d + p_d}{\bar{\gamma}(p_d, c_d)\theta_d c_d}\right), \tag{A I-1}$$

$$\bar{\gamma}(p_d, c_d) = \frac{\zeta_2 C_X^M - \sigma c_d + \left((\zeta_2 C_X^M - \sigma c_d)^2 + 4\frac{\sigma}{\theta_d}(\varepsilon_d + p_d)\right)^{\frac{1}{2}}}{2}. \tag{A I-2}$$

1.1 Concavity with respect to c_d

Set $X = \zeta_2 C_X^M - \sigma c_d$ and $Y = \left((\zeta_2 C_X^M - \sigma c_d)^2 + 4\frac{\sigma}{\theta_d}(\varepsilon_d + p_d)\right)$. Hence $\bar{\gamma} = \frac{1}{2}(X + Y^{\frac{1}{2}})$ and:

$$\frac{\partial \pi}{\partial c_d} = \frac{2N_C(\varepsilon_d + p)}{\theta_d} \times \frac{(X + Y^{\frac{1}{2}}) - c_d(\sigma + \sigma X Y^{\frac{-1}{2}})}{\left(c_d(X + Y^{\frac{1}{2}})\right)^2}, \tag{A I-3}$$

$$\begin{aligned}\frac{\partial^2 \pi}{\partial c_d^2} = \frac{2N_C(\varepsilon_d + p)}{\theta_d((X + Y^{\frac{1}{2}})c_d)^3} \times \Bigg[&\left(-\sigma c_d - \sigma c_d X Y^{\frac{-1}{2}} - (\sigma + \sigma X Y^{\frac{-1}{2}}) - (\sigma X)^2 Y^{\frac{-1}{3}}\right) \times \\ &((X + Y^{\frac{1}{2}})c_d) - \left((X + Y^{\frac{1}{2}}) - c_d \sigma(1 + X Y^{\frac{-1}{2}})\right) \times \\ &2\left(X + Y^{\frac{1}{2}} - c_d \sigma(1 + X Y^{\frac{-1}{2}})\right)\Bigg].\end{aligned} \tag{A I-4}$$

The only positive term in the numerator of above equation is $2\sigma c_d(2X + X^2 Y^{\frac{-1}{2}} + Y^{\frac{1}{2}})$ which is smaller than negative parts. Hence the objective function is always concave for all $c_d > 0$.

1.2 Concavity with respect to p_d

Set X and Y the same amount of above equations, we have:

$$\frac{\partial \pi}{\partial p_d} = N_C \left(1 - \frac{2}{c_d} \frac{(\varepsilon_d + 2p_d)(X + Y^{\frac{1}{2}}) - \frac{2\sigma p_d(\varepsilon_d + p_d)}{\theta_d} Y^{\frac{-1}{2}}}{(X + Y^{\frac{1}{2}})^2} \right), \tag{A I-5}$$

$$\begin{aligned} \frac{\partial^2 \pi}{\partial p_d^2} = \frac{2N_C}{c_d(X + Y^{\frac{1}{2}})^2} \times \Bigg[& -2(X + Y^{\frac{1}{2}}) - 4\frac{\sigma}{\theta_d} p_d(\varepsilon_d + p_d) Y^{\frac{-3}{2}} \\ & - \frac{\left(\frac{2\sigma}{\theta_d}\right)^2 p_d(\varepsilon_d + p_d) Y^{-1}}{X + Y^{\frac{1}{2}}} + \frac{\sigma}{\theta_d}(\varepsilon_d + 2p_d) Y^{\frac{-1}{2}} \Bigg]. \end{aligned} \tag{A I-6}$$

The positive part $\frac{\sigma}{\theta_d}(\varepsilon_d + 2p_d)Y^{\frac{-1}{2}}$ can be written as the sum of $\frac{2\sigma}{\theta_d}(\varepsilon_d + p_d)Y^{\frac{-1}{2}}$ and a negative excess. Setting the equivalent $\frac{2\sigma}{\theta_d}(\varepsilon_d + p_d) = \frac{1}{2}(Y - X^2)$ we have:

$$\frac{\partial^2 \pi}{\partial p_d^2} = \frac{2N_C}{c_d(X + Y^{\frac{1}{2}})^2} \times \left[-2(X + Y^{\frac{1}{2}}) + \frac{1}{2}(Y - X^2) Y^{\frac{-1}{2}} - \ldots \right]. \tag{A I-7}$$

Considering the values of X and Y, the above equation is clearly negative for all values of p_d and concavity is proved.

APPENDIX II

PROOFS OF THE PROPOSITIONS IN CHAPTER 2

1. Proof of Theorem 1

Since the possible actions, P_i, for each provider form a compact and convex set and the profit function is continuous over p_i, all we need is proving the quasi-concavity of the profit function over p_i. If $p_i = \alpha_i$, then the profit will be $-c_i(G_i)$. For the large value of $p_i = G_i K$ the usage will be zero (based on (2.8)) and again the profit goes toward $-c_i(G_i)$. Hence, there are two possible types of providers when $p_i \in [\alpha_i, G_i K]$:

Type 1) Providers that cannot attract any user. Hence, their only strategy in a competitive market is choosing $p_i = \alpha_i$ and their profit is always $-c_i(G_i)$.

Type 2) The providers for which $p_i \in [\alpha_i, G_i K]$ leads to a positive number of new registered users, e.g., $I_i = [s_{1i}, s_{2i}]$, and $p_i = \alpha_i + \varepsilon,\ \ \varepsilon \to 0$, leads to an increase in the provider profit that becomes greater than $-c_i(G_i)$. Now with increasing price p_i, the user set remains constant or reduced, but the profit is increased until a price level (denoted by p_i^M) in which increasing the price by the provider, causes no change or a reduction in the provider profit. Price levels higher than p_i^M lead to continuous decrease on provider profit (due to smaller user-set and lower maximum usage level). With the help of the π_i continuity, we can conclude that the profit functions of these providers are concave or quasi-concave. Hence, for the providers who are involved in the game and their actions can make a change in their profit, the profit functions are concave or quasi-concave and the game has a pure strategy Nash equilibrium.

2. Proof of Theorem 2

We first prove the sufficiency.

a. The first derivative of the payoff with respect to the usage d leads to :

$$\frac{\partial U_C^j(d)}{\partial d} = \frac{K\,(\sum_{i=1}^{|C|} G_i\, s_i^j)}{1+K\,d} - \frac{\sum_{i=1}^{|C|} p_i\, G_i}{\sum_{i=1}^{|C|} G_i}, \tag{A II-1}$$

which is a decreasing function of d and confirms the first property.

b. The second property can be readily achieved by using the same price and speed for all providers.

c. The payment to the Provider k in MPP is $\frac{p_k\, G_k}{\sum_{i=1}^{|C|} G_i}\, d$ which is the proportion of the coverage area of provider k to the sum of all coverage area sizes of all the providers in the coalition and this is consistent with Property 3.

To complete the proof, it must be shown that the three properties also lead to the same payoff function. In what follows, we prove the case for coalitions with 2 providers. The proof for coalitions of 3 or more providers can be constructed by following the same. When a user selects the service powered by a two-provider coalition, the probability of utilizing Network 1 is G_1 and respectively G_2 for Network 2. If the user is under the coverage of Network 1, then by consuming a relatively small amount such as Δd, it gets the utility amount $s_1 \frac{K\Delta d}{1+K\Delta d}$ and pays $p_1 \Delta d$ for it. However, the payment has a conditional probability, that means the user is charged under price p_1 if it uses network 1 rather than network 2 and vice versa. Hence, the probability of being charged under price p_1 (if user already consumed the amount Δd) is $\frac{G_1}{G_1+G_2}$. Combining the three properties, we have the following expected payoff:

$$U_C(\Delta d) = G_1\,(s_1\,\frac{K\Delta d}{1+K\Delta d}) + G_2\,(s_2\,\frac{K\Delta d}{1+K\Delta d}) - \frac{G_1}{G_1+G_2} p_1 \Delta d - \frac{G_2}{G_1+G_2} p_2 \Delta d. \tag{A II-2}$$

Now, suppose that the same user wants to consume an additional amount of Δd, here every element of (A II-2) remains the same except for the gained utility. Based on the first property which is the *law of diminishing marginal utility*, the marginal utility of user decreases by more consumption, which means the subscriber is most satisfied with the first amount of data usage and so forth. Hence, the user gains $s_i \frac{K\Delta d}{1+2K\Delta d}$ instead of $s_i \frac{K\Delta d}{1+K\Delta d}$ and we have the conditional payoff:

$$U_C(2\Delta d|\Delta d) = G_1(s_1 \frac{K\Delta d}{1+2K\Delta d}) + G_2(s_2 \frac{K\Delta d}{1+2K\Delta d}) - \frac{G_1}{G_1+G_2} p_1 \Delta d - \frac{G_2}{G_1+G_2} p_2 \Delta d.$$

The overall payoff is the accumulation of all small payoffs. By setting the desired overall usage d as the final consuming amount and letting $\Delta d \to 0$, the summation $\sum_{i=1}^{[\frac{d}{\Delta d}]} U_C(i\Delta d|(i-1)\Delta d)$ becomes an integral which leads to (2.39).

3. Proof of proposition 3

The proof is similar to the one that is given for Proposition 2.4 in (Bloch, 1996). Let the profit of provider i in each period be multiplied by a period *discount factor*, $0 < \sigma < 1$, which is inversely proportional to the churn rate (lower churn rates lead to the higher values of σ). Hence, the profit of the provider i in the period n is: $\pi_i^n = \sigma^n \pi_i$, where π_i is the profit at the first period. Let indicate the sequential coalition game by *Seq* and the new game (with the discount factor) be Seq_σ. The game Seq_σ is an infinite horizon game with *continuity at infinity*, which means with infinite repetition of coalition offers (negotiation sequences), the gained payoff in the higher periods becomes less important (due to the discount factor of profits). For such a game the *one stage deviation principle* always holds, that is, if γ is a sub-game perfect strategy and h^n is the history of game until period n, then there is no other sequential strategy like $\hat{\gamma}_i$ and history $h^{\bar{n}}$ in which $\hat{\gamma}_i$ has the same actions as γ_i except in one period, and $\hat{\gamma}_i$ is a better response to γ_{-i} conditional reaching the same history $h^{\bar{n}}$. Hence, by applying this principle, the SPE can be found. We refer the readers to the Theorem 4.2 of (Fudenberg & Tirole, 1991) for the proof of “one stage deviation principle".

APPENDIX III

PROOFS OF THE PROPOSITIONS IN CHAPTER 3

1. Proof of Proposition 5

Comparing types e and r contents, if Type e content belongs to Type III category of applications, we have $\lim_{d\to 0} I^e(d) > I^r(d)$. For $d \to 0$ ($p < \max(\beta_e, \beta_r)$), the data usage for each content type i and user j is indicated by $\frac{\alpha_i^j \beta_i}{p} - 1$. With this condition, there are two types of users; firstly, the group of users with $\alpha_e\beta_e > \alpha_r\beta_r$ who prefer the application Type e over r and the group with $\alpha_e\beta_e < \alpha_r\beta_r$. As stated above, to have content e as a Type III application, when $d \to 0$, the overall usage of the first group should be greater than the second group, which means at near zero usage, the number of users in favor of content e should be greater than the other group, that is:

$$\int_{\alpha_e=0}^{1}\int_{\alpha_r=0}^{\frac{\beta_e\alpha_e}{\beta_r}} f(\alpha_r)f(\alpha_e)d\alpha_r d\alpha_e > \int_{\alpha_r=0}^{1}\int_{\alpha_e=0}^{\frac{\beta_r\alpha_r}{\beta_e}} f(\alpha_e)f(\alpha_r)d\alpha_e d\alpha_r \to$$
$$\frac{\beta_e}{\beta_r} > \frac{\beta_r}{\beta_e} \to \beta_e^2 > \beta_r^2, \tag{A III-1}$$

since both values are positive, the above inequality gives $\beta_e > \beta_r$. □

2. Proof of Proposition 6

We must prove the concavity of the profit function for $D_e \geq 1$. The profit of SP in moderate price regime has a quadratic form with first and second derivatives as follows:

$$\pi_m^{SP}(\gamma^{SP} = 0, p) = N_T\left(\frac{p^2}{2}\left(\frac{1}{\beta_r} - \frac{1}{\beta_e}\left((1+D_e)^2 - 1\right)\right) + (D_e - 1)p + \frac{\beta_r}{2}\right),$$

$$\frac{\partial \pi_m^{SP}(\gamma^{SP}=0,p)}{\partial p} = N_T\left(p\left(\frac{1}{\beta_r}-\frac{1}{\beta_e}\Big((1+D_e)^2-1\Big)\right)+(D_e-1)\right), \tag{A III-2}$$

$$\frac{\partial^2 \pi_m^{SP}(\gamma^{SP}=0,p)}{\partial p^2} = N_T\left(\left(\frac{1}{\beta_r}-\frac{1}{\beta_e}\Big(D_e(D_e+2)\Big)\right)\right). \tag{A III-3}$$

The first derivative has one extreme point at $p=\frac{\beta_e\beta_r(D_e-1)}{\beta_r((1+D_e)^2-1)-\beta_e}$. To have this point as a global maximum, we can prove that for $D_e \geq 1$, the extreme point is always positive and the second derivative in (A III-3) is always negative:

$$N_T\left(\left(\frac{1}{\beta_r}-\frac{1}{\beta_e}\Big(D_e(D_e+2)\Big)\right)\right)<0\rightarrow$$

$$\beta_r > \frac{\beta_e}{D_e(D_e+2)} \tag{A III-4}$$

which is always true, since the threshold order is $\beta_r > \frac{\beta_e}{D_e+1}$. The above inequality also proves that the denominator of extreme point is always positive. Since $D_e \geq 1$, we have a positive extreme point with negative second derivative. Hence the extreme point is a global maximum for all $D_e \geq 1$, otherwise, for all $D_e < 1$ the extreme point is negative and the maximum of profit function occurs at the lower limit of price $\frac{\beta_r}{D_r+1}$. □

3. Proof of Proposition 7

First we show the optimum value of p^{CP} and then prove the boundary value of ζ. By taking the equations $\pi^{SP}(\gamma^{SP}=0)$ from (3.10), $\pi^{SP}(\gamma^{SP}=1)$ from (3.11), $\pi^{CP}(\gamma^{CP}=0)$ from (3.22), $\pi^{CP}(\gamma^{CP}=0)$ from (3.23), and putting into the NBS objective function (3.26), we achieve (A III-5). Based on the feasibility condition of (3.24) both parts of (A III-5) are always positive. The first derivative of objective function in (A III-6) has one extreme point in $p_b^{CP}=\zeta\eta-\frac{(\zeta(\eta+p^o)-p^o)\int_{\alpha_e=0}^{1}d_e(\alpha_e,p^o)\,d\alpha_e}{D_e}$. The second derivative of objective function with respect to p^{CP} is given by (A III-7) and is always negative. Hence the extreme point is a global maximum. For the lower and upper limits of ζ, we just check the given global maximum p_b^{CP} with $\varepsilon_1(p_b^{CP}) > 0$ and $\varepsilon_2(p_b^{CP}) > 0$ in (A III-5). This gives us the boundary condition $\frac{p^o(D_e - \int_{\alpha_e=0}^{1} d_e(\alpha_e, p^o)\, d\alpha_e)}{\eta D_e - (\eta + p^o)\int_{\alpha_e=0}^{1} d_e(\alpha_e, p^o)\, d\alpha_e} \leq \zeta \leq 1$ for the relative bargaining power of SP. $\square$

$$\begin{aligned} f(p^{CP}) &= N_T^2 \left(\pi^{CP}(\gamma^{CP} = 1, p^{CP}) - \pi^{CP}(\gamma^{CP} = 0)\right)^{\zeta} \\ &\quad \times \left(\pi^{SP}(\gamma^{SP} = 1, p^o) - \pi^{SP}(\gamma^{SP} = 0, p^o)\right)^{1-\zeta} \\ &= N_T^2 \underbrace{\left(D_e(\eta - p^{CP}) - \eta \int_{\alpha_e=0}^{1} d_e(\alpha_e, p^o)\, d\alpha_e\right)^{\zeta}}_{\varepsilon_1(p^{CP})^{\zeta}} \\ &\quad \times \underbrace{\left(D_e p^{CP} - p^o \int_{\alpha_e=0}^{1} d_e(\alpha_e, p^o)\, d\alpha_e\right)^{1-\zeta}}_{\varepsilon_2(p^{CP})^{1-\zeta}}, \end{aligned} \tag{A III-5}$$

$$\frac{\partial f(p^{CP})}{\partial p^{CP}} = N_T^2 D_e \varepsilon_1^{\zeta} \times \varepsilon_2^{-\zeta} \times \left(-\zeta \varepsilon_1^{-1} \varepsilon_2 + 1 - \zeta\right), \tag{A III-6}$$

$$\frac{\partial^2 f(p^{CP})}{\partial (p^{CP})^2} = -N_T^2 D_e^2 \zeta (1-\zeta) \varepsilon_1^{\zeta} \times \varepsilon_2^{-\zeta} \times \left(\varepsilon_1^{-2} \varepsilon_2 + 2\varepsilon_1^{-1} + \varepsilon_2^{-1}\right). \tag{A III-7}$$

4. Proof of Proposition 8

Considering the content Type e, we have the total profit of SFC program as $v(12) = v(SPCP) = \eta D_e$. Taking Shapely value of (3.29) and substituting the profit of content Type e from $\pi^{SP}(\gamma^{SP} = 0)$ (3.10) for $v(1)$ and $\pi^{CP}(\gamma^{CP} = 0)$ of (3.22) for $v(2)$, we have the following profit share for

CP and SP:

$$\Phi^{SP} = \frac{N_T}{2}\left(\eta\left(D_e + \int_{\alpha_e=0}^{1} d_e(\alpha_e, p^o)\, d\alpha_e\right) - p^o \int_{\alpha_e=0}^{1} d_e(\alpha_e, p^o)\, d\alpha_e\right), \qquad \text{(A III-8)}$$

$$\Phi^{CP} = \frac{N_T}{2}\left(\eta\left(D_e - \int_{\alpha_e=0}^{1} d_e(\alpha_e, p^o)\, d\alpha_e\right) + p^o \int_{\alpha_e=0}^{1} d_e(\alpha_e, p^o)\, d\alpha_e\right), \qquad \text{(A III-9)}$$

since Φ^{SP} is defined as the side-payment from CP to SP, we can achieve p^{CP} as:

$$p^{CP} = \frac{\Phi^{SP}}{N_T D_e} = \frac{1}{2}\left(\eta + \frac{(\eta - p^o)\left(\int_{\alpha_e=0}^{1} d_e(\alpha_e, p^o)\, d\alpha_e\right)}{D_e}\right), \qquad \text{(A III-10)}$$

which is the NBS price in (3.27) with $\zeta = \frac{1}{2}$ $\square$

BIBLIOGRAPHY

Vodafone Spain and Orange spain update fibre sharing agreement. Consulted at http://www.vodafone.com/content/index/media/vodafone-group-releases/2014/vodafone-spain-orange-spain-fibre-sharing-agreement.html.

Acemoglu, D., Ozdaglar, A. & Srikant, R. (2004). The marginal user principle for resource allocation in wireless networks. *Decision and control, 2004. cdc. 43rd ieee conference on*, 2, 1544–1549.

Alcaraz, J. J. & van der Schaar, M. Coalitional games with intervention: Application to spectrum leasing in cognitive radio.

Andrews, J. G., Ghosh, A. & Muhamed, R. (2007). *Fundamentals of wimax: understanding broadband wireless networking*. Pearson Education.

Andrews, M., Bruns, G. & Lee, H. (2014). Calculating the benefits of sponsored data for an individual content provider. *Information sciences and systems (ciss), 2014 48th annual conference on*, pp. 1–6.

Andrews, M., Jin, Y. & Reiman, M. I. (2016). A truthful pricing mechanism for sponsored content in wireless networks. *Computer communications, ieee infocom 2016-the 35th annual ieee international conference on*, pp. 1–9.

Anglano, C., Guazzone, M. & Sereno, M. (2014). Maximizing profit in green cellular networks through collaborative games. *Computer networks*, 75, Part A, 260 - 275. doi: http://dx.doi.org/10.1016/j.comnet.2014.10.003.

Apt, K. R. & Witzel, A. (2009). A generic approach to coalition formation. *International game theory review*, 11(03), 347–367.

Aram, A., Sarkar, S. & Singh, C. (2009). Provider-customer coalitional games. *Game theory for networks, 2009. gamenets' 09. international conference on*, pp. 180–189.

AT&T. (2016). AT&T sponsored data. Consulted at https://www.att.com/att/sponsoreddata/en/index.html.

AT&T Data. (2016). AT&T data plans. Consulted at https://www.att.com/shop/wireless/data-plans.html.

Aumann, R. & Dreze, J. (1974). Cooperative games with coalition structures. *International journal of game theory*, 3(4), 217-237. doi: 10.1007/BF01766876.

Aumann, R. & Myerson, R. (1988). Endogenous formation of links between players and coalitions: an application of the shapley value. *The shapley value*, 175–191.

Başar, T. & Srikant, R. (2002). A stackelberg network game with a large number of followers. *Journal of optimization theory and applications*, 115(3), 479–490.

Bell. (2016). Bell data plans. Consulted at http://www.bell.ca/Mobility/Cell_phone_plans/Turbo-Stick-and-Turbo-Hub-data-plans.

Bloch, F. (1996). Sequential formation of coalitions in games with externalities and fixed payoff division. *Games and economic behavior*, 14(1), 90–123.

Bonald, T. & Proutière, A. (2003). Wireless downlink data channels: user performance and cell dimensioning. *Proceedings of the 9th annual international conference on mobile computing and networking*, pp. 339–352.

BSNL. BSNL free roaming for mobile users. Consulted at http://www.bsnlteleservices.com/2015/06/bsnl-free-roaming-mobile-queries.html.

Chan, Y.-W., Chien, F.-T., Chang, R. Y., Chang, M.-K. & Chung, Y.-C. (2013). Spectrum sharing in multi-channel cooperative cognitive radio networks: a coalitional game approach. *Wireless networks*, 19(7), 1553–1562.

Chen, Y., Duan, L., Huang, J. & Zhang, Q. (2015a). Balancing income and user utility in spectrum allocation. *Mobile computing, ieee transactions on*, 14(12), 2460-2473. doi: 10.1109/TMC.2015.2393309.

Chen, Y., Duan, L., Huang, J. & Zhang, Q. (2015b). Balancing income and user utility in spectrum allocation. *Ieee transactions on mobile computing*, 14(12), 2460–2473.

Cisco. (2016). Cisco visual networking index:global mobile data traffic forecast update, 2015–2020 white paper. Consulted at http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html.

ComScore. (2016). The 2016 U.S. Mobile App Report. Consulted at http://www.comscore.com/Insights/Presentations-and-Whitepapers/2016/The-2016-US-Mobile-App-Report.

Courcoubetis, C. & Weber, R. (2003a). *Pricing communication networks: Economics, technology and modelling (wiley interscience series in systems and optimization)*. John Wiley & Sons.

Courcoubetis, C. & Weber, R. (2003b). *Pricing communication networks: Economics, technology and modelling*. Wiley Online Library.

Driessen, T. (1988). *Cooperative games, solutions and applications*. Kluwer Academic Publishers Dordrecht.

Duan, L., Huang, J. & Shou, B. (2013a). Optimal pricing for local and global WiFi markets. *Infocom, 2013 proceedings ieee*, pp. 1088-1096. doi: 10.1109/INFCOM.2013.6566899.

Duan, L., Huang, J. & Shou, B. (2013b). Optimal pricing for local and global wifi markets. *Infocom, 2013 proceedings ieee*, pp. 1088–1096.

Duan, L., Huang, J. & Walrand, J. (2013c). Economic analysis of 4G network upgrade. *Infocom, 2013 proceedings ieee*, pp. 1070–1078.

Duan, L., Gao, L. & Huang, J. (2014). Cooperative spectrum sharing: a contract-based approach. *Mobile computing, ieee transactions on*, 13(1), 174–187.

Dugar, C., Jain, A., Rajawat, A. & Bhattacharya, S. (2015). Dynamic pricing of call rates: Bayesian approach. *Information processing letters*, 115(2), 237 - 242. doi: http://dx.doi.org/10.1016/j.ipl.2014.09.020.

ElDelgawy, R. & La, R. J. (2015). Interaction between a content provider and a service provider and its efficiency. *2015 ieee international conference on communications (icc)*, pp. 5890–5895.

ElNashar, A., El-saidny, M. & Sherif, M. (2014). *Design, deployment and performance of 4G-LTE networks: A practical approach*. John Wiley & Sons.

Ericsson. (2016). Ericsson Mobility Report, June 2016. Consulted at https://www.ericsson.com/res/docs/2016/ericsson-mobility-report-2016.pdf.

Ezziane, Z. (2005). Charging and pricing challenges for 3G systems. *Ieee communications surveys tutorials*, 7(4), 58-68. doi: 10.1109/COMST.2005.1593280.

Fudenberg, D. & Tirole, J. (1991). Game theory. 1991. MIT Press.

Gajic, V., Huang, J. & Rimoldi, B. (2009). Competition of wireless providers for atomic users: Equilibrium and social optimality. *Communication, control, and computing, 2009. allerton 2009. 47th annual allerton conference on*, pp. 1203-1210. doi: 10.1109/ALLERTON.2009.5394532.

Garetto, M., Salonidis, T. & Knightly, E. W. (2008). Modeling per-flow throughput and capturing starvation in csma multi-hop wireless networks. *Ieee/acm transactions on networking (ton)*, 16(4), 864–877.

Gibbens, R., Kelly, F., Cope, G. & Whitehead, M. (1991). Coalitions in the international network. *Teletraffic and data tra c in a period of change, proc. 13th int. teletraffic congress (jensen and iversen eds), amsterdam: Elsevier*, pp. 93–98.

Gizelis, C. A. & Vergados, D. D. (2011). A survey of pricing schemes in wireless networks. *Ieee communications surveys tutorials*, 13(1), 126-145. doi: 10.1109/SURV.2011.060710.00028.

Ha, S., Joe-Wong, C., Sen, S. & Chiang, M. (2012a). Pricing by timing: innovating broadband data plans. *Spie opto*, pp. 82820D–82820D.

Ha, S., Sen, S., Joe-Wong, C., Im, Y. & Chiang, M. (2012b). Tube: time-dependent pricing for mobile data. *Acm sigcomm computer communication review*, 42(4), 247–258.

Hall, R. E. & Lieberman, M. (2012). *Microeconomics: Principles and applications*. Cengage Learning.

Han, Z. & Liu, K. R. (2008). *Resource allocation for wireless networks: basics, techniques, and applications*. Cambridge university press.

Han, Z. & Poor, H. (2009). Coalition games with cooperative transmission: a cure for the curse of boundary nodes in selfish packet-forwarding wireless networks. *Communications, ieee transactions on*, 57(1), 203-213. doi: 10.1109/TCOMM.2009.0901.060661.

Hart, S. & Kurz, M. (1983). Endogenous formation of coalitions. *Econometrica: Journal of the econometric society*, 1047–1064.

He, L. & Walrand, J. (2003). Pricing internet services with multiple providers. *Proceedings of the annual allerton conference on communication control and computing*, 41(1), 140–149.

Heidarpour, B., Dziong, Z. & Lau, W. C. (May 2017). Volume-based pricing with guaranteed data rate. *Accepted in ieee infocom workshop on smart data pricing*.

Hou, J., Yang, J. & Papavassiliou, S. (2002). Integration of pricing with call admission control to meet qos requirements in cellular networks. *Ieee transactions on parallel and distributed systems*, 13(9), 898-910. doi: 10.1109/TPDS.2002.1036064.

Huang, C.-I. (2008). Estimating demand for cellular phone service under nonlinear pricing. *Quantitative marketing and economics*, 6(4), 371–413.

Ismail, M., Abdrabou, A. & Zhuang, W. (2013). Cooperative decentralized resource allocation in heterogeneous wireless access medium. *Wireless communications, ieee transactions on*, 12(2), 714–724.

Jang, J. & Lee, K. B. (2003). Transmit power adaptation for multiuser ofdm systems. *Selected areas in communications, ieee journal on*, 21(2), 171–178.

Jia, J. & Zhang, Q. (2008). Competitions and dynamics of duopoly wireless service providers in dynamic spectrum market. *Proceedings of the 9th acm international symposium on mobile ad hoc networking and computing*, pp. 313–322.

Joe-Wong, C., Ha, S. & Chiang, M. (2015). Sponsoring mobile data: An economic analysis of the impact on users and content providers. *2015 ieee conference on computer communications (infocom)*, pp. 1499–1507.

Kong, H. (2016). Hong kong, government Wi-Fi programme. Consulted at http://www.gov.hk/en/theme/wifi/program/.

Konishi, H. & Ray, D. (2003). Coalition formation as a dynamic process. *Journal of economic theory*, 110(1), 1–41.

Lai, H.-Q., Chen, Y. & Liu, K. R. (2014). Energy efficient cooperative communications using coalition formation games. *Computer networks*, 58, 228 - 238. doi: http://dx.doi.org/10.1016/j.comnet.2013.10.007.

Lazar, A. A. & Semret, N. (1998). The progressive second price auction mechanism for network resource sharing. *8th international symposium on dynamic games, maastricht, the netherlands*.

Liverpool. (2016). Liverpool launches its free Wi-Fi in the city centre. Consulted at http://www.liverpool.nsw.gov.au/services/free-wi-fi.

Lotfi, M. H., Sarkar, S., Sundaresan, K. & Khojastepour, M. A. (2015). The economics of quality sponsored data in non-neutral networks. *arxiv preprint arxiv:1501.02785*.

Lotfi, M. H., Sarkar, S. & Kesidis, G. (2016). Is non-neutrality profitable for the stakeholders of the internet market? *arxiv preprint arxiv:1611.10191*.

Low, S. H. & Varaiya, P. P. (1993). A new approach to service provisioning in atm networks. *Ieee/acm trans. netw.*, 1(5), 547–553. doi: 10.1109/90.251913.

MacKie-Mason, J. K. & Varian, H. R. (1995). Public access to the internet (ch. Pricing the Internet, pp. 269–314). Cambridge, MA, USA: MIT Press. Consulted at http://dl.acm.org/citation.cfm?id=212764.212778.

Maillé, P. & Tuffin, B. (2010). Price war in heterogeneous wireless networks. *Computer networks*, 54(13), 2281–2292.

Mathur, S., Sankar, L. & Mandayam, N. B. (2008). Coalitions in cooperative wireless networks. *Selected areas in communications, ieee journal on*, 26(7), 1104–1115.

McKelvey, R. D., McLennan, A. M. & Turocy, T. L. (2013). Gambit: Software tools for game theory.

Michalak, T., Tyrowicz, J., McBurney, P. & Wooldridge, M. (2009). Exogenous coalition formation in the e-marketplace based on geographical proximity. *Electronic commerce research and applications*, 8(4), 203–223.

Michiardi, P. & Molva, R. (2005). Analysis of coalition formation and cooperation strategies in mobile ad hoc networks. *Ad hoc networks*, 3(2), 193 - 219. doi: http://dx.doi.org/10.1016/j.adhoc.2004.07.006. Ad Hoc Networking for Pervasive Systems.

Mogensen, P., Na, W., Kovacs, I. Z., Frederiksen, F., Pokhariyal, A., Pedersen, K. I., Kolding, T., Hugl, K. & Kuusela, M. (2007, April). Lte capacity compared to the shannon bound. *2007 ieee 65th vehicular technology conference - vtc2007-spring*, pp. 1234-1238. doi: 10.1109/VETECS.2007.260.

Montreal. (2016). MtlWiFi. Consulted at http://ville.montreal.qc.ca/portal/page?_pageid=5977,142099204&_dad=portal&_schema=PORTAL.

Murphy, J., Murphy, L. & Posner, E. C. (1994). Distributed pricing for embedded atm networks. *Proc. international teletraffic congress itc-14, antibes, france*, pp. 1053–1063.

Musacchio, J., Schwartz, G. & Walrand, J. (2011). 18 network economics: neutrality, competition, and service differentiation. *Next-generation internet: Architectures and protocols*, 378.

Myerson, R. B. (1991). Game theory: analysis of conflict. *Harvard university*.

Nash Jr, J. F. (1950). The bargaining problem. *Econometrica: Journal of the econometric society*, 155–162.

Neumark, S. (2014). *Solution of cubic and quartic equations*. Elsevier.

Nisan, N., Roughgarden, T., Tardos, E. & Vazirani, V. V. (2007). *Algorithmic game theory*. New York, NY, USA: Cambridge University Press.

Niyato, D. & Hossain, E. (2006). A cooperative game framework for bandwidth allocation in 4G heterogeneous wireless networks. *Communications, 2006. icc '06. ieee international conference on*, 9, 4357-4362. doi: 10.1109/ICC.2006.255766.

Niyato, D. & Hossain, E. (2009). Dynamics of network selection in heterogeneous wireless networks: an evolutionary game approach. *Vehicular technology, ieee transactions on*, 58(4), 2008–2017.

Osborne, M. J. & Rubinstein, A. (1994). *A course in game theory*. MIT Press.

Pantisano, F., Bennis, M., Saad, W., Verdone, R. & Latva-aho, M. (2011). Coalition formation games for femtocell interference management: A recursive core approach. *Wireless communications and networking conference (wcnc), 2011 ieee*, pp. 1161-1166. doi: 10.1109/WCNC.2011.5779295.

Rahwan, T., Michalak, T. P., Jennings, N. R., Wooldridge, M. & McBurney, P. (2009). Coalition structure generation in multi-agent systems with positive and negative externalities. *Ijcai*, 9, 257–263.

Ray, D. & Vohra, R. (1999). A theory of endogenous coalition structures. *Games and economic behavior*, 26(2), 286–336.

Rogers. (2016). Rogers data plans. Consulted at http://www.rogers.com/consumer/wireless.

Rubinstein, A. & Osborne, M. (1990). Bargaining and markets. Academic Press California.

Saad, W., Han, Z., Basar, T., Debbah, M. & Hjorungnes, A. (2009a). A selfish approach to coalition formation among unmanned air vehicles in wireless networks. *Game theory for networks, 2009. gamenets '09. international conference on*, pp. 259-267. doi: 10.1109/GAMENETS.2009.5137409.

Saad, W., Han, Z., Debbah, M. & Hjorungnes, A. (2009b). A distributed coalition formation framework for fair user cooperation in wireless networks. *Wireless communications, ieee transactions on*, 8(9), 4580-4593. doi: 10.1109/TWC.2009.080522.

Saad, W., Han, Z., Debbah, M. & Hjørungnes, A. Coalitional game theory for communication networks: A tutorial. *Ieee signal processing magazine*.

Sandholm, T., Larson, K., Andersson, M., Shehory, O. & Tohmé, F. (1999). Coalition structure generation with worst case guarantees. *Artificial intelligence*, 111(1–2), 209 - 238. doi: http://dx.doi.org/10.1016/S0004-3702(99)00036-3.

Sarkar, S., Singh, C. & Kumar, A. (2008). A coalitional game model for spectrum pooling in wireless data access networks. *Information theory and applications workshop, 2008*, pp. 310–319.

Seim, K. & Vaiard, V. B. (2011). The effect of market structure on cellular technology adoption and pricing. *American economic journal: Microeconomics*, 3(2), 221–251.

Sen, S., Joe-Wong, C., Ha, S. & Chiang, M. (2013a). Smart data pricing (SDP): Economic solutions to network congestion. *Recent advances in networking*, 1, 221–274.

Sen, S., Joe-Wong, C., Ha, S. & Chiang, M. (2013b). A survey of smart data pricing: Past proposals, current plans, and future trends. *Acm computing surveys (csur)*, 46(2), 15.

Sengupta, S., Chatterjee, M. & Ganguly, S. (2007). An economic framework for spectrum allocation and service pricing with competitive wireless service providers. *New frontiers in dynamic spectrum access networks, 2007. dyspan 2007. 2nd ieee international symposium on*, pp. 89–98.

Sesia, S., Toufik, I. & Baker, M. (2009). *Lte: the umts long term evolution*. Wiley Online Library.

s.Ferguson, T. (2007). Games in coalitional form.

Shakkottai, S. & Srikant, R. (2006). Economics of network pricing with multiple isps. *Ieee/acm transactions on networking (ton)*, 14(6), 1233–1245.

Shapley, L. S. (1952). *A value for n-person games*.

Shenoy, P. P. (1979). On coalition formation: a game-theoretical approach. *International journal of game theory*, 8(3), 133–164.

Singh, C., Sarkar, S. & Aram, A. (2011). Provider–customer coalitional games. *Networking, ieee/acm transactions on*, 19(5), 1528–1542.

Singh, C., Sarkar, S., Aram, A. & Kumar, A. (2012a). Cooperative profit sharing in coalition-based resource allocation in wireless networks. *Ieee/acm transactions on networking (ton)*, 20(1), 69–83.

Singh, C., Sarkar, S., Aram, A. & Kumar, A. (2012b). Cooperative profit sharing in coalition-based resource allocation in wireless networks. *Ieee/acm transactions on networking (ton)*, 20(1), 69–83.

Spagna, S., Liebsch, M., Baldessari, R., Niccolini, S., Schmid, S., Garroppo, R., Ozawa, K. & Awano, J. (2013). Design principles of an operator-owned highly distributed content delivery network. *Ieee communications magazine*, 51(4), 132–140.

Sprint. (2016). Sprint data plans. Consulted at https://www.sprint.com/shop/plan-wall/#!/?plan=individual.

Subbiah, B. & Uzmi, Z. A. (2001). Content aware networking in the internet: issues and challenges. *Communications, 2001. icc 2001. ieee international conference on*, 4, 1310–1315.

T-Mobile. (2016). T-Mobile Binge On. Consulted at http://www.t-mobile.com/offer/binge-on-streaming-video.html.

Tan, C., Chuah, T. & Tan, S. (2014). Subcarrier and power allocation for ofdma-based multicast cellular networks using a coalitional game. *Computer networks*, 64, 180 - 194. doi: http://dx.doi.org/10.1016/j.comnet.2014.02.012.

Telus. (2016). Telus data plans. Consulted at http://www.telus.com/en/qc/mobility/plans/.

Tijms, H. C. (2003). *A first course in stochastic models*. John Wiley and Sons.

Toronto. (2016). Wireless toronto. Consulted at http://wirelesstoronto.ca/.

Verizon. (2016). Verzion, introducing FreeBee data. Consulted at http://freebee.verizonwireless.com/business/freebeedata.

Verizon-Data. (2016). Verizon data plans. Consulted at http://www.verizonwireless.com/landingpages/cell-phone-plans/.

Viterbo, E. & Chiasserini, C. F. (2001). Dynamic pricing in wireless networks. *International symposium on telecommunications*, 5(2), 385–388.

Wallenius, E. & Hämäläinen, T. (2002). Pricing model for 3G/4G networks. *Personal, indoor and mobile radio communications, 2002. the 13th ieee international symposium on*, 1, 187–191.

Wright, J. (2002). Access pricing under competition: an application to cellular networks. *The journal of industrial economics*, 50(3), 289–315.

Xiao, Y. (2005). Ieee 802.11n: enhancements for higher throughput in wireless lans. *Wireless communications, ieee*, 12(6), 82-91. doi: 10.1109/MWC.2005.1561948.

Yi, S.-S. (1997). Stable coalition structures with externalities. *Games and economic behavior*, 20(2), 201–237.

Yilmaz, O. & Chen, R. (2009). Utilizing call admission control for pricing optimization of multiple service classes in wireless cellular networks. *Computer communications*, 32(2), 317–323.

Yin, H., Sun, Y., Cui, B., Hu, Z. & Chen, L. (2013). Lcars: a location-content-aware recommender system. *Proceedings of the 19th acm sigkdd international conference on knowledge discovery and data mining*, pp. 221–229.

York, N. (2016). LinkNYC. Consulted at https://www.link.nyc/.

Youtube. (2016). Youtube Statistics. Consulted at https://www.youtube.com/yt/press/statistics.html.

Zhang, L., Wu, W. & Wang, D. (2014). Time-dependent pricing in wireless data networks: Flat-rate vs. usage-based schemes. *Proc. of ieee infocom*.

Zhang, T., Chen, W., Han, Z. & Cao, Z. (2013). Coalitional game theoretic approach for cooperative transmission in vehicular networks. *Communications (icc), 2013 ieee international conference on*, pp. 6179–6183.

Zhou, H., Liu, B., Liu, Y., Zhang, N., Gui, L., Li, Y., Shen, X. & Yu, Q. (2014). A cooperative matching approach for resource management in dynamic spectrum access networks.

Zhu, K., Hossain, E. & Niyato, D. (2014). Pricing, spectrum sharing, and service selection in two-tier small cell networks: A hierarchical dynamic game approach. *Ieee transactions on mobile computing*, 13(8), 1843–1856.