Design of Network Architectures: Role of Game Theory and Economics

by

Nikhil Gopinath Shetty

A dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy
in

Engineering - Electrical Engineering and Computer Sciences

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Jean Walrand, Chair
Professor Pravin Varaiya
Professor John Chuang

Fall 2010

Design of Network Architectures:
Role of Game Theory and Economics

Abstract

Design of Network Architectures:
Role of Game Theory and Economics

by

Nikhil Gopinath Shetty

Doctor of Philosophy in Engineering - Electrical Engineering and Computer Sciences

University of California, Berkeley

Professor Jean Walrand, Chair

The economics of the market that a network architecture enables has a important bearing on its success and eventual adoption. Some of these economic issues are tightly coupled with the design of the network architecture. A poor design could end up making certain markets very difficult to enable, even if they are in the better interest of society. The analysis of these cross-disciplinary problems requires understanding both the technology and the economic aspects. This thesis introduces three major recurring themes in these problems - revenue maximization, welfare maximization and missing markets - and provides enlightening examples for them. It then delves deeper into three problems representative of these three themes and provides a complete analysis and discussion for each of them.

First, the thesis studies user incentives for the adoption of femtocells or home base stations and their resulting impact on network operator revenues. The thesis develops a model of a monopolist network operator who offers the option of macrocell access or macro+femtocell access to a population of users who possess linear valuations for the data throughput. The study compares the revenues from two possible spectrum schemes for femtocell deployment; the split spectrum scheme, where femtocells and macrocells operate on different frequencies and do not interfere, and the common spectrum scheme, where they operate on the same frequencies (partially or fully) and interfere. The results suggest that the common spectrum scheme that creates heavy interference for the macrocell still performs comparably to the split spectrum scheme for revenue maximization. This suggests that the common spectrum scheme with good interference management may be the pathway to better femtocell adoption.

Second, the thesis investigates the impact of the provision of two classes of service in the Internet on the surplus distribution between users and providers. The study considers multiple competing Internet Service Providers (ISPs) who offer network access to a fixed user base, consisting of end-users who differ in their quality requirements and willingness to pay for the access. User-ISP interactions are modeled as a game in which each ISP makes capacity and pricing decisions to maximize his profits and the end-users only decide which service to buy (if any) and from which ISP. The model provides pricing for networks with

single- and two-service classes for any number of competing ISPs. The results indicate that multiple service classes are socially desirable, but could be blocked due to the unfavorable distributional consequences that it inflicts on the existing Internet users. The research proposes a simple regulatory tool to alleviate the political economic constraints and thus make multiple service classes in the Internet feasible.

The third topic is a problem involving missing markets for cyber-security insurance. The study explains why insurance markets for Internet security fail to take off due to a number of factors including information asymmetry, efficiency losses due to network externalities and competition. The interdependent nature of security on the Internet causes a negative externality that results in under-investment in technology-based defences. The research investigates how competitive cyber-insurers affect network security and user welfare. The model explores a general setting, where the network is populated by identical users with arbitrary risk-aversion and network security is costly for the users. The user's probability to incur damage (from being attacked) depends on both his security and the network security. First, the model considers cyber-insurers who cannot observe (and thus, affect) individual user security. This asymmetric information causes moral hazard. If an equilibrium exists, network security is always worse relative to the no-insurance equilibrium. Though user utility may rise due to a coverage of risks, total costs to society go up due to higher network insecurity. Second, the study considers insurers with full information about their users' security. Here, user security is perfectly enforceable (zero cost). Each insurance contract stipulates the required user security and covers the entire user damage. Still, for a significant range of parameters, network security worsens relative to the no-insurance equilibrium. Thus, although cyber-insurance improves user welfare, in general, competitive cyber-insurers may fail to improve network security.

To dearest Amma, Mummy, Papa and Shonu.

Contents

List of Figures

List of Tables

Acknowledgements

I would like to take this opportunity to thank all the people who have been a contributing factor both to this dissertation and to my life as a PhD student. First, I would like to extend my deep gratitude to my advisor, Jean Walrand, whose guiding thoughts are responsible for shaping my interest in the topic of this dissertation. Over the course of the PhD, he has always been available for advice and his advice has always enabled me to determine the right path. I appreciate very much his approach of letting his students explore but at the same time always being there to correct our missteps. This has afforded me the right amount of freedom necessary to explore the interdisciplinary nature of the dissertation while at the same time being grounded and producing results at a regular pace.

I would like to thank Professors Pravin Varaiya, John Chuang and Shachar Kariv for advising my research proposal and final dissertation. I would also like to thank all the other professors, who have taught me during my stay as a PhD student at Berkeley. In particular, I would like to mention Prof. David Tse's wireless communications class and Prof Reza Moazzami's classes in the Haas school of Business, which I found to be great learning experiences. I hope all my professors found teaching me as interesting as I found learning from them.

I would like to thank everyone who have collaborated with me for all my research projects. Here, I would like to thank Dr. Shyam Parekh, who has been an excellent support for me during my research career. He helped set up internships at Alcatel-Lucent Bell Labs, which were great learning experiences. I also thank Galina Schwartz, who collaborated on multiple projects including QoS and security economics. She has been a wonderful research partner and has helped me navigate the large and complex field of economics and game theory. Mark Felegyhazi was also very influential during the research on security economics and I would like to thank him for that. I express gratitude to Assane Gueye, with whom I had a great collaboration effort during the bottleneck analysis research. We shared great chemistry and believe we both missed our collaboration effort. Without Sofie Pollin and Przemek Pawelczak, I would not have been able to complete our work on machine learning in cognitive radio. I thank them for this partnership. Lastly, research on IPv6 adoption was another interesting collaboration effort, which I thank Dilip Joseph for involving me.

I would also like to thank all other members of the NetEcon group including Jiwoong Lee, Libin Jiang, Christophe Choumert, John Musacchio for their helpful comments and suggestions. Thanks are due for alumni like Wilson, Antonis and Teresa, who led Jean's research before our current batch. Thanks are due also for Ruth Gjerde for helping me navigate the bureaucratic requirements at Berkeley. Last, I express my gratitude to Koushik Pal, Vinayak Nagpal, Sharmodeep Bhattacharya, Vikram Vijay, Nipun Misra, Vikram Rentala, Pulkit Grover, Marghoob Mohiyuddin, and others for contributing to the excellent 5 years of my life at Berkeley.

Chapter 1

Introduction

Networking researchers have long known and argued [9, 66, 24, 39] that the economics of the markets enabled by their network architecture have a huge bearing on its success and eventual adoption. Some of these economic issues are tightly coupled with their design of the network architecture. A poor design could end up making certain markets very difficult to enable even though enabling such markets could be in the better interest of society. In many cases, even when the technologies are mature and well-researched, adoption is either very slow or just plain fails in the face of the economics. Examples of such failures include Asynchronous Transfer Mode (ATM)[1], the slow adoption of IPv6[2] and end-to-end quality of service on the Internet.

Studying how a change in architecture affects user and provider welfare is important because this analysis ultimately determines the success or failure of a new invention. This analysis requires an understanding of both the technology and the resulting economics of the market enabled by it. On one hand, the performance offered by a new architecture can be analyzed using tools like queueing theory, systems theory, etc. This analysis provides a sense of the utility that users derive from the usage of that architecture. On the other hand, problems involving incentives are usually handled well using tools from game theory. These tools provide us the ability to analyze the welfare obtained from the adoption of a certain architecture, taking into consideration that users and providers may be independent decision makers. It is this confluence of technology and game theory that this dissertation aims to study.

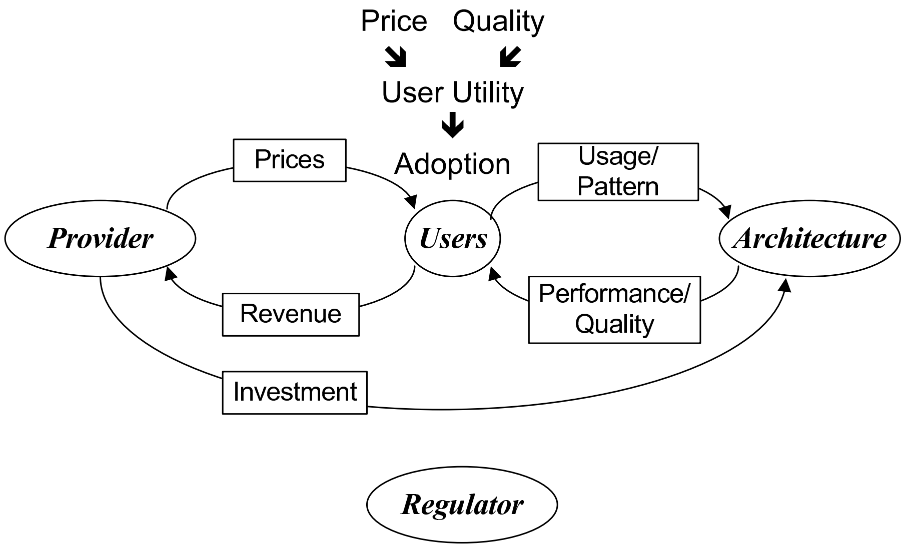


Figure 1.1. Framework

1.1 Framework

Let us first start by introducing a general framework into which most of the problems can be placed. Fig. 1.1 represents the block diagram into which most problems can be cast [99]. Network providers invest in their chosen network architecture and then charge prices to users. Users observe the prices and choose to adopt the services, if any. This usage of the network may affect the performance that is observed by the users. This combination of prices and performance determines the utility observed by a user which, in turn, determines user adoption and associated provider revenues. In addition, providers may compete or need to coordinate with each other to provide certain services - this may also determine the investment, equilibrium prices and user utility. The external agent in this framework is the government or the regulator whose concern it is to maximize the welfare of the system as a whole and to ensure that innovation continues unimpeded. The government may also have special powers to mandate certain behaviors from the providers which may help eliminate certain coordination problems, improve social welfare and create a favorable distribution of the surplus.

To see an example of how this framework applies, let us look at the case of provision of multiple service classes in the Internet. Internet service providers invest in an architecture that enables QoS - this could be DiffServ, IntServ or some other new architecture [100]. The providers then charge prices for the different service classes and users choose which service class to use, if any. The observed user performance depends upon their network usage and the

[1] An internetworking standard that was a competitor to IP.
[2] IPv6 adoption is still very slow though IPv4 addresses are close to exhaustion.

policies implemented by the ISPs. Users may now observe this performance and determine whether they are satisfied with the quality of the service obtained. If not, they may switch to a different service from any of the ISPs or choose not to take up any service at all. In equilibrium, no user will want to shift to a different service option and no ISP will want to charge different prices. A regulator observing this situation might want to set policies regarding such multiple classes of service to suitably protect user welfare. His policy must also consider whether future ISP investment will be impacted which in turn could affect the long term utility of the users. Alternatively, if ISPs cite coordination issues due to a lack of common standards as a reason for no such market, the regulator may step in to encourage such coordination using a mandate.

The framework thus provides us with a clear mental picture of the problem. It enables us to separate out the technological and economic parts of the problem and attack each portion separately. The framework also gives us an insight into how the various parameters interact with each other to determine the final equilibrium.

1.2 Research Problems

There are many interesting research problems at this interdisciplinary intersection. In this section, I will describe a few aspects that show up often in these problems. Most problems will incorporate multiple aspects from those described below.

1.2.1 Revenue Maximization

Maximizing revenues is the concern of network providers. Providers want a good return on their investment and they choose an architecture that allows them to maximize their profits. Though this may seem like a simple cost-benefit analysis, in many cases, calculation of costs and benefits is complicated by the fact that users may create negative externalities for other users by congestion. A lot of research has been performed on how this congestion pricing must be performed. For example, [70, 39] suggest that multiple service classes must be simply isolated and the price charged to users must determine the congestion and the performance in the class. [6] also performed experiments on real Internet users to determine how demand changes with the price. [54, 17] suggest changes in the congestion control mechanism to charge for packets marked during periods of congestion. On the other hand, some recent work [88] suggests that the monetary gains from multiple service classes may not be large enough for certain user demands to justify investing in a QoS architecture. These examples suggest that revenue maximization is a complex process requiring knowledge of both user demand and the choice of an appropriate architecture to implement congestion control and QoS.

Managing congestion and pricing it appropriately is becoming extremely important in today's wireless networks, where bandwidth is scarce and supporting heavy data usage requires massive investments in the infrastructure. Wireless congestion management is also complicated by the fact that channel quality depends upon the user's position in the cell and can vary rapidly. How should resources be allocated to different users and how should the users be charged for those resources become interesting questions in such a dynamic scenario.

1.2.2 Welfare Maximization

Welfare maximization usually takes the perspective of maximization of total system utility, including both the providers and end users of a technology. Welfare increases when all user types and requirements are satisfied by the adoption of a certain architecture. If the architecture prevents serving a certain essential user requirement due to the way it is structured, this results in lowered welfare. For example, researchers [82] argue that the lack of an in-built per-flow charging mechanism in the Internet may have led to a missing market for enhanced quality of service guarantees at a flow level. Though the increase in network capacity has allowed Voice over IP (VoIP) services to take off without this per-flow requirement (though not with assured guarantees), video conferencing is now facing the same issues that voice did a few years back. Though enterprises have taken care of this issue by buying reserved bandwidth, such services are hardly available to end users. I believe that there will always be applications on the horizon that cannot be handled purely using the statistical guarantees provided by the Internet and will require some guaranteed services.

Efficiency Losses

The operating point that maximizes total system welfare is called the social optimum. Such a social optimum can only be reached by a central planner who can enforce the behavior of all the entities in the system. However, very rarely do large systems have such centralized decision-making. Behavioral decisions are usually made by entities with their own selfish interest in mind. If such behavior is associated with externalities, i.e., if one user's behavior affects the utility observed by others in the system, these selfish choices result in an equilibrium that is much worse than the social optimum. This may be interpreted as a loss of efficiency or, when rephrased in terms of costs, is also known as the "price of anarchy" [55]. [51] have derived the worst case loss of efficiency for certain resource allocation games in networks where agents make decentralized decisions . Similarly, [78] have derived the price of anarchy in routing games, where routing decisions are made independently by users. Interestingly, under linear utility/cost assumptions, both these papers suggest that efficiency losses are bounded by a factor of 3/4.[3]

Another interesting example of efficiency losses is the case of Internet security. Computer security is interdependent and when users under-invest in security, this creates a relatively more insecure environment for the other users. Such users may not necessarily see the direct

[3] In [51], this bound assumes that the users can bid on individual links. If they bid on a path, the POA can be arbitrarily large.

costs of such risky behavior though they enjoy the lower costs associated with their lower investment. When all users apply the same logic, the result in under-investment in security throughout the network relative to the socially optimal level [97]. This is also referred to sometimes as the free-riding problem. Similar results have been derived by [62, 49] under constrained network topologies and corresponding externalities. In Chapter 4, I will present a model for network security with full connectivity and a general user utility and replicate the results mentioned here.

Surplus Distribution

At times, just maximizing total social welfare may not be enough. Regulators may also be interested in how the surplus is actually distributed between the participants. For example, in a fiercely competitive scenario, a large part of the surplus remains with the users and providers may end up with zero profits (no surplus). This explains why regulators encourage policies that foster competition in provider markets. Other times, when new technologies enable providers to extract more surplus from users, regulators may be concerned about the backlash this may generate from such losing users. Though the new technologies may also expand the market and increase availability, regulators may still be concerned about the effects of the transition on these existing users. An example of this case is the discussion regarding network neutrality where an important part of the debate is a consideration whether two classes of service create distortions in the distribution of surplus between users and providers. In this case, the regulator needs to balance the protection of current user surplus with the future incentives to invest in greater capacity. This example is analyzed in detail in Chapter 3.

1.2.3 Missing Markets

Missing markets are markets that increase both user and provider welfare (Pareto) and yet, fail to take off in reality. Such missing markets are a result of many factors, a few of which we will discuss here.

Information Asymmetry

Information asymmetry exists when information regarding quality of the service or product is not available equally well to two entities involved in a transaction. Information asymmetry makes contracts difficult to enforce, resulting in a collapse of markets. The seminal work in the Economics literature includes the lemon market problem, where it was shown that the market for used cars could collapse if lemons could not be suitably distinguished from the good cars [5]. Similar problems exist in the network and in fact, may be accentuated by the choice of the network architecture. For example, if a certain data path traverses multiple ISPs. With the current IP architecture, if there is a problem with the quality of service experienced by the flow, there is no way of knowing which ISP is the source. Such information is local and private to the ISPs and without some monitoring systems [60], end-to-end quality of service contracts across multiple ISPs will be impossible.

More examples exist in the computer security market. If security products cannot be sufficiently distinguished in their quality of protection, this could result in a lemon market for such products [7]. Similarly, if insurance companies cannot sufficiently gauge how well an enterprise has protected its network from external attacks, problems like moral hazard and adverse selection show up and can result in a collapse of the insurance market [76, 77]. This example of insurance for network security is analyzed in detail in Chapter 4.

Network Effects and Switching Costs

When the utility from adoption of a technology depends upon the number of other users of that technology, we say that network effects are in play [29, 28]. Network effects create lock-in into network architectures. This lock-in can be extremely strong, as can be seen in the poor adoption of IPv6 in the Internet. IPv4 is the principal protocol on the Internet. Suppose a new content provider offers his services on the Internet. Lets say she has a choice of IPv6 or IPv4. Her standalone benefits from adopting IPv6 are not hugely higher than her benefits from adopting IPv4.[4] However, due to being the incumbent standard, her network benefits from adopting IPv4 are huge compared to that from IPv6. If she chooses IPv6, she still has to provide converters for users who access her services using IPv4, which can further raise her costs. Here, the role of converters/gateways from one architecture to another and the standalone benefits of adoption become very important [53, 50]. Also important is a critical mass of early adopters who can push the rest of the market to the new architecture. A regulator could encourage such adoption by mandating government installations and offices to adopt the new architecture, thus providing this critical mass.

A similar example exists for the transmission control protocol TCP. Though newer flavors of TCP have been suggested that provide improvements in throughput for fat pipes [101] or rely not on network losses but on latency to adjust rates [16], these attempts have failed at adoption because either they do not work well in the presence of TCP Reno or they impact TCP Reno flows negatively, thus inviting the network provider's wrath. These network effects are so huge that researchers have proposed "TCP-friendly" rate control algorithms with the goal to improve TCP performance while still maintaining fairness when combined with the usual TCP flows [32].

Transaction Costs

In his seminal work on transaction costs, economist Ronald Coase [23] suggests that all economic transactions involve costs and these costs must be accounted for in analyzing

[4] IPv6 does offer some improvements in security and user management.

whether a particular economic arrangement is efficient.[5] Transaction costs can be both external and internal to a firm. Here, we are only concerned with external transaction costs that come into play when a firm interacts with other firms to deliver a product or service. These transaction costs involve the cost of acquiring the service information from the partners, cost of writing up complicated service contracts involving a large number of variables and the costs of monitoring and enforcing those contracts. If the costs of performing these activities are higher than the profits made by providing the service, firms may decide not to offer such a service at all, resulting in a missing market. Certain network architectures may raise these enforcement costs and make certain markets very difficult to enable. In the following paragraphs, I will provide examples of manifestations of this problem of transaction costs.

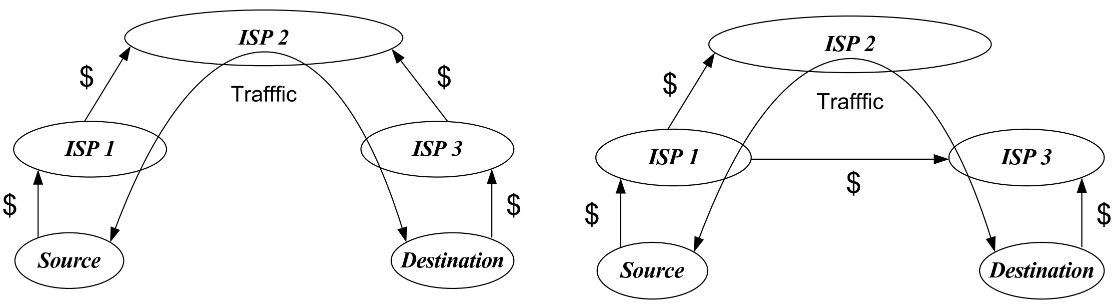
Coordination:
Coordination in network upgrades is a major problem that providers face. Usually, this problem arises because there are positive externalities associated with the upgrade process of a single provider. When a single operator upgrades his network, he may indirectly provide better performance for other operators' traffic flowing through his network. This results in the free-rider problem, wherein the incentive to upgrade for his rival decreases, especially if the cost of upgrade is high. This may result in a contest, where operators wait for others to adopt the new technology, thus delaying adoption for inordinately long amounts of time. [48] investigated the conditions under which ISPs upgrade their networks by setting up a repeated game model for the value derived from a network upgrade. Such coordination problems may require external stimulus by providing some mechanism by which ISPs can coordinate their network investments. This could involve government mandates or schedules set by standard bodies for network upgrades.

Competition:
Sometimes, competitive pressures may preclude investments in new technologies. A good example of this may be network virtualization [8, 20, 68]. Network virtualization enables operators to experiment and program new architectures in software over an existing infrastructure, thus reducing the need for frequent network upgrades. However, virtualized network equipment may come at a higher cost or lower performance compared to equipment that cannot support future architectures but handles all existing architectures very well.[6] Though a virtualized architecture is obviously better for society because of the accelerated innovation ability it provides, it is not clear that a provider running virtualized architecture is in a better competitive position to one who does not. Though such an operator can bring innovations to market quicker, the higher costs of operating such an architecture may negate the potential benefits from being the first to market. Additionally, there is also the risk associated with betting on the success of future innovations.[7] In this case, the coordination mechanism to ensure that all competitive operators invest in such a virtualized architecture may be too costly to implement.

[5] Other interesting readings include [103, 69, 104]

[6] Hardware implementations are usually much faster than software ones.

[7] This is work in progress. The author and his collaborators have pinned down the rate of new successful innovations and the rate of network upgrades as important factors underlining the decision to choose a virtualized infrastructure.



(a) IP Market Structure (b) Telephone Market Structure

Figure 1.2. Market Structure

Market Structure:
Market structure, that is a result of the network architecture, may also play a big role in the adoption of new services. An example is the current market structure of the Internet[11, 92, 95]. The structure of contracts is bilateral, i.e., ISPs set up traffic contracts with their neighbors only [60]. In addition, "access" ISPs always pay "transit" ISPs for traffic flowing in either direction - they are paying for connectivity to the "rest of the Internet." This is different from telephone networks where contracts are source-based, i.e., the source sets up contracts with each of the networks that the traffic lands up on and termination fees are paid to each of these networks. Some argue that this modification keeps contracts simple and local [89]. But, this market structure only works well for best-effort traffic where quality requirements are lax. However, for guaranteed services, this market structure creates incentive problems. Fig. 1.2 shows this difference in payment structure. ISP 3, on whom high-priority traffic is terminated, still has to pay for this traffic (maybe higher price) even though no session-specific revenues are shared with him. Also, if the end users do not pay for any differentiated services, there may be no incentive for him to give better service to this flow. He may just instruct ISP 2 to remark all packets as normal priority.

Another interesting problem related to market structure is the role it plays in innovation. [61] investigate the resultant market structure from a virtualized network, where infrastructure providers and service providers are different entities. As an example, consider the cable TV providers. This can be envisioned as an infrastructure provider who manages the cable system and a service provider who provides TV service over this system. In this two-level system, users may end up paying the infrastructure provider or the service provider or both. [61] concludes that, since infrastructure providers are few, a market structure where the user pays directly to the service provider will be most competitive and encourage newer innovations both in the network and in the service layer. In fact, they believe that a system where payments happen through the infrastructure provider alone will not generate the necessary innovation incentive. [61] only look at independently generated innovations - it would be interesting to investigate how their results change when network providers need to coordinate and move to new architectural standards enmasse and not independently.

1.3 Organization

In this dissertation, I will attack three problems, each focusing mainly on one of the aspects discussed above. First, I will investigate a problem involving revenue maximization that arises with the adoption of femtocells[8] for 4G networks. This will be the topic of Chapter 2. It is expected that with heavy network usage, 4G networks will require some offloading of data traffic onto femtocells. Here, the operators have a choice of two possible spectrum schemes for femtocell deployment; the split spectrum scheme, where femtocells and macrocells operate on different frequencies and do not interfere, and, the common spectrum scheme, where they operate on the same frequencies (partially or fully) and interfere. We model a monopolist network operator who offers the option of macrocell access or macro+femtocell access to a population of users who possess linear valuations for the data throughput and compare the revenues from the two schemes. The results will suggest that common spectrum schemes that create heavy interference for the macrocell still perform comparably to the split spectrum scheme for revenue maximization. This suggests that common spectrum schemes with good interference management may be the pathway to better femtocell adoption.

Second, I will investigate a problem involving welfare maximization where I estimate the effect on surplus distribution between users and providers from provision of two classes of service in the Internet. This problem has relevance to the much-larger debate on network neutrality. This will be the topic of Chapter 3. In this case, I take the view of a regulator to determine under what conditions multiple service classes should be permitted. Here, I consider multiple competing Internet Service Providers (ISPs) who offer network access to a fixed user base, consisting of end-users who differ in their quality requirements and willingness to pay for the access. ISPs make capacity and pricing decisions to maximize their profits and the end-users only decide which service to buy (if any) and from which ISP. I will show that competition per se does not preclude the provision of service classes, i.e., there is no missing market due to competitive pricing instability. Though the results indicate that multiple service classes are also socially desirable in the long run, they could be blocked due to the unfavorable distributional consequences that it inflicts on some existing Internet users and the short run considerations of a regulator.

Third, I will investigate a problem involving missing markets. Here, I will look at why insurance markets for Internet security fail to take off due to a number of factors including information asymmetry, efficiency losses due to network externalities and competition. This will be the topic of Chapter 4. I will utilize a general setting, where the network is populated by identical users with arbitrary risk-aversion and network security is costly for the users. A user's probability to incur damage (from being attacked) depends on both his security and the network security, thus creating an externality. Thus, there are efficiency losses and network security is worse than the socially optimal level. Total costs to society go up due to this higher network insecurity. However, even with competitive security insurance, the situation is no better. There are two cases - case with information asymmetry, i.e., lack of an enforceable security contract, between the insurer and the insured and the case without.

[8] Femtocells or home base stations are a proposed solution to the problem of degraded indoor service from the macrocell base station in future 4G data networks.

With information asymmetry, an equilibrium rarely exists, i.e., there is a missing market. Even when it exists, network security is always worse relative to the no-insurance equilibrium. This suggests that the architecture must provide strong enforcement mechanisms that reduce this asymmetry. With no asymmetric information, an equilibrium exists but network security still worsens relative to the no-insurance equilibrium for a range of parameters. Thus, the results suggest competitive cyber-insurers may fail to improve network security, which implies that an enforceable mechanism is required to ensure that cyber-insurers maintain high network security even in the face of competitive pressures.

Chapter 2

Revenue Maximization - Femtocells

4G networks, especially those operating at high frequencies, are expected to face the problem of poor connectivity inside the users' home. This is mainly due to the high attenuation suffered at these frequencies. To circumvent poor reception inside such built-up areas, the industry has proposed tiny base stations for homes called femtocells [73, 46, 18, 74].[1] These femtocells not only enable high-quality use of mobile devices in the user's home but also allow the user to seamlessly move his calls and data sessions between the macrocell and his femtocell.[2]

From the point of view of the network operator, femtocells appear advantageous since femtocell usage reduces the load on the macrocell network and allows more users to be served, which helps raise revenues. In addition, network operators may be able to price discriminate and extract a higher value from femtocell users. However, an operator's use of femtocells is not devoid of costs. In this chapter, we do not consider an increase in operational costs (like the additional costs of managing an integrated macro-femto network, customer support, etc.) due to the provision of femtocells. We only consider the opportunity costs of the network operator due to the (in)efficient use of spectrum in the hybrid macro-femto network. Note that these opportunity costs would not exist if femtocells were to operate in free spectrum, like those in the 2.4 and 5.8 GHz bands. However, due to the prolific number of devices (that the operator himself does not control) present in these frequencies, no quality of service (QoS) guarantees can be given. To provide QoS, the operator must utilize his own spectrum, adding opportunity costs to his femtocell operations.

Various spectrum deployment options have been proposed for the deployment of femtocells [44]. The authors in [44] suggest 3 possible spectrum schemes - the "separate carrier" deployment where the spectrum is divided into two parts and a dedicated fraction is used for

[1]The previous incarnation, picocells, operated at higher powers and were costlier, making them unsuitable for wide deployment in homes.

[2]We do not consider WiFi hotspots as examples of femtocells since seamless mobility when moving from the macrocell to the WiFi access point has not yet been achieved.

femtocells, the "shared carrier" deployment where the macrocell and the femtocells operate on the same frequencies and the "partially-shared carrier" deployment where the femtocells operate only on a fraction of the spectrum used by the macrocell. In this chapter, we will consider two spectrum schemes only. The first scheme - which we term "split spectrum" - will be similar to the separate carrier scheme. The second scheme - which we term "common spectrum" - will model both the shared carrier and the partially-shared carrier cases.

Both femtocell schemes provide gains via increased macrocell capacity due to lower congestion. However, the two schemes impose different costs on the macrocell. With split spectrum, macrocell capacity is not affected due to interference, but, dedicating spectrum for femtocell usage directly reduces the capacity of the macrocell. With common spectrum, there is no loss due to dedicated capacity, but increased adoption of femtocells leads to increased interference for macrocell users and decreases its capacity. Previous research [21, 43, 74, 19, 65] has focused on how this interference affects macrocell capacity and service quality and have suggested that common spectrum deployments are as feasible as split spectrum ones.

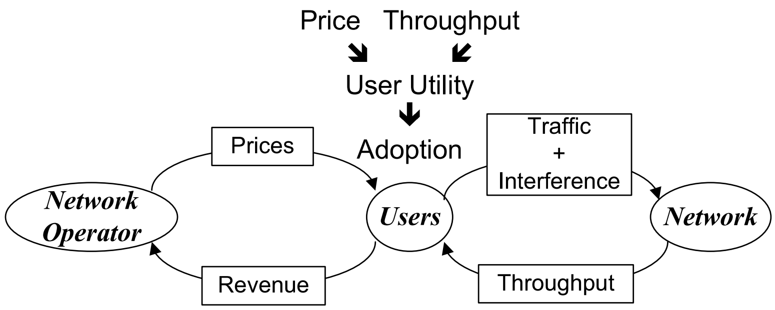


Figure 2.1. High Level Model

The focus of this chapter is to study the impact of the complex interplay of interference and service pricing on user adoption of femtocells. Fig. 2.1 depicts the high-level model that we analyze in this chapter. Analysis of femtocell adoption is not straightforward. With split spectrum, as femtocell adoption increases, the pure macrocell service becomes more attractive to the users. Hence, user incentive to adopt femtocells decreases as femtocell adoption rises. Due to the presence of this positive externality, adoption of femtocells may be hindered. With common spectrum, this effect reappears, but is further complicated by the addition of interference for the macrocell users. In this chapter, we analyze these effects from the point of view of a monopolist network operator's revenues. Under his optimum choice of prices, we will compare the revenues from the competing spectrum schemes. Though some papers [22] have focused on the financial aspects of femtocells, they focus on the network deployment costs only. To our best knowledge, no previous work focuses on determining the network operator's optimal pricing choices and the resulting users' incentives for femtocell adoption.

The rest of the chapter is organized as follows. Section 4.2 describes the model with subsections 2.1.1 and 2.1.2 describing the network model and the interference model respectively. Section 2.2 describes how the network operator revenues are determined. We present the numerical results in Section 2.3 and conclude in Section 4.5.

2.1 Model

Consider a monopolist wireless network operator who offers mobile services to a population of users N. Assume that this operator has a fixed amount of spectrum to deploy. Assuming that the operator wants to deploy femtocells, he has the following two options: deploy femtocells under a split spectrum scheme or under a common spectrum scheme. In both these cases, we will assume that the operator only provides two service options - a mobile-only service m that allows the user to access the macrocell only and a mobile-plus-femto service f that permits the additional usage of a home-based femtocell. Let p_m and p_f be the prices charged for the services m and f respectively. Once the prices are charged, users are free to choose their preferred service, if any. Let the operator's objective be to maximize his revenue V given by:

$$V = p_m X_m + p_f X_f, \tag{2.1}$$

where X_m and X_f are the number of users who adopt services m and f respectively. Also, define

$$X = X_m + X_f,\ x = \frac{X}{N} \text{ and } \alpha = \frac{X_f}{X}. \tag{2.2}$$

Next, we model the user demand for services m and f. Let T_j be the instantaneous data throughput received by a user from service $j = m, f$. We assume that a user derives an instantaneous benefit $\gamma f(T_j)$ from the service where γ represents the user's valuation for this throughput and $f(\cdot)$ is a concave function with $f' > 0$ and $f'' \leq 0$. Further, we assume that the user population consists of users of type $\gamma \in (0, \gamma_{max}]$ and let the cumulative distribution function (cdf) of the user types be Γ, satisfying the usual conditions $\Gamma(\gamma \leq 0) = 0$ and $\Gamma(\gamma \geq \gamma_{max}) = 1$.

The instantaneous throughput T_j varies with the user's position, the time of access, and the congestion in the network (i.e. the access times of other users). We obtain the user's expected benefit $\gamma E\big[f(T_j)\big]$ from adoption of the service j by taking the expectation over all possible user trajectories and network access times. Next, we make the simplifying assumption that $E\big[f(T_j)\big]$ is a function of x and α only and does not depend upon a specific user. Let this dependence be given by the function $g_j(\alpha, x)$ for $j = m, f$. Then, a type γ user's utility from adopting service $j = m, f$ will be

$$U_j^{\gamma} = \gamma g_j(\alpha, x) - p_j. \tag{2.3}$$

For any given x and α, we assume

$$g_f(\alpha, x) > g_m(\alpha, x), \tag{2.4}$$

i.e., a user derives a higher benefit from service f than service m. Hence, if $p_f \le p_m$, $\alpha = 1$, i.e., users choose to buy the service f only, if any. Note that α may be 1 even when p_f is higher than p_m.

From (2.3) and (2.4), for $j = m, f$, if for some $\tilde{\gamma}$, $U_j^{\tilde{\gamma}} > 0$, then, $U_j^{\gamma} > 0$ for all $\gamma > \tilde{\gamma}$. This implies that there is a threshold user type beyond which all users (with a higher valuation for the throughput) adopt some service. Define γ_m to be the critical user type beyond which all users buy some service. Then, the fraction of users who adopt some service will be:

$$x = \big[1 - \Gamma(\gamma_m)\big]. \tag{2.5}$$

Next, suppose $\alpha < 1$. From (2.3) and (2.4), if for some $\tilde{\gamma}$, $U_f^{\tilde{\gamma}} > U_m^{\tilde{\gamma}}$, then, $U_f^{\gamma} > U_m^{\gamma}$ for all $\gamma > \tilde{\gamma}$. This implies that all users with type greater than a certain critical user type adopt service f over service m. Define γ_f to be the critical user type beyond which all customers buy service f. Then, given x, the fraction of customers who adopt service f will be:

$$\alpha = \frac{\big[1 - \Gamma(\gamma_f)\big]}{x}. \tag{2.6}$$

Fig. 2.2 depicts γ_m, γ_f, α and x.

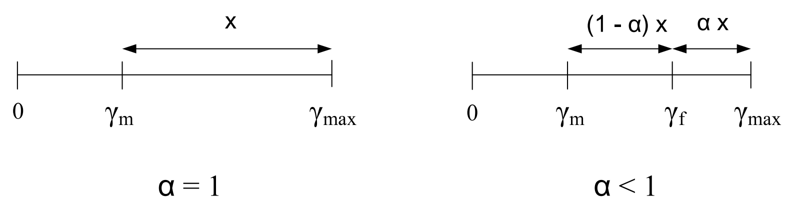


Figure 2.2. Relationship between γ_m, γ_f, α and x

Theorem 2.1.1 *For any given p_m and p_f, the values of x and α in equilibrium are determined from (2.5) and (2.6) with γ_m and γ_f given as below:*

$$\gamma_m = \begin{cases} \frac{p_f}{g_f(1,x)}, & \alpha = 1 \\ \frac{p_m}{g_m(\alpha,x)}, & \alpha < 1, \textit{ and,} \end{cases} \tag{2.7}$$

$$\gamma_f = \frac{p_f - p_m}{g_f(\alpha, x) - g_m(\alpha, x)}, \ \textit{if } \alpha < 1. \tag{2.8}$$

Proof See Appendix.

Corollary 2.1.2 *If $\frac{g_f(\alpha,x)}{g_m(\alpha,x)} \ge \frac{g_f(1,x)}{g_m(1,x)}$ $\forall \alpha < 1$, then, for any x, $\alpha = 1 \Leftrightarrow \frac{p_f}{p_m} \le \frac{g_f(1,x)}{g_m(1,x)}$.*

Proof See Appendix.

From Corollary 2.1.2, in equilibrium, if $\alpha = 1$, the operator must charge $p_f = \frac{g_f(1,x)}{g_m(1,x)} p_m$ to maximize his revenue. Hence, even when $\alpha = 1$ in equilibrium, $p_f > p_m$, i.e., service f will be costlier than service m. Substituting the values of x and α from Theorem 2.1.1 in (2.1) and optimizing w.r.t. p_m and p_f, we obtain the monopolist's optimal choice.

2.1.1 Model for $g(\cdot)$

In this section, we present an approximation to the function g. First, we assume that the user's benefit $f(T)$ is proportional to the instantaneous throughput T that he receives [80]:

$$f(T) = kT,$$

where k is the constant of proportionality. We discuss the implication of this assumption in Section 4.5. Thus, the utility for the user type γ if he adopts service $j = m, f$ becomes

$$U_j^\gamma = \gamma k E[T_j] - p_j,$$

which, from (2.3), gives us

$$g_j(\alpha, x) = kE[T_j], \; j = m, f. \tag{2.9}$$

In the rest of the discussion, we will restrict our attention to the throughput obtained in the downlink only. A similar analysis for the uplink may be performed.

Next, let users spend a fraction $f_i (f_o = 1 - f_i$ and $f_o < f_i)$ of their time inside their home. Then, we assume

$$E[T_f] = f_i E[T_b] + f_o E[T_m], \tag{2.10}$$

where T_b is the throughput obtained by the user from his broadband connection via the femtocell. We assume that $E[T_b]$ is fixed and independent of the user's position in his home, the interference from femtocells and macrocell users in the neighborhood (see Fig. 2.3 for the only interference that is modeled) and the spectrum scheme being employed. $E[T_b]$ depends on the congestion in the wired network and is *not* the maximum supportable data rate of the femtocell. It is conceivable that $E[T_b] > E[T_m]$ for the near future and this gives us our desired condition: $g_m(\alpha, x) < g_f(\alpha, x)$. Note that (2.10) underestimates $E[T_f]$. Since femtocell adopters use macrocell services only when they are outside, they may receive better expected throughput $E[T_m]$ than pure macrocell users. Yet, this impact will be low since $f_i > f_o$ and $E[T_b] > E[T_m]$.

T_m depends upon the position-varying and time-varying channel conditions, interference from the femtocells and the congestion in the network. Analyzing data rate variations due to channel conditions is beyond the scope of this chapter. Hence, to simplify, we define a *macrocell data rate* $R(\alpha x)$, that captures all channel variations and depends only upon the fraction of population that has adopted femtocells. $R(\alpha x)$ can be conceived as an average throughput received from the macrocell if exactly one user were to move around the macrocell, both outside and inside his home, in the presence of αx fraction of femtocells. Note that

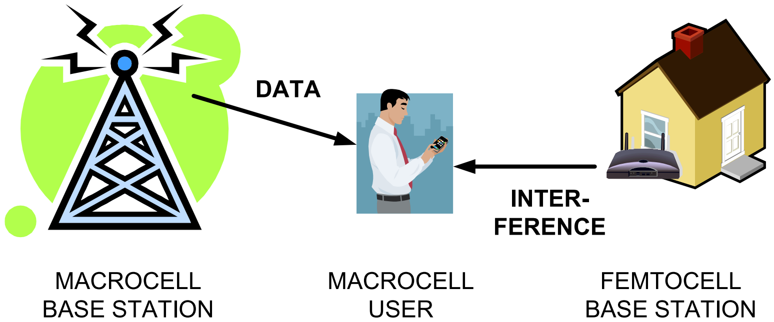


Figure 2.3. Interference

$R(\alpha x)$ depends on the spectrum scheme being employed (see Section 2.1.2). With $R(\alpha x)$ thus defined, the throughput T_m depends upon the congestion in the network only.

To determine $E[T_m]$, we model the congestion in the macrocell network as follows. First, we assume that user population is distributed identically across all cells. Then, if X_{cell} are the adopters and N_{cell} are the number of users in any macrocell, we let $x = \frac{X_{cell}}{N_{cell}} = \frac{X}{N}$. Next, let users generate i.i.d. requests for downloads following a Poisson process of rate λ_o when they are outside their homes and rate λ_i when they are inside. Define the activity ratio β as

$$\beta = \frac{\lambda_o}{\lambda_i}. \tag{2.11}$$

Next, assume that the file lengths are exponentially distributed. Then, if this download is served at rate $R(\alpha x)$, it would take random exponential amount of time of mean $1/\mu$, where

$$\mu = \frac{R(\alpha x)}{\text{Mean File Length}}. \tag{2.12}$$

When l (≥ 0) downloads are simultaneously active, and each download shares the macrocell data rate equally, each download will be served in random time given by an exponential distribution with mean $\frac{l}{\mu}$. With this, we can now generate a Markov chain of the number of active downloads in the system. In this Markov chain, at any state l, let λ_t be the rate at which new downloads are added and μ_t be the rate at which downloads are removed from the system. The Markov chain thus generated is identical to a processor-sharing queue and is depicted in Fig. 2.4. Next, we determine λ_t and μ_t.

Fig. 2.5 depicts how λ_t is modeled. Suppose $f_o X_{cell}$ users are outside their homes, and they generate download requests for the macrocell at rate λ_o irrespective of whether they have adopted service m or f. Of the remaining $f_i X_{cell}$ users who are inside their homes, a fraction α have adopted the femtocell and do not generate any requests for the macrocell while the rest generate requests at a rate λ_i. Note that this does not capture correlated behavior (like peak hour) but only an average sense of the traffic load. Then, we have,

$$\lambda_t = [f_i(1-\alpha) + \beta f_o] x \lambda_i N_{cell}. \tag{2.13}$$

When there are l downloads in parallel for the Markov chain, each taking i.i.d. exponential

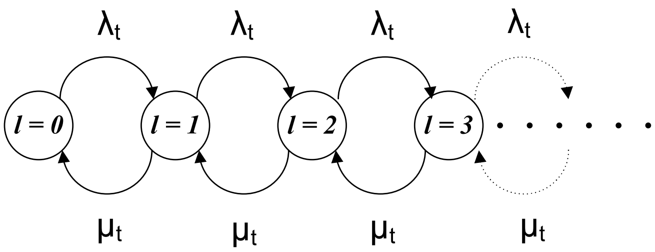


Figure 2.4. Markov Chain

amount of time with mean $\frac{l}{\mu}$, the rate at which the system exits from state l is given by

$$\mu_t = \frac{\mu}{l} \times l = \mu = \frac{R(\alpha x)}{\text{Mean File Length}}. \tag{2.14}$$

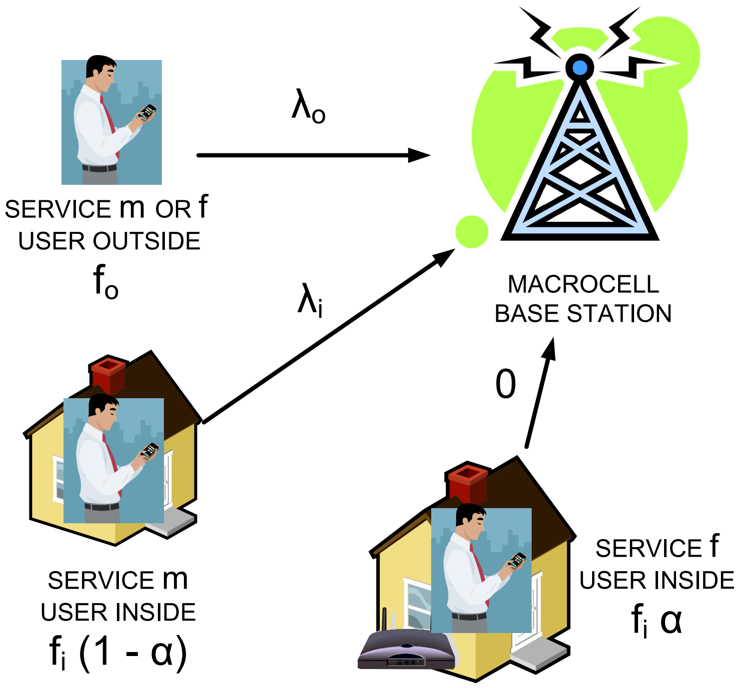


Figure 2.5. Macrocell Download Arrival Rate

Assume $\lambda_t < \mu_t$. Then, in the stationary state, the probability that the system is in state l is given by

$$P(l) = \frac{1}{1 - \lambda_t/\mu_t}\left(\frac{\lambda_t}{\mu_t}\right)^l = \frac{1}{1-\rho_t}(\rho_t)^l,$$

where $\rho_t = \lambda_t/\mu_t$. When the system is in state l, individual requests obtain a data rate of $\frac{R(\alpha x)}{l}$ from the system. Thus, letting $E[T_m]$ be the expected rate at which the system serves individual requests, when requests are present, we have

$$E[T_m] = \frac{\sum_{l=1}^{\infty} \frac{R(\alpha x)}{l} P(l)}{\sum_{l=1}^{\infty} P(l)}.$$

Note that $\sum_{l=1}^{\infty} P(l) \neq 1$; it represents the fraction of time that the macrocell is busy serving a download. On simplification, we have

$$\frac{E[T_m]}{R(\alpha x)} = \tau(\alpha, x) = \left(1-\rho_t\right)\frac{-\log\left(1-\rho_t\right)}{\rho_t}. \tag{2.15}$$

From (2.14), since μ_t depends on $R(\alpha x)$, we note that τ also depends upon the spectrum scheme being used. Henceforth, we use the subscripts s and c to denote the quantities specific to the split and common spectrum schemes respectively.

2.1.2 Model for $R(\alpha x)$

Assume that the operator has a total spectrum availability of $1.2W$. We will assume that the macrocell data rate is proportional to the employed spectrum. Accordingly, if the operator employs this complete spectrum for the macrocell, let him obtain the macrocell data rate $1.2R_0$. Let the corresponding service rate as defined in (2.14) be $1.2\mu_0$.

Split Spectrum

In this scheme, assume that the operator chooses to split his spectrum as follows - W for the macrocell and $0.2W$ for the femtocells.[3] In this case, we assume there will be no interference due to the femtocells and the macrocell data rate will be R_0 (correspondingly μ_0):

$$R^s(\alpha x) = R_0. \tag{2.16}$$

Common Spectrum

In this scheme, the operator chooses to operate both the macrocell and the femtocell in the same $1.2W$ MHz spectrum. In this case, when no femtocells are adopted, the macrocell data rate will be $1.2R_0$ (correspondingly $1.2\mu_0$). As femtocell adoption rises, interference from the femtocell downlink reduces throughput for macrocell users by affecting the downlink macrocell rate (see Fig. 2.3). Next, assuming that all femtocell users contribute equally to degradation of the macrocell data rate, we let the macrocell data rate decrease linearly in the number of users who adopt the femtocell.

$$R^c(\alpha x) = \max\{1.2R_0(1-d\alpha x), 0\}, \tag{2.17}$$

[3]This closely models the solution proposed by Clearwire/Sprint where 5 MHz will be reserved for femtocells and 30 MHz for the macrocell.

where $d > 0$ is the coefficient of degradation and $x = \frac{X_{cell}}{N_{cell}} = \frac{X}{N}$. For any network, d may be estimated as follows. If $R(\alpha) \neq 0$ for $\alpha < 1$, i.e., if the macrocell rate does not go to 0 before every user adopts the femtocell, then $d = 1 - \frac{R(1)}{R(0)} \leq 1$. Else, $d = \frac{1}{\alpha_{min}} > 1$ where $\alpha_{min} = \underset{R(\alpha)=0}{\arg\min}\, \alpha$.

Model for Γ

Γ gives us the distribution for the user valuations. For $\gamma \leq 0$, $\Gamma(\gamma) = 0$ and for $\gamma > \gamma_{max}$, $\Gamma(\gamma) = 1$. In this chapter, we will assume only a uniform distribution of users.

$$\Gamma(\gamma) = \frac{\gamma}{\gamma_{max}}, \;\; \gamma \in (0, \gamma_{max}]. \tag{2.18}$$

2.2 Operator Revenues

To simplify expression, we define the broadband rate factor b and the macrocell capacity c_0 as

$$b = \frac{E[T_b]}{R_0}, \;\; c_0 = \frac{\mu_0}{\lambda_i N_{cell}}, \tag{2.19}$$

and normalize the values of p_m, p_f, g_m, g_f w.r.t. kR_0. From (2.1) and (2.7), revenue has the same units as $\gamma g(\cdot)X \sim \gamma k R_0 x N$. Henceforth, we normalize the revenues w.r.t. kR_0N.

2.2.1 Femtocell: Split Spectrum

From (2.9), (2.15) and (2.16), we have (normalizing g_m^s and g_f^s)

$$g_m^s = \tau^s, \;\; g_f^s = (f_i b + f_o \tau^s), \tag{2.20}$$
$$\tau^s = (1 - \rho_t^s)\frac{-\log(1-\rho_t^s)}{\rho_t^s} \text{ where } \rho_t^s = \frac{[f_i(1-\alpha)+\beta f_o]x}{c_0}.$$

Solving (2.5), (2.6), (2.7) and (2.8) using (2.18) and (2.20), we can obtain the values of x and α for any given p_m and p_f. Substituting these values in (2.1) and optimizing w.r.t. p_m and p_f, we get the monopolist's optimal choice.

2.2.2 Femtocell: Common Spectrum

From (2.9), (2.15) and (2.17), we have (normalizing g_m^c and g_f^c)

$$g_m^c = 1.2\tau^c[1 - d\alpha x]^+, \;\; g_f^c = (f_i b + f_o g_m^c), \tag{2.21}$$
$$\tau^c = (1 - \rho_t^c)\frac{-\log(1-\rho_t^c)}{\rho_t^c} \text{ where } \rho_t^c = \frac{[f_i(1-\alpha)+\beta f_o]x}{1.2(1-d\alpha x)c_0}.$$

Solving (2.5), (2.6), (2.7) and (2.8) using (2.18) and (2.21), we can obtain the values of x and α for any given p_m and p_f. Substituting these values in (2.1) and optimizing w.r.t. p_m and p_f, we get the monopolist's optimal choice.

2.2.3 Base case: No Femtocell

In this case, the entire spectrum is used for the macrocell and no femtocells are deployed. Hence, this is similar to the common spectrum femtocell deployment with $\alpha = 0$. From (2.5), (2.7) and (2.21), γ_m is a solution of

$$\gamma = \frac{p_m}{g_m^c(0, 1-\Gamma(\gamma))} = \frac{p_m}{1.2kR_0\tau^c}. \tag{2.22}$$

Normalized operator revenue is $\frac{p_m X}{kR_0 N} = 1.2\gamma_m\tau^c\big[1-\Gamma(\gamma_m)\big]$.

2.3 Results

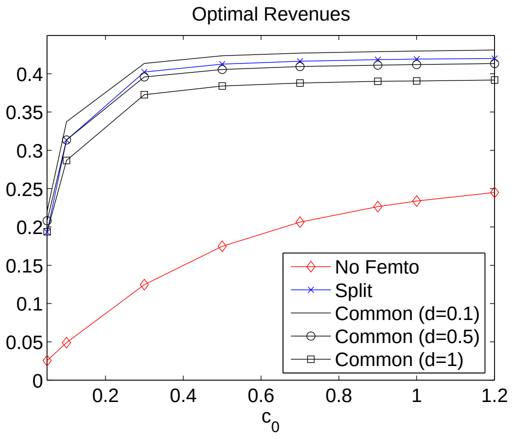
We carried out the numerical analysis[4] in MATLAB®. For each scenario, we varied p_m, the price of service m, and p_f, the price of service f. For each such pair (p_m, p_f), we determined the values of x and α in equilibrium using the fixed point approach. Table 2.1 lists the parameter values[5] used for the numerical analysis. Unless otherwise specified (or varied), the parameter values used in the numerical analysis are the ones listed in Table 2.1.

Table 2.1. Table of default parameter values

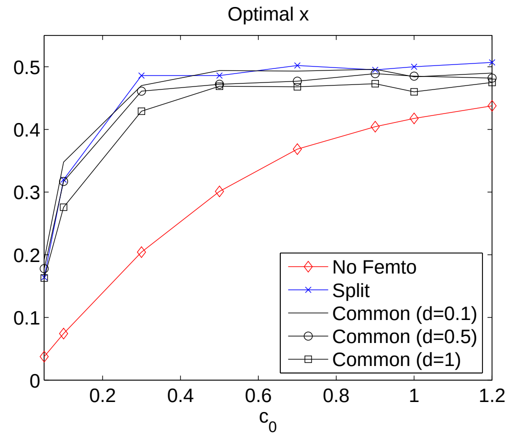
Parameter	Description	Value
$f_o = 1 - f_i$	Fraction outside	0.3
β	Activity Ratio $\left(\frac{\lambda_o}{\lambda_i}\right)$	1
b	Broadband factor $\left(\frac{T_f}{R_0}\right)$	2
c_0	Macrocell capacity $\left(\frac{\mu_0}{\lambda_i N_{cell}}\right)$	0.5
d	Degradation coefficient	0.5
γ_{max}	Maximum User Type	1

Figures 2.6(a), 2.6(b) and 2.6(c) depict the optimal values of the normalized revenues, x and α varying with the network capacity c_0 (defined in (2.19)). For each value of c_0, p_m and p_f have been chosen optimally. From Fig. 2.6(a), all femtocell schemes yield much higher revenues than with no femtocells. Further, this revenue gain is relatively more pronounced when the macrocell capacity is low. However, at these low levels of capacity, even the common

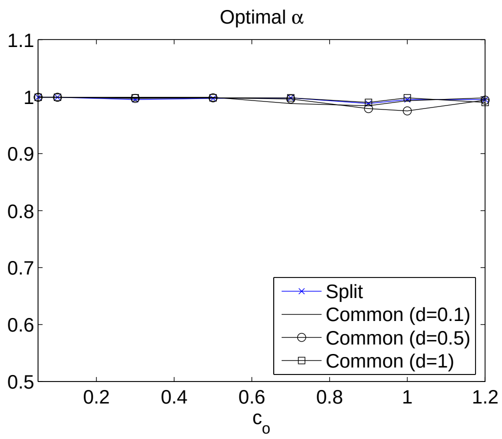
[4] Code is available upon request from nikhils[AT]eecs.berkeley.edu.
[5] Refer [73] for the choices of f_o and β.



(a) Optimal Revenues



(b) Optimal x



(c) Optimal α

Figure 2.6. Revenues, x and α vs c_0

spectrum scheme with $d = 1$ earns revenues comparable to the split spectrum deployment. As the degradation coefficient (d) increases, revenues from common spectrum deployments strictly decrease as expected. However, the common spectrum scheme with $d = 0.1$ earns higher revenues than the split spectrum scheme for all c_0, which confirms that a common spectrum scheme with low enough d may be superior to the split spectrum scheme. In fact, even with $d = 0.5$, the revenues from a common spectrum scheme are only marginally lower for all c_0. From Fig. 2.6(b), all femtocell schemes serve a larger number of users in equilibrium with the split spectrum scheme performing the best. Finally, from Fig. 2.6(c), at low macrocell capacity, almost full adoption is optimal for all femtocell deployments. At high macrocell capacity, optimal $\alpha < 1$ for common spectrum schemes with high degradation. Here, macrocell degradation impacts the number of users served. To keep the number of users (and the resulting revenues) high enough, α is reduced via a higher p_f.

2.4 Conclusion

In this chapter, we provide an economic framework for the analysis of adoption of femtocells. We compared the economic viability of two spectrum schemes - split spectrum and common spectrum - for deployment of femtocells in a 4G network. We assumed that a single monopolist network operator sets prices for both the mobile-only service and the femto+mobile service. Users were assumed to possess linear utility for data throughput and have different valuations for data throughput. Our results suggest that the optimal pricing scheme always charges a higher price for the femtocell service. Further, at the optimal prices, almost full adoption of femtocells is achieved in most cases. As expected, if the degradation coefficient is sufficiently low, the revenues from the common spectrum scheme are always higher than with the split spectrum scheme. However, interestingly, when the macrocell capacity is low, though all femtocell deployments bring in higher revenues, the revenues from common spectrum schemes are comparable to the split spectrum even when they heavily degrade the macrocell capacity.

Though we assumed that the user benefit is linear in throughput, in reality, we expect it to be concave. The linearity impacts our results in two ways, yet we will argue that it does not markedly change our results. One, high femtocell throughput does not result in proportionally higher revenues from femtocells. Including this effect will reduce the viability of all spectrum schemes equally, without affecting the relative performance. Two, users lose utility if the throughput varies considerably during usage. Though it appears that the common spectrum scheme suffers more from this effect than the split spectrum scheme due to the random interference from femtocells, this may not necessarily be true. Since the macrocell in the split spectrum scheme has lower capacity (due to the lower spectrum availability), the higher congestion may cause large throughput variations and wipe out any gains in the user utility.

We do not consider the additional operational cost of deploying femtocells and the probable femtocell revenue share with wireline providers. These factors will further reduce the viability of both schemes, but, the relative performance of the two schemes may not be affected. Though the linear interference model may be simplistic, it provides a framework for estimating the real-life parameters for such an environment.

We carry out the analysis from the point of view of service provider revenues only and do not consider the total utility accrued to users as a result of the scheme adopted. It is conceivable that the scheme that provides maximal revenues does not maximize the total utility for the users. The restriction to a monopolist is also a drawback of the analysis. As part of future work, the authors wish to investigate the performance of the two schemes in the presence of operator competition.

Appendix

Proof of Theorem 2.1.1

The user with type γ_m will be indifferent between buying some service and not buying, i.e., $U_m^{\gamma_m} = 0$ or $U_f^{\gamma_m} = 0$ depending upon the value of α.

If $\alpha = 1$, the critical user type γ_m prefers to buy service f over service m, and he is indifferent between buying service f and not buying anything:

$$\gamma_m g_f(1, x) - p_f = 0, \text{ and } \gamma_m g_m(1, x) - p_m \leq 0,$$

which gives us the desired result: $\gamma_m = \frac{p_f}{g_f(1,x)} \leq \frac{p_m}{g_m(1,x)}$.

If $\alpha < 1$, the critical user type γ_m is indifferent between buying service m and not buying anything and his utility from buying service f is strictly lower:

$$\gamma_m g_m(\alpha, x) - p_m = 0, \text{ and } \gamma_m g_f(\alpha, x) - p_f < 0,$$

which gives us the desired result: $\gamma_m = \frac{p_m}{g_m(\alpha,x)} < \frac{p_f}{g_f(\alpha,x)}$. Note that if $\gamma_m > \gamma_{max}$, then $x = 0$.

For any given x, if $\alpha < 1$, the critical user type γ_f is indifferent between service m and service f, i.e., $U_m^{\gamma_f} = U_f^{\gamma_f}$:

$$\gamma_f g_m(\alpha, x) - p_m = \gamma_f g_f(\alpha, x) - p_f$$

which gives us the desired result: $\gamma_f = \frac{p_f - p_m}{g_f(\alpha,x) - g_m(\alpha,x)}$. Note that if $\gamma_f > \gamma_{max}$, then $\alpha = 0$.

Proof of Corollary 2.1.2

If $\alpha = 1$, for any x, $\gamma_m = \frac{p_f}{g_f(1,x)} \leq \frac{p_m}{g_m(1,x)}$. Then, it must hold in equilibrium that $\frac{p_f}{p_m} \leq \frac{g_f(1,x)}{g_m(1,x)}$. If $\alpha < 1$, for any x, $\gamma_m = \frac{p_m}{g_m(\alpha,x)} < \frac{p_f}{g_f(\alpha,x)}$. Then, it must hold in equilibrium that $\frac{p_f}{p_m} > \frac{g_f(\alpha,x)}{g_m(\alpha,x)}$. If $\frac{p_f}{p_m} \not> \frac{g_f(\alpha,x)}{g_m(\alpha,x)}$, $\alpha \not< 1$, which gives us $\alpha = 1$ and the desired result.

Monotonicity of $g(\cdot)$

Let $\rho = \frac{\lambda_t}{\mu_t}$. From $\lambda_t < \mu_t$, we have $0 \leq \rho \leq 1$. Then,

$$\begin{aligned}
\frac{\partial \tau(\alpha, x)}{\partial \rho} &= \frac{\partial}{\partial \rho}((1-\rho)\frac{-\log(1-\rho)}{\rho}) \\
\therefore \frac{\partial \tau(\alpha, x)}{\partial \rho} &= \frac{1-\rho}{\rho}\frac{1}{1-\rho} + \frac{-1}{\rho^2}(-\log(1-\rho)) \\
&= \frac{1}{\rho} + \frac{\log(1-\rho)}{\rho^2} \\
&= \frac{\rho + \log(1-\rho)}{\rho^2}
\end{aligned}$$

The numerator $\rho + \log(1-\rho)$ is 0 at $\rho = 0$. For $\rho > 0$, the slope of the numerator is $1 - \frac{1}{1-\rho} = \frac{-\rho}{1-\rho} < 0$ which means it is a decreasing function. Hence, the numerator is always negative. Since the denominator is always positive,

$$\frac{\partial \tau(\alpha, x)}{\partial \rho} < 0. \tag{2.23}$$

From (2.13) and (2.14), we know that $\rho = \frac{[f_i(1-\alpha)+\beta f_o]\lambda_i N_{cell}}{\mu} x$, which gives us $\frac{\partial \rho}{\partial x} > 0$ and $\frac{\partial \rho}{\partial \alpha} < 0$. Hence, from (2.23),

$$\frac{\partial \tau(\alpha, x)}{\partial X} = \frac{1}{N}\frac{\partial \tau(\alpha, x)}{\partial x} < 0 \text{ and } \frac{\partial \tau(\alpha, x)}{\partial \alpha} > 0. \tag{2.24}$$

Chapter 3

Welfare Maximization - Quality of Service

In today's Internet, despite the technological possibility of providing network-wide differentiated services (QoS),[1] no such services are actually offered by the ISPs [26]. Although ISPs offer multiple contracts (rate tiers), these rate tiers are hardly service classes - they only specify the peak data rate on the access link and the maximum volume of data (aggregate per user). Indeed, no contract is backed by congestion guarantees throughout the network. Thus, although some contracts may improve performance if the access link was the network bottleneck, the user's data may still be congested at other points in the network. In this chapter, we do not explicitly consider these service tiers, though our analysis will reflect their presence. Recently, some ISPs have begun to offer services that prioritize the user's packets throughout their individual networks [1]. But such prioritization is limited by the amount of time (via restricting the volume of data) during which a file's download is sped up. We believe that these modest attempts to improve QoS are inadequate and reflect the ISPs' concerns about the imposition of network neutrality regulations.

In the broadest sense, network neutrality is about "the rules of the game" (standards, laws and regulations) between all networked parties. The network neutrality debate includes a wide array of issues about ISPs' rights and responsibilities with respect to network pricing and management, and interactions with content providers. We make no attempt at summarizing the issues. For extensive coverage, see [87].[2] The question whether current ISP practices should continue and be mandated by law, or ISPs ought to be allowed to charge users for QoS is part of this larger debate on network neutrality. This chapter addresses only this aspect of the network neutrality debate.

[1] For brevity, the term quality of service (QoS) refers to such services.

[2] Also see [96] for a discussion of relevant pros and cons.

In general, lack of QoS could be driven by numerous demand and supply considerations. Indeed, from the demand side, the ISPs could choose no QoS provision due to high uncertainty about demand for bandwidth,[3] and meager end-user demand for premium QoS, which does not justify the necessary up-front expenses.[4] In this chapter, we do not address these demand-side reasons.

From the supply side, four reasons are worth mentioning. First, the lack of QoS could be driven by difficulties of QoS pricing due to ISP competition [38, 93, 41]. Second, QoS provision could be an inferior investment relative to plain capacity expansion.[5] Third, contractual difficulties between the ISPs also undermine ISP incentives for QoS, that is if QoS were dependent on a single ISP, it would be profitable to offer,[6] but the end-to-end QoS guarantees could become impractical due to contractual and informational imperfections. Lastly, fourth, the threat of network neutrality regulations hampers ISPs' incentives for QoS. Indeed, at present, the ISPs are "at their best behavior," i.e., they suffer from self-imposed constraints, and these constraints may preclude the ISPs from investing into developing QoS [30, 105, 90].

In practice, ISPs are increasingly 'managing' network congestion via differentiated treatment of traffic from heavy network users. ISP investment into network equipment that enables them to implement such policy decisions indicate that they view plain capacity expansion as an inferior option. Contractual difficulties could be an important cause for the lack of end-to-end quality of service. But, increasingly, a large fraction of network traffic is being served via Content Distribution Network (CDN) servers and data centers located at the edge of the network [59]. Hence, only a small fraction of traffic traverses ISP boundaries. This implies that ISPs can substantially improve service quality by offering differentiated services within their own domains. End-to-end differentiation is an open problem and will not be tackled in this chapter. To sum up, this chapter addresses supply-side reasons, focusing on the first and fourth reasons only.

We build on the classical industrial organization literature pioneered by [64, 56], where both capacities and prices are chosen strategically. [56] demonstrated that when the firms first compete in capacities, and thereafter engage in price competition (Bertrand), the Cournot-like outcome is an equilibrium. Several recent papers [52, 3, 2, 79] apply these ideas to large-scale communications networks, and most address QoS issues (congestion) only from a routing perspective. None of these papers considers two service classes.

Our pricing model is based on the network architecture similar to the Paris Metro proposal (PMP) [70]. Other closely related papers modeling PMP are [33, 27]. Both [33, 27] focus on ISP competition, with network access provided by duopolists, but include no analysis of the effects of ISP choices on user welfare. We assume that capacity is costly, and provide complete analysis of ISP capacity choices and its division between service classes. We calculate end-user welfare for any number of competing ISPs, and compare equilibria for

[3] This uncertainty is so profound that demand estimation posits difficulties [37, 98, 6].

[4] From analysis in [25], with high upfront costs, only primitive QoS mechanisms are viable, which may be insufficient to achieve a meaningful quality increase.

[5] For example, [71] asserts that improving QoS by investing in capacity is more profitable than investing in provision of multiple service classes.

[6] We demonstrate in [84], that QoS provision is indeed profitable for a monopolist.

a single- and two-service classes networks. Our results indicate that ISP competition per se does not preclude QoS provision. We find that even with perfect competition between ISPs, two service classes remain optimal. Thus, in contrast with [33], we do not view ISP competition alone as a valid explanation for the lack of QoS in the Internet.

Instead of appealing to ISP competition to explain the current lack of QoS, we suggest that the lack of QoS could be driven by political economic considerations. Indeed, in our setting, when the number of competing ISPs is small, a high fraction of ISP capacity is allocated to premium service. In this case, the fraction of existing (who used to buy access in a single service class regime) users forced to buy premium access is also high, because the quality of basic service does not satisfy their needs. For these users, the welfare is lower than it was with a single service class, and this adverse distributional effect could bring the discontent of such users. The ISPs may fear that this will justify the imposition of neutrality. Thus, at low levels of ISP competition, the threat of network neutrality regulations could be used to explain the current lack of QoS on the Internet. This effect becomes less significant with increased ISP competition, and disappears when the number of competing ISPs is high. Even with highly competitive ISPs, our analysis indicates a superiority of two service classes for both ISPs and end-users. Still, even in this case, driven by the ISP fear of neutrality regulation, ISPs may constrain themselves from offering two service classes.

In our related work [85], we explore an inexpensive regulatory tool that alleviates investment disincentives of ISPs by securing their property rights over a pre-specified fraction of their capacity. This tool achieves two goals. It reduces negative distributional effects of adopting multiple service classes; also, it eliminates the threat of imposition of network neutrality for a pre-specified fraction of the ISPs' capacity and restores the ISPs' incentives for QoS deployment.

The chapter is organized as follows. In Section 3.1, we outline our model. In Sections 3.2.1 and 3.2.2, we analyze the networks where each ISP provides a single-service class and two-service classes respectively. In Section 3.3, we present our results and in Sections 3.4 and 3.5, we discuss our findings and conclude. The technical details are relegated to the Appendix.

3.1 Model

3.1.1 The environment

To start, let us consider a single service class network. We assume that M identical competing ISPs (where M is fixed) offer connectivity to a user base of fixed size. Let N (which we assume to be large) be the total number of end-users. Here, and below, we use the superscript $m = 1, ..., M$ to denote the variables of the m-th ISP. First, each ISP chooses his capacity $C^m \geq 0$ that he builds at a constant unit cost $\tau > 0$. Investment in capacity is irreversible. Second, once the capacity is sunk, each ISP makes his pricing decision p^m, after which *the end-user price for network connection* p (access price p for short) is determined by

$$p = \min_{m=1,..M} p^m. \tag{3.1}$$

From (3.1), when one of the ISPs announces a price lower than the other ISPs, due to ISP competition, all others must lower their prices as well. Indeed, the frequently occurring provision "If you find an offer with a lower price, we will be happy to match it" amounts exactly to (3.1).

Each user decides whether to purchase the service, and from which ISP. The m-th ISP's objective is to maximize profit Π^m_{total} which equals his revenue net of his capacity expense:

$$\Pi^m_{total} = \max_{C^m, p^m} \{pZ^m - \tau C^m\},$$

where Z^m is the number of users who adopt the service from the m-th ISP. Also, let $Z = \sum_m Z^m$ be the aggregate number of end users who purchase the service.

Next, let us consider two service classes l and h. Let the m-th ISP allocate capacities C^m_i, and quote prices p^m_i with $p^m_h > p^m_l$ for service $i = l, h$. Also, let $C_i = \sum_m C^m_i$ be the aggregate capacity for service i. Similar to (3.1), the access prices p_i are determined by

$$p_i = \min_{m=1,..M} p^m_i, \; i = l, h. \tag{3.2}$$

We call h *premium service* (the service with a higher access price), and we call l *basic service* (the service with a lower access price). Then, the m-th ISP's objective becomes:

$$\Pi^m_{total} = \max_{C^m_i, p^m_i} \left\{ \sum_i p_i Z^m_i - \tau C^m \right\},$$

where $C^m = \sum_{i=l,h} C^m_i$ and Z^m_i is the number of users who adopt service i from the m-th ISP. Let $Z_i = \sum_{m=1,..M} Z^m_i$ be the aggregate number of end-users adopting service $i = l, h$ and $C_i = \sum_m C^m_i$ be the aggregate capacity. The access price p_i of each service i is determined by (3.2).

We define the *quality of service* q observed by users as $q = 1 - Z/C$, if Z users are multiplexed in capacity C. This definition of quality reflects the common perception about service quality [33]. As Z decreases and capacity remains the same, the quality of service improves, i.e., as the capacity per user increases, so does the quality. Finally, we assume that each user contributes equally to the loss of quality, i.e. each user generates an identical unit amount of traffic. Let $z = Z/N$ denote the fraction of the users who purchase the service and $c = C/N$ denote the capacity per user in the base. Then, using these normalized values, $q = 1 - z/c$. Similarly, for $i = l, h$, let $z^m_i = Z^m_i/N$, $c^m_i = C^m_i/N$ and $q_i = 1 - z_i/c_i$.

Let each user in the user base be characterized by his type θ, which we assume to be a random variable with support $[0, 1]$. For a user with type θ, the lowest *acceptable* service quality is $q = \theta$; and his highest *affordable* access price is $p = \theta$. Thus, a user buys a service only if this service is acceptable and affordable, i.e., $p < \theta \leq q$ [99]. Here, we assume that the user willingness-to-pay and the quality requirement coincide, i.e., a user with a high quality requirement also has a high willingness-to-pay. One can imagine cases violating this assumption. For example, a user with low willingness-to-pay could have high requirement for quality. Still, we expect that, in most cases, price and quality requirements are highly correlated, and our model is well-suited for such a scenario. For the user with type θ, the surplus U_θ is given by

$$U_\theta = (\theta - p)I(q - \theta), \text{ where } I(y) = \begin{cases} 1 \text{ if } y \geq 0 \\ 0 \text{ if } y < 0 \end{cases}. \tag{3.3}$$

The parameter θ can also be interpreted as the network quality required for the most quality-intensive application that this user utilizes. Indeed, when a user adopts a service for e-mail only, he gains no extra benefit if his network quality permits someone else to use streaming video, which makes (3.3) a good fit. Our user preferences are primitive, but similar to routinely used preferences in such contexts [2, 102]. In these papers, user utility depends explicitly on q, which they term "congestion cost," but all users are homogeneous. Our users differ in type, but the preferences depend on q as a step function only.

In general, for a distribution $p(\theta)$ of user types $\theta \in [0,1]$, the aggregate user surplus can be expressed as $U_{total} = \int_0^1 U_\theta p(\theta) N d\theta$. With an assumption of user types uniformly distributed in $[0,1]$, we have $U_{total} = \int_0^1 U_\theta N d\theta$. Although the uniformity assumption is restrictive, it is common [42]. We impose this assumption for analytical tractability. In contrast to this assumption, [4, 88] impose fewer restrictions on the distribution of user types but provide their analyses for a monopolistic ISP only. Let $U = \frac{U_{total}}{N}$ be the surplus per user in the base. Then,

$$U = \int_0^1 U_\theta d\theta. \tag{3.4}$$

Similarly, let $\Pi^m = \frac{\Pi^m_{total}}{N}$. With a single service class, the m-th ISP objective is:

$$\Pi^m = \max_{c^m, p^m} \left\{ p z^m - \tau c^m \right\},$$

and with two service classes, his objective becomes:

$$\Pi^m = \max_{c^m, x^m, p^m_{i=l,h}} \left\{ \sum_{i=l,h} p_i z_i^m - \tau c^m \right\}. \tag{3.5}$$

Per user in the base, the social surplus S is the sum of user surplus and provider profit: $S = U + \Pi$, where $\Pi = \sum_{m=}^{M} \Pi^m$.

We assume that, for each service, the number of end-users who purchase a service from the m-th ISP is proportional to his share of capacity dedicated to provision of that service.[7] Then, for a single service class we have:

$$z^m = s^m z, \text{ where } s^m = \frac{c^m}{c} \tag{3.6}$$

[7]This assumption may be relaxed and was imposed for expositional reasons.

and for two- service classes we have:

$$z_l^m = s_l^m z_l \text{ and } z_h^m = s_h^m z_h, \text{ where } s_l^m = \frac{c_l^m}{c_l}, s_h^m = \frac{c_h^m}{c_h}. \tag{3.7}$$

We justify (3.6) and (3.7) by the following provision routinely present in the end-user contracts: "you can cancel your contract any time during the first 30 days". Indeed, if one of the ISPs serves a higher share of end-users than his respective share of total capacity, his users would experience a lower quality. Then, the marginal user (for whom this quality difference matters) will use this provision to switch to another provider with disproportionately (relative to his installed capacity) lower number of end-users. In other words, the provision permits to recreate the situation in which the ISPs' investments are observable by end-users.

3.1.2 Network Regulations

Let x^m denote the fraction of the m-th ISP capacity allocated to the premium service. Then we have:

$$c_l^m = (1 - x^m)c^m \text{ and } c_h^m = x^m c^m. \tag{3.8}$$

From (3.8), one can easily switch between the use of (c_l^m, c_h^m) and (c^m, x^m) as choice variables. In fact, we will freely switch between these two notations.

We will say that the network is regulated when the regulator restricts capacity division between the service classes. The regulator's choice variable is $\bar{\mathbf{x}} = (\bar{x}^1, \dots \bar{x}^m \dots, \bar{x}^M)$, i.e., the regulator only affects the m-th ISP by constraining him from dedicating more than a fraction $\bar{x}^m$ of his capacity to service h. We assume that the regulatory constraint is identical for all ISPs $\bar{x}^m = \bar{x}$. Then, with regulations, the m-th ISP's objective becomes

$$\Pi^m = \max_{c^m, p_l^m, p_h^m} \left\{ p_l z_l \frac{c_l^m}{c_l} + p_h z_h \frac{c_h^m}{c_h} - \tau c^m \right\} \text{and } c_h^m \le \bar{x} c^m.$$

The case of a single service class is identical to the imposition of $\bar{x} = 0$. We do not consider explicit regulations in the case of single service class, but we believe that the lack of QoS provision by ISPs in the current Internet reflects the tacit presence of such a regulatory threat. The ongoing network neutrality debate confirms that this threat is indeed real. We argue that this regulatory threat makes the ISPs to act as if $\bar{x} = 0$ is imposed. This regulatory threat could explain why QoS is not provided currently (see Introduction). Thus, we use the surplus of the single service class users as a proxy for the surplus of the current Internet users.

We consider three regulatory scenarios. Regulator 1 (a social planner) maximizes social surplus (sum of aggregate user surplus and ISP profit), regulator 2 maximizes user surplus and regulator 3 maximizes the surplus of the users who are served under a single service class. For the regulators 1 - 3, the respective objectives S_1, S_2 and S_3 are:

$$S_1 = \max_x \{U + \Pi\}; \ S_2 = \max_x U; \ S_3 = \max_x U|_{\theta \in \Theta}, \tag{3.9}$$

where U is defined in (3.4), and Θ denotes the set of end-users served in the network with a single-service class only. Let $\bar{x}_1$, $\bar{x}_2$, $\bar{x}_3$ be the values chosen by the regulators 1-3 respectively.

3.1.3 The Order of Moves

In the unregulated two-service class case, we assume the following order of moves. First, the ISPs simultaneously and independently invest in irreversible capacity $\mathbf{c} = (c^1, ..., c^M)$, and observe $\mathbf{c}$. Second, the ISPs simultaneously and independently choose $\mathbf{x} = (x^1, \ldots, x^M)$, i.e., the capacity division between the services. Let $\mathbf{c}_i = (c_i^1, ..., c_i^M)$ denote the vector of ISP capacities dedicated to the provision of service $i = l, h$ with $c_h^m = x^m c^m$. Next, the ISPs play a subgame $G(\mathbf{c}, \mathbf{x})$, in which they make pricing decisions p_i^m and the access price p_i of each service i is determined by (3.2).

With the regulator present, we assume that he makes the first move and announces $\bar{x}$. After the ISPs observe $\bar{x}$, they simultaneously and independently invest in irreversible capacity $\mathbf{c}$. Next, upon observing the capacities the ISPs play the game $G(\mathbf{c}, \mathbf{x})$. In contrast with the unregulated case, where $\mathbf{c}$ is chosen *before* $\mathbf{x}$, with a regulator, $\bar{x}$ is chosen *before* capacities are sunk. Notice, that single-class service equilibrium can be obtained as the case in which all ISPs must choose $x^m = 0$, and this restriction is imposed before the ISPs invest in capacities.

Another possibility is that the regulator announces $\bar{x}$ after the ISPs' capacity is installed. Then, in the short-run, since ISP capacity investments are irreversible, the regulator could impose an $\bar{x}$ that is extremely favorable to users. However, in the long-run, as traffic requirements increase for all users, ISPs will aver from capacity upgrades and the new equilibrium will correspond to the order of moves we have described here. In the rest of the text, we analyze only a regulator concerned about the long-term impact of his choices. We refer an interested reader to our working paper [86] for a complete analysis and comparison of both the long-run and the short-run cases.

With two service classes, in both cases, with and without regulator, we assume that the ISPs choose their prices *after* they observe capacities $\mathbf{c}$. We justify this assumption by the scale of the required initial investments. Capacities of the ISPs tend to be longer-term investments in infrastructure, and thus are harder to adjust.

In the unregulated case, the ISPs' capacity divisions between service classes $\mathbf{x}$ is assumed fixed and observable prior to prices being chosen by the ISPs. We justify this assumption by letting prices adjust quickly to changes in $\mathbf{x}$. Indeed, consider the situation where the m-th ISP alters his x^m after the prices are chosen. In reality, such an alteration would be immediately followed by price adjustments and ISPs would choose the new prices optimally, given the new $\mathbf{x}$. Hence, we can also view this subgame as a game being replayed by the ISPs until a stable capacity division vector $\mathbf{x}$ and the corresponding optimal prices are reached. Modeling this as the result of a game where $\mathbf{x}$ is chosen before prices is hence justified.

3.1.4 Subgame $G(\mathbf{c}, \mathbf{x})$

Consider the subgame $G(\mathbf{c}, \mathbf{x})$, which occurs after ISPs' capacities have been sunk and their divisions are chosen. In $G(\mathbf{c}, \mathbf{x})$, each ISP maximizes his gross revenue.

We analyze $G(\mathbf{c}, \mathbf{x})$ when the ISPs provide a single- ($\mathbf{x} = 0$) or two service classes:

(a) Single service class: The ISPs simultaneously choose p^m, and the access price is determined by (3.1).

(b) Two service classes: The ISPs simultaneously choose (p_l^m, p_h^m), and the access prices are determined by (3.2).

In $G(\mathbf{c}, \mathbf{x})$, the m-th ISP objective $R^m = R^m_{total}/N$ is:

$$R^m = \max_{p_h, p_l}\left(\sum_{i=l,h} s_i^m p_i z_i^m\right) = \frac{c_l^m}{c_l} p_l z_l + \frac{c_h^m}{c_h} p_h z_h. \tag{3.10}$$

Short-run Game

In addition, we will consider a modified game to address the short-run effects of transition. We will define this short-run game as a game where ISP capacities are fixed at the equilibrium level of the single service class ($\mathbf{c}^{\dagger}$). With this capacity fixed, the ISPs engage in provision of two service classes with or without regulation imposed. Accordingly, the regulator also assumes a fixed $\mathbf{c}^{\dagger}$ in his optimization.

3.2 Analysis

Although a single service class can also be viewed as a regulatory restriction $\bar{x} = 0$, to ease the exposition we consider the case of single service class separately.

3.2.1 Single Service Class

Assume that all ISPs provide a single service class only. This means that the entire capacity c is offered at the access price p, determined by (3.1).

Theorem 3.2.1 *With a single service class, there exists a unique Pareto efficient equilibrium in the game between M ISPs. This equilibrium is symmetric; aggregate capacity, number of adopting users, and service quality are increasing in M.*

Proof See Appendix for details.

Let us summarize the intuition behind the proof of Theorem 3.2.1. Let $M = 1$. From (3.3), a user with type θ will adopt the access (service) if and only if $p < \theta \leq q$, where $q = 1 - z/c$, with z being the fraction of users who adopt the service. Clearly, the service is affordable to all users with type $\theta > p$. As more users adopt the service, z increases and q decreases until it becomes equal to the user type at some critical value of θ. Let users with types $\theta \in (\underline{\theta}, \bar{\theta}]$ adopt the service. Then, we obtain (see Appendix)

$$\underline{\theta} = p \text{ and } \bar{\theta} = \frac{p + c}{1 + c}, \text{ and } z = \frac{c}{1 + c}(1 - p), \tag{3.11}$$

and the monopolist maximizes his revenue given by

$$R = pz = \frac{c}{1 + c}p(1 - p). \tag{3.12}$$

The revenue maximizing price is $p = 1/2$.

Next, let $M > 1$, and the m-th ISP's capacity be fixed at c^m. Once this capacity is sunk, the ISP's objective is to maximize his revenue $R^m(\mathbf{c}, p)$:

$$R^m(\mathbf{c}, p) = pz^m, \text{ where } z^m = s^m z. \tag{3.13}$$

Since s^m are fixed once capacities $\mathbf{c}$ are sunk, revenue maximization becomes identical for all ISPs: $\max_{p^m} pz$. Therefore, it is optimal for each ISP to choose the price coinciding with the monopolist's access price $p = 1/2$, as this maximizes the total revenue. Indeed, if an ISP deviates and quotes a lower price, his revenue decreases because his share of revenue remains the same, but the aggregate revenue becomes lower.

Note that any price $p < 1/2$ is an equilibrium in this subgame. However, in all these equilibria, the revenue earned by the ISPs is lower than the monopolist's access price 1/2. Hence, these equilibria are not Pareto-optimal. Thus, for any M, and fixed $\mathbf{c}$, the Pareto efficient equilibrium price is identical to the one in the game with $M = 1$.

Hence, for any c, the aggregate revenue is maximized at $p^\dagger = 1/2$, and equals $R(\mathbf{c}, 1/2) = \frac{c}{4(1+c)}$, which permits us to simplify the m-th ISP objective to

$$\Pi^m = \max_{c^m} \frac{c^m}{4(1 + c)} - \tau c^m.$$

This objective resembles the players' objectives under Cournot competition, similar to [56]. Henceforth, we will use the superscript † to designate the ISPs' optimal choices in the single service class case. In Appendix, we derive the aggregate equilibrium capacities for any M:

$$c^\dagger(\tau, M) = \frac{(M - 1) + \sqrt{(M - 1)^2 + 16\tau M}}{8\tau M} - 1, \tag{3.14}$$

for $\tau \in (0, 0.25)$, and $c^\dagger(\tau, M) = 0$, for $\tau \geq 0.25$. Also,

$$p^\dagger = 1/2;\ \underline{\theta}^\dagger = \frac{1}{2} \text{ and } \bar{\theta}^\dagger = \frac{\frac{1}{2} + c^\dagger}{1 + c^\dagger}. \tag{3.15}$$

From (3.14), for a monopolist and for perfect competition:

$$c^\dagger(\tau, 1) = \frac{1}{2\sqrt{\tau}} - 1 \text{ and } c^\dagger(\tau, \infty) = \frac{1}{4\tau} - 1. \tag{3.16}$$

From (3.14) and (3.15), we have $U^\dagger$ and $S^\dagger$ increasing, and $\Pi^\dagger$ decreasing with M. Due to peculiarities of our model, price does not depend on M for single service class. Still, note that equilibria differ with M because capacities increase in M. In general, prices should depend on M, and with two service classes, they do.

3.2.2 Two Service Classes

To start, we consider a monopolistic ISP ($M = 1$) who provides two service classes. From (3.5), his objective is:

$$\Pi = \max_{c,x,p_h,p_l} (\sum_{i=l,h} p_i z_i - \tau c),$$

where $c = c_l + c_h$ and from (3.8), $c_l = (1-x)c$ and $c_h = xc$. Henceforth, we will denote the ISPs' optimal choices in the case of two service classes by ‡.

Theorem 3.2.2 *For $M = 1$, there exists a unique equilibrium in $G(c, x)$:*

$$p_l^\ddagger(c, x) = \tfrac{1}{2} - \tfrac{c_h c_l}{2[(1+c_l)(1+c_h)c_l + c_h]},\ p_h^\ddagger = \tfrac{p_l + c_l}{1+c_l}, \tag{3.17}$$

$$\textit{and, } \underline{\theta}_l = p_l^\ddagger,\ \bar{\theta}_l = \underline{\theta}_h = p_h^\ddagger,\ \bar{\theta}_h = \tfrac{p_h^\ddagger + c_h}{1+c_h}. \tag{3.18}$$

The users with types $\theta \in (\underline{\theta}_l, \bar{\theta}_l]$ and $\theta \in (\underline{\theta}_h, \bar{\theta}_h]$ adopt services l and h respectively.

Proof See Appendix for details.

Let us summarize the intuition behind the proof of Theorem 3.2.2. First (Lemma 3.6.1), we note that, analogous to the case of a single service class, users with type $\theta \in (\underline{\theta}_l, \bar{\theta}_l]$ adopt service l, irrespective of price p_h. Second (Lemma 3.6.2), we prove that, for any given p_l, the ISP's revenue is maximized at some $p_h \geq \bar{\theta}_l$. The result follows from (3.3), since introducing a service h priced at $p_h < \bar{\theta}_l$ has no effect on the users of service l; no such users will shift to service h. Also, there is no effect on the number of users adopting service h. Thus, a lower price results in a lower revenue, from which $p_h \geq \bar{\theta}_l$ follows.

Third (Lemma 3.6.3), we show that in the ISP optimum, $\bar{\theta}_l = \underline{\theta}_h = p_h$. This implies that there is "no gap" between service classes, i.e., $(\bar{\theta}_l,\ \underline{\theta}_h)$ is an empty interval. Assume to the contrary that there is a gap between the service classes. Then, one can view each of the two service classes as separate networks, each providing a single service class. The monopolistic ISP will price each class independently to maximize his profit. From (3.12) and (3.13), in a single service class network, the ISP revenue is concave in price and the optimal price is unique and equal to 1/2. Hence, if there is a gap, $p_h = p_l = 1/2$, which contradicts $p_h > p_l$. Thus, indeed, there is no gap and $\bar{\theta}_l = \underline{\theta}_h = p_h$. From Lemmas 3.6.1 - 3.6.3, we obtain (3.18) which can be expressed in terms of c and x using (3.8). This permits us to express the ISP revenue as a function of c, x, and p_l only:

$$R(c, x, p_l) = \frac{c_l}{1 + c_l} p_l(1 - p_l) + \frac{c_h}{1 + c_h} \frac{p_l + c_l}{1 + c_l} \frac{1 - p_l}{1 + c_l}. \tag{3.19}$$

Thus, with fixed c_l and c_h, the ISP revenue maximization can be expressed as an optimization in just one variable – p_l. Maximizing (3.19) with respect to p_l, we obtain (3.17) (see Appendix), which completes the proof of Theorem 3.2.2.

From Theorem 3.2.2, we have $\bar{\theta}_h > \bar{\theta}_l$, i.e., the service h with a higher price ($p_l < p_h$), has a higher quality ($q_l < q_h$) too.

Theorem 3.2.3 *For $M > 1$, there exists a unique symmetric Pareto efficient equilibrium in $G(\mathbf{c}, \mathbf{x})$; the prices and end-user types served in each service class are identical to those of $G(c, x)$ with $M = 1$.*

Proof Consider $G(\mathbf{c}, \mathbf{x})$ in which the ISPs invest symmetrically ($c^m = \frac{c}{M}$), and divide their capacities identically ($x^m = x$). From (3.2), (3.7) and (3.10), at any given p_l and p_h, in each service class, each ISP's share of the total revenue equals to his capacity share $1/M$. Thus, irrespective of the ISPs' price choices, the revenue is shared equally. Therefore, the ISPs' optimal prices coincide with the monopolist's access prices (3.17), as this maximizes the aggregate revenue. Indeed, if an ISP deviates from these prices, his revenue decreases, because his share of revenue remains the same, but the aggregate revenue becomes lower.

Note that any combination of access prices, where both services l and h or either of them have a lower access price than the monopolist's optimal (3.17), forms an equilibrium in this subgame. However, in all these equilibria, the revenue earned by the ISPs is lower than the one with the monopolist's optimal access prices. Hence, these equilibria are not Pareto-optimal. Thus, for any M, and fixed (and symmetric) $\mathbf{c}$ and $\mathbf{x}$, the symmetric Pareto efficient equilibrium price is identical to the one in the game $G(c, x)$ with $M = 1$.

Corollary 3.2.4 *For any fixed c and x and any M, we have*

$$p_l^{\ddagger}(c, x) < \frac{1}{2} \textit{ and } p_h^{\ddagger}(c, x) > \frac{1}{2}. \tag{3.20}$$

Proof See Appendix.

From (3.3) and (3.20), in the case of a transition from a single service class to two service classes, all existing single-service class users who adopt service l gain surplus, and those who adopt service h lose surplus.

Next, we combine (3.10) with the result of Theorem 3.2.3 to obtain the m-th ISP's equilibrium revenue in $G(\mathbf{c}, \mathbf{x})$:

$$R^m = \frac{c_l^m}{1+c_l} p_l^\ddagger (1 - p_l^\ddagger) + \frac{c_h^m}{1+c_h} \frac{p_l^\ddagger + c_l}{1+c_l} \frac{1-p_l^\ddagger}{1+c_l}, \tag{3.21}$$

where $p_l^\ddagger$ is given by (3.17). Expression (3.21) is too cumbersome to carry out further investigation analytically, necessitating a numerical analysis (see Section 3.3). The uniqueness results in the following section justify this numerical analysis.

3.2.3 Uniqueness of Equilibria

In Appendix, we prove the following uniqueness results.

Theorem 3.2.5 *For $M = 1$, at any capacity c, there exists a unique x at which the ISP's profit is maximized. The ISP's revenue increases with x for $x \in (0, x^\ddagger)$.*

Theorem 3.2.6 *In the limit of $M \to \infty$, at any aggregate capacity c, there exists a unique x in the equilibrium of $G(c)$.*

Theorem 3.2.7 *For any fixed c, there exists a unique symmetric Pareto efficient equilibrium in the game of M competing ISPs.*

The proof of Theorem 3.2.7 starts by establishing the existence of a unique equilibrium capacity division ($x^\ddagger(c, M)$) in each of the games with $M = 1$ ($x^\ddagger(c, 1)$) and $M = \infty$ ($x^\ddagger(c, \infty)$). For any fixed capacity c, we obtain $x^\ddagger(c, 1) > x^\ddagger(c, \infty)$, i.e., a monopolistic ISP reserves a higher fraction of his capacity to the premium service than a perfectly competitive ISP does. Further, we show that, for any $M > 1$,

$$x^\ddagger(c, M) \in (x^\ddagger(c, \infty), x^\ddagger(c, 1)). \tag{3.22}$$

From Theorem 3.2.5, for any $x \leq x^\ddagger(c, 1)$, aggregate (and therefore each ISP's) profit increases with x. Combining with (3.22), we obtain that a unique Pareto optimal $x^\ddagger(c, M)$ exists for any fixed c, which leads to Theorem 3.2.7.

Theorem 3.2.8 *There exists a unique Pareto efficient equilibrium in the game of M ISPs competing in the presence of a regulator. This equilibrium is symmetric.*

Proof Under regulation, x is fixed before capacity is sunk. The prices in $G(\mathbf{c}, \mathbf{x}$ are the same as in the monopolist's case. Note that the single-service class network is identical to a regulated two-service class network with $\bar{x} = 0$. Hence, once $\bar{x}$ is non-zero, the resulting capacity game is similar to the capacity game with a single-service class (with $x \neq 0$ and prices $p_l^{\ddagger}$ and $p_h^{\ddagger}$). The proof for the unique Pareto-efficient equilibrium is identical to the proof of Theorem 3.2.1.

3.3 Results

In this section, we present the core results of our model. We compare the equilibrium of the game in which the ISP(s) provide(s) a single service class (denoted by †) with the equilibria of the games in which the ISP(s) provide(s) two service classes for the unregulated ISP(s) (denoted by ‡) and the ISP(s) constrained by regulators. We consider 3 regulators - a social welfare maximizer, a user welfare maximizer and an existing user welfare maximizer (denoted by the superscripts 1, 2 and 3 respectively). Note that we model the unregulated and regulated scenarios using different games. Indeed, while in the unregulated scenario, the x is chosen ex post (after the capacities are sunk), with the regulator, the restriction on capacity division is announced ex ante (prior to investments in capacity).

We have obtained a closed form solution for the equilibrium with perfect competition, i.e. in the limit of $M \to \infty$, only (see Appendix). As mentioned before, for any finite M, the expression (3.21) is too cumbersome to carry out further investigation analytically. Hence, we solve the ISPs' optimization problem numerically[8] using MATLAB®.

For the unregulated scenario, the determination of equilibrium capacity and its division x is nested in the following four-step procedure. First, we let all ISPs, except one, have identical capacity $\tilde{c}$ and consider the remaining ISP's (w.l.o.g. the 1'st ISP) choice of capacity, c^1. Second, holding these capacities constant, we let all ISPs, except the 1'st ISP, have identical $\tilde{x}$. We determine the the 1'st ISP's best response x^1 by maximizing the revenue (3.21). Similarly, we determine the other ISPs' best responses $\tilde{x}$ to x^1. The Nash equilibrium is found as the point where these best responses coincide. This gives us an equilibrium capacity division for the capacities described in step 1. Third, we iterate step 2 varying $\tilde{c}$ with a step of 0.01. We determine the 1'st ISP's best response capacity to any capacity $\tilde{c}$ by maximizing the profit (3.5). Last, we obtain the equilibrium capacity by finding the value of $\tilde{c}$ which coincides with the best response capacity of the 1'st ISP.

For each regulated scenario, x is fixed by the regulator. First, we fix x, and as before, let all ISPs, except one (w.l.o.g. the 1'st ISP), have identical capacity $\tilde{c}$. Let the 1'st ISP's choice of capacity be c^1. Second, we use the x fixed at step 1 to determine the 1'st ISP's best response capacity to any fixed capacity $\tilde{c}$ by maximizing the profit (3.5). Third, we obtain the equilibrium capacity $c(x)$ by finding $\tilde{c}$ which coincides with the 1'st ISP's best response. Fourth, we vary x from 0 to 1 with a step of 0.01 and for each x, determine the

[8]The code is available upon request from nikhils@eecs.berkeley.edu.

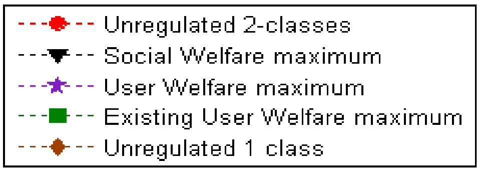


Figure 3.1. Legend

ISPs' capacity $c(x)$ using the steps described above. Using $c(x)$, we calculate welfare and profits. Finally, for each regulator, we determine his optimal x from these quantities.

From (3.16), for a single service class scenario, non-zero capacity is optimal only if $\tau \in (0, 0.25)$. Our results are presented for $\tau \in [0.01, 0.15]$ only, as when τ approaches 0 or 0.25, the computations involve division by terms approaching zero. We obtain optimal values of the functions of interest by cycling over the steps described above for $\tau \in [0.01, 0.15]$ with a step size of 0.01. The legend for all figures is depicted in Fig. 3.1.

Fig. 3.2 depicts how x (the fraction of capacity that each ISP allocates for service h) varies with the cost of capacity. Figures 3.2(a) - 3.2(c) depict x for unregulated and regulated two-service classes scenarios, for different structure of industry competition, i.e., $M = 1, 2, 4, \infty$.

As we expect, the x chosen by an unregulated monopolistic ISP exceeds x chosen by all regulators, thus decisively showing the necessity of regulation, that is, limiting the monopolist's fraction of capacity for premium service h. Let x_0 denotes capacity division chosen by the monopolistic ISP. From (3.9), it is intuitive that:

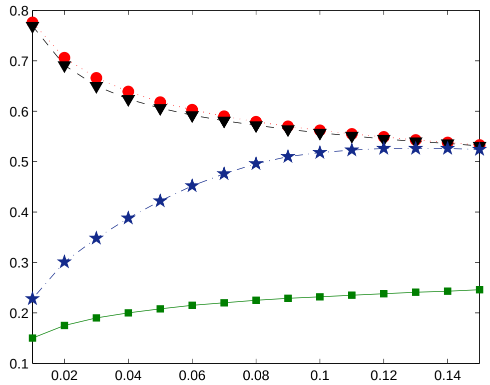
$$\bar{x}_3 \leq \bar{x}_2 \leq \bar{x}_1 \leq x_0.$$

Indeed, the more the regulator cares about the existing users' welfare, the lower is the fraction of capacity that he allocates for the premium service.

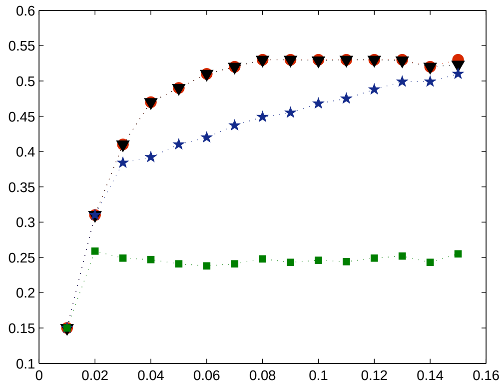
For all other structures of market competition, i.e., $M > 1$, due to competitive effects, the x chosen by the ISPs approaches 0 as capacity cost approaches 0. Except at low capacity costs, for both $M = 2$ and $M = 4$, the ISPs choose an x higher than the one chosen by both the social welfare and user welfare maximizers (regulators 2 and 3). At low capacity costs, we expect that the regulation will not be binding, even when the number of ISPs is low. In such cases, regulation may be unnecessary, but there is no welfare loss if the regulation is imposed.

The range of τ for which competitive ISPs choose a higher x than regulators 2 and 3 decreases with M. In fact, in the limit of $M \to \infty$, except at high capacity costs, no regulation is binding. Note that, in this case, the ISP profits are 0 and hence, $\bar{x}_1$ and $\bar{x}_2$ (chosen by regulators 1 and 2 respectively) coincide.

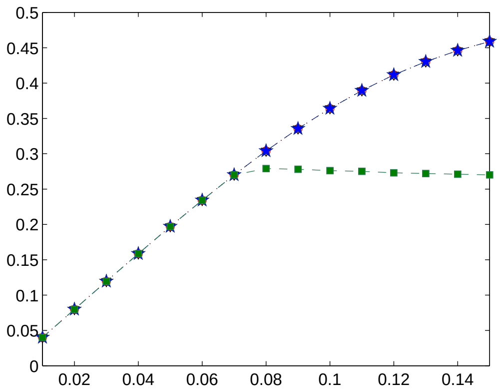
Fig. 3.3 depicts how aggregate capacity varies with τ for $M = 1, 2, 4, \infty$ and $\tau \in [0.05, 0.15]$. Predictably, under all scenarios, the aggregate capacity decreases with τ. Further, when ISPs compete (i.e. Figures 3.3(b) and 3.3(c)), at low τ, the capacity in the unregulated two service class scenario coincides with capacity under regulation since the regulation is not binding. Whenever the regulation is not binding, any restriction on x also



(a) $M = 1$



(b) $M = 2$



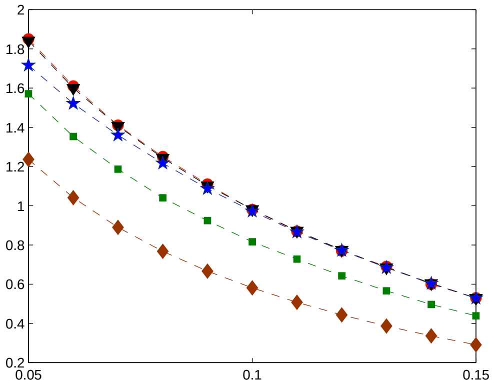
(c) $M = \infty$

Figure 3.2. X as a function of Capacity Cost τ. Legend: Fig. 3.1

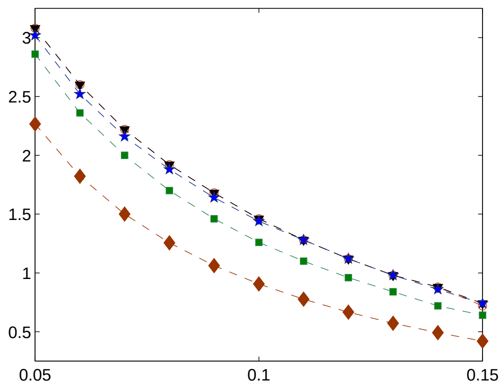
strictly reduces capacity investment. However, irrespective of whether the regulation is binding or not, the capacity in all two-class scenarios always exceeds the capacity in the single service class scenario. Notice that if the ISPs would have chosen x with their capacities fixed at the single service class equilibrium level, our welfare analysis would have been quite different.

Figures 3.4(b) - 3.4(d) depict respectively the values of user welfare, the existing user welfare and the percentage of users who lose surplus due to transition from single- to two service classes. The values are presented at $\tau = 0.05$ for $M = 1, 2, 3, 4, \infty$. In all these figures, the last data point depicts the values for the perfectly competitive ISPs.

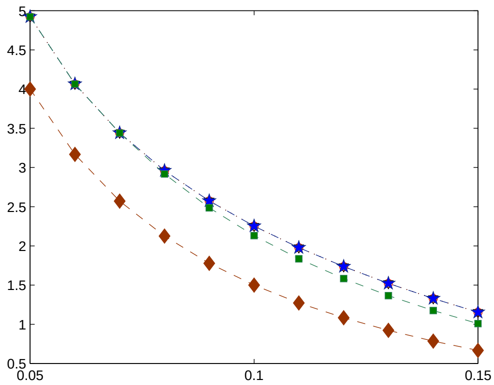
From Figures 3.4(b) and 3.4(c), both, the user welfare and the welfare of existing single-class users, are increasing with competition. For all two-class scenarios, the user welfare is higher than for the single service class scenario. Except for the monopolistic ISP, the existing user welfare is also higher for all two-service class scenarios.



(a) $M = 1$



(b) $M = 2$



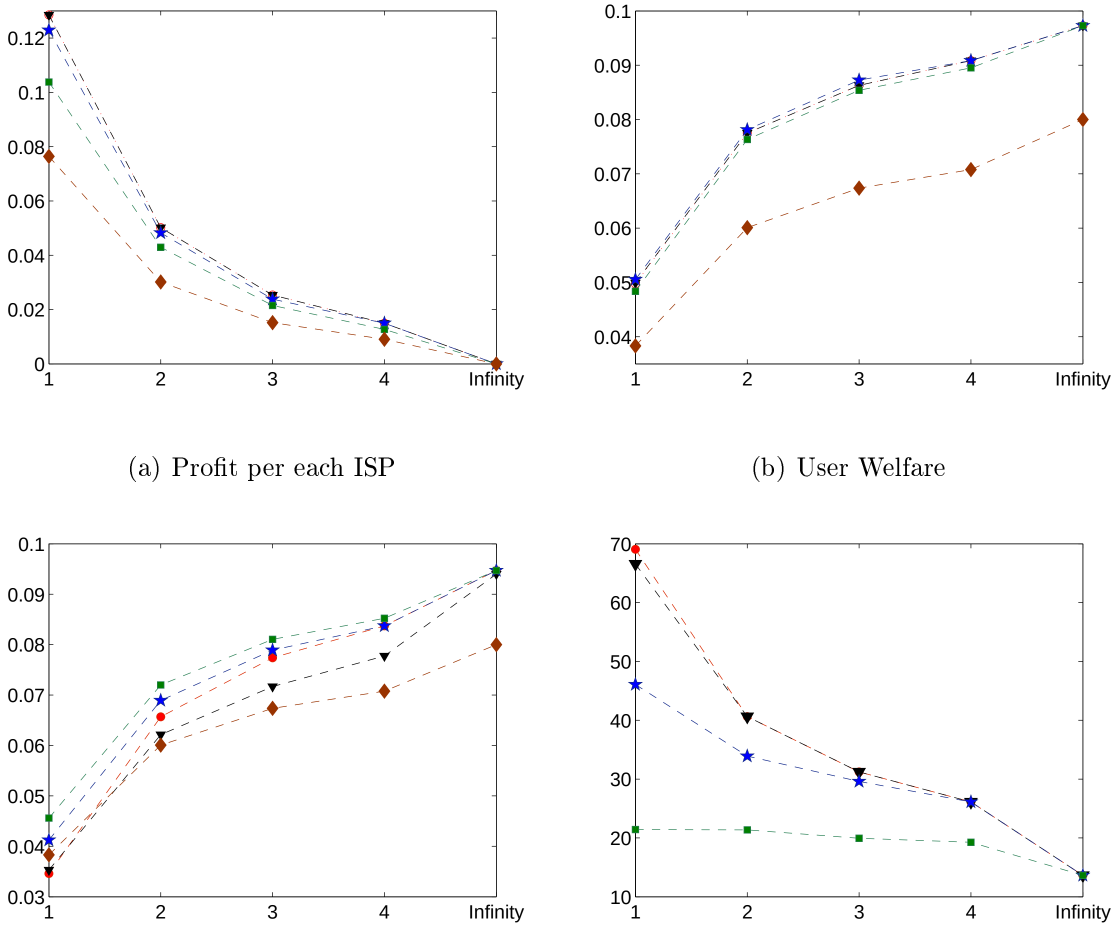
(c) $M = \infty$

Figure 3.3. Aggregate Capacity (c) vs Cost of Capacity (τ). Legend: Fig. 3.1

The percentage of users who lose surplus with two-service classes is depicted in Fig. 3.4(d). We observe that this percentage decreases with competition. This percentage is about 70 for the unregulated monopolist, but even with perfectly competitive ISPs, the percent of surplus losing users remains strictly (and substantially) positive, and exceeds 15%. With regulator 3 (existing user welfare maximizer), the percent of surplus losing users remains under 25% irrespective of ISP competition.

Figures 3.4(a) - 3.4(d) depict respectively the values of per ISP profit, user welfare, the existing user welfare and the percentage of users who lose surplus due to transition from single- to two service classes. The values are presented at $\tau = 0.05$ for $M = 1, 2, 3, 4, \infty$. In all these figures, the last data point depicts the values for the perfectly competitive ISPs.

Fig. 3.4(a) depicts how per ISP profit varies with ISP competition. Intuitively, as competition intensifies, the profit decreases and approaches zero in the limit of perfect competition.



(a) Profit per each ISP

(b) User Welfare

(c) Existing User Welfare

(d) Percent Sufferers

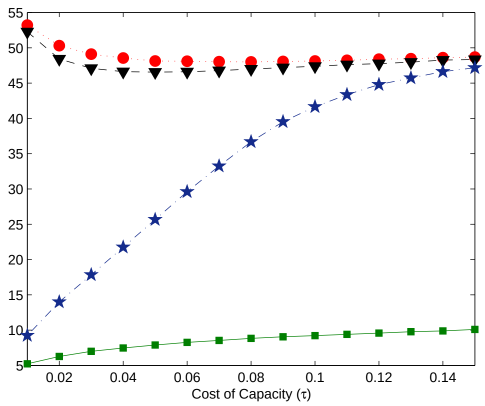
Figure 3.4. Results for various M, with capacity cost fixed at $\tau = 0.05$. Legend: Fig. 3.1

However, under any regulator, for any finite M, per ISP profit with two-service classes is higher than for a single service class.

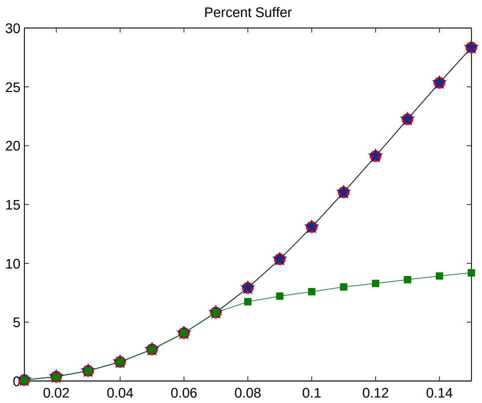
From Figures 3.4(b) and 3.4(c), both, the user welfare and the welfare of existing single-class users, are increasing with competition. For all two-class scenarios, the user welfare is higher than for the single service class scenario. Except for the monopolistic ISP, the existing user welfare is also higher for all two-service class scenarios.

The percentage of users who lose surplus due to the transition to two-service classes is depicted in Fig. 3.4(d). We observe that this percentage decreases with competition. This percentage is about 70 for the unregulated monopolist, but even with perfectly competitive ISPs, the percent of surplus losing users remains strictly (and substantially) positive, and exceeds 15%. With regulator 3 (existing user welfare maximizer), the percent of surplus losing users remains under 25% irrespective of ISP competition.

Let us stress that this welfare analysis is carried out at the equilibrium capacity levels in the single and two class scenarios. These capacity levels are markedly different (see Fig. 3.3). If we perform the analysis at a fixed capacity, the welfare implications will not be the same.



(a) Percent Sufferers Vs τ for $M = 1$



(b) Percent Sufferers Vs τ for $M = \infty$

Figure 3.5. Percent Users with Surplus Loss in the Long Run. Legend: Fig. 3.1

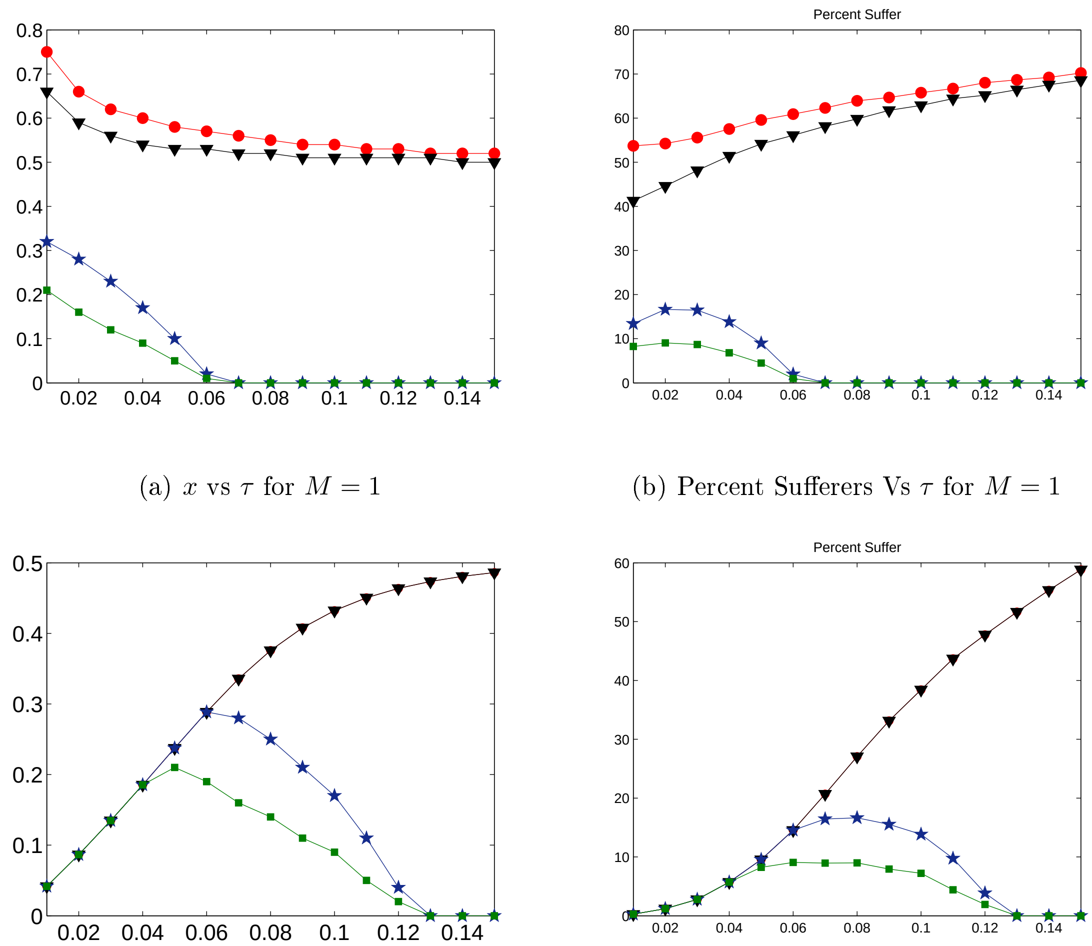
The Short-Run

Fig. 3.6 depicts the short run choice of x and the resulting percentage of existing users who lose surplus as a function of the capacity cost τ for $M = 1, 2, \infty$.

From Figures 3.6(a) and 3.6(c), x for the unregulated ISPs becomes lower with capacity cost which is similar to the long run (see Fig. 3.2). In the short run, when the capacity cost τ becomes high, only regulator 1 (social welfare maximizer), divides capacity, that is regulators 2 and 3 choose $x = 0$. Even when τ in low, and non-zero x becomes optimal for all regulators, for regulators 2 and 3, the short run x is lower than the long run one.

From Figures 3.6(b)and 3.6(d), the percentage of existing single class users who suffer a loss of surplus in the short run (due to the transition to two service classes) is higher than that in the long run (see Fig. 3.5) and closely resembles the graph for the choice of x.

Thus, in both cases, short-run and long-run, a substantial fraction of existing Internet users lose as a result of transition, and the percentage of such surplus losing end-users is higher in the short-run. Therefore, short-run considerations only exacerbate the political economic constraints for transition. Indeed, from our results, in the short-run, in aggregate, the existing end-user welfare gains from transition are lower, and their losses are higher that in the long run. To ease this socially desirable transition to multiple service classes, regulatory intervention could be recommended.



(a) x vs τ for $M = 1$

(b) Percent Sufferers Vs τ for $M = 1$

(c) x vs τ for $M = \infty$

(d) Percent Sufferers Vs τ for $M = \infty$

Figure 3.6. x and Percent Users with Surplus Loss in Short Run. Legend: Fig. 3.1

3.4 Discussion

Our model relies on five key assumptions: (i) irreversibility of investment in capacity, (ii) the ISP commitment to the declared prices, (iii) the uniformity of user type distribution, (iv) a simplified user demand (given by (3.3)), and (v) observability of each ISP's capacity and its division by all ISPs. On one hand, (iii) could be relaxed as our numerical results will work with other distributions as well; on the other hand, such numerical analysis is hard to justify since our uniqueness results require (iii).

From (3.3), we assume that user willingness to pay (highest affordable price (p)) and the lowest acceptable quality (q) are coincident for each user type θ. In general, one expects these requirements to differ. For example, a business user may value the promptness of his e-mail far more than a student user. In fact, one could argue that we have to define a two-dimensional distribution of user types over two separate quantities - willingness to pay and quality requirement. Our simplification, which imposes identical willingness to pay and quality requirements, will be a good description of the case where these distributions are highly correlated.

We assume that ISP capacity costs are identical but our setting permits to consider ISPs with different costs of capacity. Intuitively, when one of the ISP's capacity cost is higher than for others, his capacity investment will be lower. Accordingly, he will allocate a lower capacity fraction to the premium service since the basic service yields higher revenues per unit capacity. Thus, when capacity costs differ, the resulting capacity investment and its division between service classes will also differ.

Further, we made another simplifying assumption that the quality observed by the user depends only on the number of users multiplexed within a given capacity. Though the actual quality observed by the users depends upon the end-to-end variables like delay, jitter, etc., even if a single ISP divides capacity, users might see an improvement in quality. If, in addition, all ISPs coordinate the adoption of two service classes, the user will perceive a marked improvement in quality.

Networking researchers agree that end-to-end issues are a major obstacle to QoS. (These problems relate to both strategic issues associated with surplus-sharing (agency type of conflict) and the coordination problem.) Prior research explains the lack of QoS by the impossibility to price differentiated services when ISPs compete. If this were true, coordination would not matter. To the best of our knowledge, the study in this chapter is the first successful demonstration of QoS pricing for any provider competition. Thus, this study is a necessary step in investigating the coordination problems. But, coordination between ISPs is outside the scope of this thesis and will be the focus of our future work.

3.5 Conclusion

We make the following three contributions to the literature. First, we develop a model for social welfare in a network with two service classes. Second, we investigate the political economic considerations that may constrain the feasibility of adopting the network with QoS. Third, we propose a simple regulatory tool that permits to alleviate the political economic constraints for the network with two service classes.

Our pricing model is based on the network architecture similar to the Paris Metro proposal (PMP) [70]. We extend the model developed in [84] to the case of multiple ISPs. Other closely related papers modeling pricing with PMP network features are [33, 27]. In contrast to these papers, we demonstrate pricing for the network with two service classes for any number of competing ISPs. Thus, from our results, ISP competition per se does not preclude the QoS provision.

Specifically, from our analysis, a network with two service classes is socially desirable, but it could be blocked due to unfavorable distributional consequences. In Section 3.3, we demonstrated that in the absence of regulation and considerable ISP market power (small M), a sizable fraction of the current network users will experience a surplus loss with two

Table 3.1. Table of Notation

Notation	Explanation
p	Price of single service class
p_l	Price of service l
p_h	Price of service h
$\bar{\theta}$	Highest user type adopting the single service
$\underline{\theta}$	Lowest user type adopting the single service
$\bar{\theta}_l$	Highest user type adopting service l
$\underline{\theta}_l$	Lowest user type adopting service l
$\bar{\theta}_h$	Highest user type adopting service h
$\underline{\theta}_h$	Lowest user type adopting service h
N	Total number of users in user base
U	User surplus per user in user base
c	Aggregate capacity per user in user base
c^m	m-th ISP's capacity per user in user base
x^m	m-th ISP's capacity fraction for service h
s^m	m-th ISP's share of aggregate capacity (c^m/c)
c_h^m	m-th ISP's capacity for service h ($c^m x^m$)
c_l^m	m-th ISP's capacity for service l ($c^m(1-x^m)$)
R^m	m-th ISP's revenue per user in user base
Π^m	m-th ISP's profit per user in user base

service classes. Thus, the imposition of regulation, which lowers the fraction of users who lose surplus, improves the feasibility of the two-service class regime.

3.6 Appendix

Proof of Theorem 3.2.1

Let users with types $\theta \in (\underline{\theta}, \bar{\theta}]$ adopt a service with quality q at the price p. Then, from (3.4), the user surplus can be written as

$$U = \frac{1}{2}((\bar{\theta} - p)^2 - (\underline{\theta} - p)^2). \tag{3.23}$$

From (3.3), a user with type θ will adopt the access (service) if and only if $p < \theta \leq q$, where $q = 1 - z/c$, with z being the fraction of users who adopt the service. Clearly, the service is affordable to all users with type $\theta > p$. As more users adopt the service, z increases and q decreases until it becomes equal to the user type at some critical value of θ. Hence, $\underline{\theta}= p$ and $\bar{\theta} = 1 - z/c$ where $z = \bar{\theta}-\underline{\theta}$. Then, we have:

$$\bar{\theta} = 1 - \frac{\bar{\theta} - \underline{\theta}}{c} \iff \bar{\theta} = \frac{c + \underline{\theta}}{c+1} = \frac{c+p}{c+1}, \tag{3.24}$$

and $\underline{\theta} \leq \bar{\theta} \leq 1$ is clearly true from (3.24). Thus, we have determined a non-empty interval of user types who will adopt the service priced at $p \in [0, 1]$.

Since $p =\underline{\theta}$, from (3.23), the user surplus can be written as

$$U = \frac{1}{2}(\bar{\theta} - \underline{\theta})^2.$$

Next, we derive the m'th ISP's optimal capacity and price. From (3.24), for a given c and $p \in [0, 1]$, the provider revenue R^m is

$$R^m = s^m p(\bar{\theta} - \underline{\theta}) = \frac{c^m}{c}\frac{c}{1+c}p(1-p) = \frac{c^m}{1+c}p(1-p).$$

To find the optimal price p, we differentiate w.r.t. p, and get $p = 1/2$. This is the Pareto efficient equilibrium of the subgame $G(\mathbf{c}, \mathbf{x})$ with multiple ISPs. Here, p is independent of c, and thus, for any fixed c, the optimal R^m is

$$R^m = \frac{1}{4}\frac{c^m}{(1+c)}.$$

From the ISP objective, the optimal choice of $c^{m\dagger}$ of the m-th ISP given by

$$c^{m\dagger} = \arg\max_{c^m} \Pi^m = \arg\max_{c^m} \left\{ \frac{1}{4}\frac{c^m}{1+c} - \tau c^m \right\}.$$

Let us express the m-th ISP objective as:

$$\max_{c^m} \frac{c^m}{4(1 + c^{-m} + c^m)} - \tau c^m,$$

where

$$c^{-m}=\sum_{j\neq m} c^j \text{ and } c=\sum_j c^j \text{ and thus } c = c^{-m} + c^m.$$

To find $c^{m\dagger}$, we differentiate the expression above w.r.t. c^m and equate it to 0 (to obtain the m-th ISP FOC):

$$\frac{\partial}{\partial c^m}\left\{ \frac{1}{4}\frac{c^m}{1+c} - \tau c^m \right\} = 0,$$

$$\frac{1}{4}\left[\frac{1}{1+c} - \frac{c^m}{(1+c)^2} \right] = \tau,$$

from which for any two ISPs, $m1$ and $m2$ we have:

$$\frac{1}{4}\left[\frac{1}{1+c}-\frac{c^{m1}}{(1+c)^2}\right]=\frac{1}{4}\left[\frac{1}{1+c}-\frac{c^{m2}}{(1+c)^2}\right].$$

Thus, we have $c^{m1} = c^{m2}$, that is, in any equilibrium, the ISPs' investments are identical. Thus, we have proven that any equilibrium is symmetric.

Next, we use this to rewrite the FOCs as

$$\frac{1+(M-1)c^m}{(1+Mc^m)^2}=4\tau,$$

which we solve to express equilibrium capacities $c^{m\dagger}$ in the game of M ISPs and capacity cost τ:

$$c^{m\dagger}(M,\tau)=\frac{(M-1)+\sqrt{(M-1)^2+16\tau M}}{8\tau M^2}-\frac{1}{M}. \tag{3.25}$$

Thus, there exists a unique equilibrium of this game, and aggregate equilibrium capacity $c^\dagger$ is:

$$c^{\dagger}(M,\tau)=\frac{(M-1)+\sqrt{(M-1)^2+16\tau M}}{8\tau M}-1. \tag{3.26}$$

In equilibrium, $p^\dagger = \frac{1}{2}$ and all ISPs invest equality, with ISP equilibrium investments given by (3.25), and aggregate investment given by (3.26).

From (3.26), aggregate capacity increases with M. Since equilibrium price is identical for all M, and capacity increases with M, number of served users and service quality also increase with M, and Theorem 3.2.1 is proven.

Proof of Theorem 3.2.2

We start with the following Lemmas:

Lemma 3.6.1 *All users with type $\theta \in (\underline{\theta}_l, \bar{\theta}_l]$ adopt service l, where $\underline{\theta}_l = p_l$ and $\bar{\theta}_l = \frac{p_l+c_l}{1+c_l}$.*

Proof From (3.3), given a choice between two different *affordable* ($\theta > p$) and *acceptable* ($\theta \le q$) services, a user always chooses a cheaper service. Same as in the single service class case, for a given price p_l and capacity c_l, users with type $\theta \in (\underline{\theta}_l, \bar{\theta}_l]$ adopt the service l. Introducing a service h priced at $p_h > p_l$ has no effect on the users of service l; no such users will shift to service h. Hence, from (3.24), we have $\underline{\theta}_l = p_l$ and $\bar{\theta}_l = \frac{p_l+c_l}{1+c_l}$ and Lemma 3.6.1 is proven.

Users with type $\theta \in (\underline{\theta}_h, \bar{\theta}_h]$ adopt the service h. From Lemma 3.6.1, any $p_h > p_l$ does not affect $\underline{\theta}_l$ and $\bar{\theta}_l$. This means that $\underline{\theta}_h \ge \bar{\theta}_l$.

Lemma 3.6.2 *For any given p_l, the ISP's revenue is maximized at $p_h \geq \bar{\theta}_l$.*

Proof Assume to the contrary that $p_h < \bar{\theta}_l$. Accordingly,

$$\underline{\theta}_h = \bar{\theta}_l \text{ and } \bar{\theta}_h = 1 - z_h/c_h = 1 - \frac{\bar{\theta}_h - \underline{\theta}_h}{c_h}$$

giving $\bar{\theta}_h = \frac{\bar{\theta}_l + c}{1+c}$. Hence $z_h = \frac{c_h}{1+c_h}(1 - \bar{\theta}_l)$ is independent of p_h. Thus, the ISP's revenue from capacity c_h is equal to $p_h z_h$. Consider $\tilde{p} = \frac{p_h + \bar{\theta}_l}{2}$. Then, we have: $p_h < \tilde{p} < \bar{\theta}_l$ and thus, $p_l < \tilde{p} < \bar{\theta}_l$, implying that z_h remains the same. Since $\tilde{p} > p_h$, we have $\tilde{p} z_h > p_h z_h$, which leads to a higher revenue and contradicts our assumption and concludes the proof of Lemma 3.6.2.

Lemma 3.6.3 *In the ISP's optimum: $\bar{\theta}_l = \underline{\theta}_h = p_h$.*

Proof Formally, assume the contrary, i.e., assume that $\underline{\theta}_h = p_h > \bar{\theta}_l$. Let $\delta > 0$ and $p_h = \bar{\theta}_l + \delta$. Then, the ISP revenue (by summing up the revenues from the two classes) is

$$\begin{aligned} R &= \frac{c_l}{1+c_l} p_l(1-p_l) + \frac{c_h}{1+c_h} p_h(1-p_h) \\ &= \frac{c_l p_l(1-p_l)}{1+c_l} + \frac{c_h}{1+c_h}\left[\frac{c_l+p_l}{1+c_l} + \delta\right]\left[\frac{1-p_l}{1+c_l} - \delta\right]. \end{aligned} \tag{3.27}$$

Maximizing R w.r.t. δ gives an optimal $\delta^* = \frac{1-c_l-2p_l}{2(1+c_l)}$. Since we assumed $\delta > 0$, we have $\delta^* > 0$, thus giving

$$p_l < \frac{1-c_l}{2}. \tag{3.28}$$

We insert δ^* into (3.27) to obtain

$$\begin{aligned} R &= \frac{c_l}{1+c_l} p_l(1-p_l) + \frac{c_h}{1+c_h}\left[\frac{1+c_l}{2(1+c_l)}\right]^2 \\ &= \frac{c_l}{1+c_l} p_l(1-p_l) + \frac{1}{4}\frac{c_h}{1+c_h}, \end{aligned}$$

which is maximized at $p_l = 1/2$. But, this requires $-c_l > 0$, which contradicts (3.28). Thus, the assumption $\delta > 0$ is false and we have $p_h = \bar{\theta}_l = \underline{\theta}_h = \frac{c_l+p_l}{1+c_l}$, and Lemma 3.6.3 is proven.

From Lemmas 3.6.1 - 3.6.3, we obtain:

$$\underline{\theta}_h = \bar{\theta}_l = p_h = \frac{c_l + p_l}{c_l + 1} \text{ and } \bar{\theta}_h = \frac{c_h + p_h}{c_h + 1},$$

where c_l and c_h are given by (3.8). We can, therefore, express the ISP revenue as

$$\begin{aligned} R(c, x, p_l) &= \frac{c_l}{1+c_l} p_l(1-p_l) + \frac{c_h}{1+c_h} p_h(1-p_h) \\ &= \frac{(1-p_l)\left[A p_l + c_h c_l\right]}{B}, \end{aligned} \tag{3.29}$$

where $A = (1+c_l)(1+c_h)c_l+c_h$ and $B = (1+c_l)^2(1+c_h)$. Thus, the ISP revenue maximization problem can be expressed as an optimization in just one variable p_l. Differentiating this expression for revenue with respect to p_l and equating to zero, we get $A(1-p_l)-(Ap_l+c_hc_l) = 0$, which gives us p_l as a function of c and x:

$$p_l(c,x) = \frac{1}{2} - \frac{c_hc_l}{2A} = \frac{1}{2} - \frac{c_hc_l}{2[(1+c_l)(1+c_h)c_l+c_h]}, \tag{3.30}$$

and Theorem 3.2.2 is proven.

Proof of Corollary 3.2.4

Proof From (3.30) we have $p_l(c,x) < \frac{1}{2}$, and we obtain p_h as:

$$p_h(c,x) = \frac{p_l + c_l}{1+c_l} = \frac{1}{2} + \left[\frac{1}{2}\frac{(1+c_h)c_l^2}{(1+c_l)(1+c_h)c_l+c_h}\right].$$

The expression in the square brackets is obviously positive and hence $p_h(c,x) > 1/2$, and Corollary 3.2.4 is proven.

Proof of Theorem 3.2.5

Substituting (3.30) into (3.29), and using $c_l = (1-x)c$ and $c_h = xc$, we get

$$R(c,x) = \frac{((1+xc)(1-x)+x)^2c}{4[(1+(1-x)c)(1+xc)(1-x)+x](1+xc)}. \tag{3.31}$$

To prove the theorem, we show that $R(c,x)$ is maximized at a unique x. Since $R(c,x)$ is differentiable w.r.t. x, it is sufficient to prove that $\frac{\partial R(c,x)}{\partial x} = 0$ at a single interior x and this point is a maximum. Let

$$B = ((1+xc)(1-x)+x)^2c = (1+xc-x^2c)^2c$$

and let

$$\begin{aligned} D &= 4[(1+(1-x)c)(1+xc)(1-x)+x](1+xc) \\ &= 4[1+c+x(1-x)(2-x)c^2+x^2(1-x)^2c^3] \end{aligned}$$

Now, $R(c,x) = B/D$ and hence, differentiating w.r.t. x gives

$$\frac{\partial R(c,x)}{\partial x} = \frac{D\frac{\partial B}{\partial x} - B\frac{\partial D}{\partial x}}{D^2},$$

where the denominator is always positive. Hence, we focus on the numerator $D\frac{\partial B}{\partial x} - B\frac{\partial D}{\partial x}$ only. We find

$$\frac{\partial B}{\partial x} = 2(1 + xc - x^2c)(1 - 2x)c^2 \tag{3.32}$$

and

$$\frac{\partial D}{\partial x} = 4[(2 - 6x + 3x^2)c^2 + 2x(1 - x)(1 - 2x)c^3]. \tag{3.33}$$

From (3.32) and (3.33), after simplifying, we have

$$D\frac{\partial B}{\partial x} - B\frac{\partial D}{\partial x} = 4(1 + xc - x^2c)c^2[2 - 4x + (2x - 3x^2)c + x^3(1 - x)c^2]. \tag{3.34}$$

Note that $\forall x \in [0, 1], \forall c > 0$, we have $4(1 + xc - x^2c)c^2 > 0$. Hence, this term does not contribute any zeroes. Hence, if the second term (say $\zeta(x)$) has a single zero in [0,1] and is strictly positive in a small interval close to 0, then $\frac{\partial R(c,x)}{\partial x}$ has a unique maximum in $[0, 1]$.

$$\begin{aligned}\zeta(x) &= 2 - 4x + (2x - 3x^2)c + x^3(1 - x)c^2 \\ &= 2 + (2c - 4)x - 3cx^2 + c^2x^3 - c^2x^4.\end{aligned}$$

It is clear that $\zeta(0) = 2 > 0$. Next, we show that $\zeta(x)$ has a single zero in [0,1]. $\zeta(x)$ is a fourth order equation in x with $\zeta(0) > 0$ and $\zeta(1) = -2 - 2c < 0$. Hence, $\zeta(x)$ either has 1 root or 3 roots in [0,1]. Now,

$$\zeta'(x) = 2c - 4 - 6cx + 3c^2x^2 - 4c^2x^3; \tag{3.35}$$

$$\zeta''(x) = -6c + 6c^2x - 12c^2x^2. \tag{3.36}$$

For $\zeta(x)$ to have 3 roots in [0,1], $\zeta'(x)$ must have at least 2 roots and $\zeta''(x)$ must have at least 1.

Case 1: $0 < c < 8$
$\zeta''(x) = -6c + 6c^2x - 12c^2x^2 = -6c + 6c^2x(1 - 2x) \leq -6c + 6c^2/8 = 6c(c/8 - 1)$. For $0 < c < 8$, $6c(c/8 - 1) < 0$ giving us $\zeta''(x) < 0 \; \forall x \in [0, 1]$. Hence, for $c < 8$, $\zeta(x)$ has exactly one root in [0,1], by the strict concavity of $\zeta(x)$.

Case 2: $c \geq 8$
From case 1, we can rewrite $\zeta''(x)$ as $\zeta''(x) = -6c + 6c^2x(1 - 2x)$. For $x \in [1/2, 1]$, $\zeta''(x) < 0$. Further, $\zeta(1/2) = c/4 + c^2/16 > 0$ and $\zeta(1) = -2 - 2c < 0$ which implies that there is *exactly one root* in [1/2,1] by the strict concavity of $\zeta(x)$.

Next, we need to show that there are no roots of $\zeta(x)$ in [0,1/2]. We will prove this by showing that $\zeta'(x) \geq 0$ in [0,1/2], i.e., $\zeta(x)$ is non-decreasing in [0,1/2]. Since $\zeta(0) = 2 > 0$, this implies $\zeta(x) > 0 \; \forall x$ in [0,1/2], and hence it cannot have any roots in [0,1/2].

– Consider the interval [0,1/4].
Since $6c^2x(1 - 2x) > 0$, $\zeta''(x) > -6c$ in [0,1/4]. Hence, the fastest rate at which $\zeta'(x)$ decreases is $-6c$ anywhere in this interval. We know that $\zeta'(0) = 2c - 4 > 0$ for $c \geq 8$. Hence $\forall x \in [0, 1/4]$,
$\zeta'(x) \geq \zeta'(0) - 1/4 \times 6c = 2c - 4 - 6c/4 = c/2 - 4 \geq 0$
for $c \geq 8$.

– Consider the interval [1/4,1/2].

Since $\zeta''(x) = -6c + 6c^2x(1-2x)$, there is exactly one root of $\zeta''(x)$ in [1/4,1/2]. (The two roots are $\frac{1}{4} \pm \frac{\sqrt{1-\frac{8}{c}}}{4}$.) Now, $\zeta''(1/4) = -6c + 6c^2/8 \geq 0$ and $\zeta''(1/2) = -6c < 0$ for $c \geq 8$ which means that $\zeta'(x)$ first increases and then decreases in [1/4,1/2].

$\zeta'(1/4) = 2c - 4 - 6c/4 + 3c^2/4^2 - 4c^2/4^3 = c/2 - 4 + c^2/8 \geq 0$ and

$\zeta'(1/2) = 2c - 4 - 6c/2 + 3c^2/2^2 - 4c^2/2^3 = -c - 4 + c^2/4 > 0$ for $c \geq 8$. Hence, though $\zeta'(x)$ decreases in some interval in [1/4,1/2], it never goes to 0, giving us the required result that $\forall x \in [1/4, 1/2]$, $\zeta'(x) \geq 0$.

Corollary 1: The monopoly profits increase monotonously for $x < x(c,1)$. **Proof** $\zeta(x)$ is non-decreasing in [0,1/2] (since $\zeta'(x) \geq 0$ for $x \in [0, 1/2]$). Combining with $\zeta(0) > 0$, we have $\zeta(x) > 0$ for $x < x(c,1)$. This implies that $\frac{\partial R(c,x)}{\partial x} > 0$ for $x < x(c,1)$, which gives us our result.

Corollary 2: For any fixed c, the monopolist chooses

$$x(c,1) > 1/2. \tag{3.37}$$

Proof $\zeta(1/2) = c/4 + c^2/16 > 0$ and $\zeta(1) = -2 - 2c < 0$ which implies there is at least one root in (1/2,1). From Theorem 3.2.5, we know that the root in [0,1] is unique. Hence, this root must lie between 1/2 and 1. Thus, $\frac{\partial R(c,x)}{\partial x} = 0$ for a unique $x \in (1/2, 1)$, giving us our desired result.

Proof of Theorem 3.2.6

Let c^m be sunk, and consider the m-th ISP's choice of x^m. In optimum:

$$\frac{dR^m}{dx^m} = 0 \text{ or } \frac{dR^m_l}{dx^m} = -\frac{R^m_h}{dx^m}.$$

Since, under perfect competition, each ISP is too small to affect aggregate capacity division (c_l and c_h) and prices (p_l and p_h) we obtain:

$$\frac{p_l(1-p_l)}{1+c_l}\frac{dc^m_l}{dx^m} = -\frac{p_h(1-p_h)}{1+c_h}\frac{dc^m_h}{dx^m}$$

and using $c^m_h = x^m c^m$ and $c^m_l = (1-x^m)c^m$ we have:

$$\frac{p_l(1-p_l)}{1+c_l} = \frac{p_h(1-p_h)}{1+c_h}.$$

That is, for any ISP, average return on investment is equal in both service classes:

$$\frac{R_l}{c_l} = \frac{R_h}{c_h}, \tag{3.38}$$

where the subscript m is dropped to simplify. Since under perfect competition each ISP profit is zero:

$$\Pi = R_l + R_h - \tau(c_l + c_h) = 0,$$

we combine with (3.38), which gives us zero profit in each service class, from which

$$\Pi_i = \frac{R_i}{c_i} - \tau = 0,$$

$$\text{i.e., } \frac{R_i}{c_i} = \frac{1}{1+c_i} p_i(1-p_i) = \tau. \tag{3.39}$$

Using the result of Theorem 3.2.2 that $p_h(c,x) = \frac{c_l + p_l(c,x)}{1+c_l}$ (equation (3.17)) we infer

$$\frac{1}{1+c_l} p_l(1-p_l) = \frac{1}{1+c_h} \frac{(p_l+c_l)}{1+c_l} \frac{(1-p_l)}{1+c_l},$$

$$\text{i.e., } (1+c_l)(1+c_h)p_l = (p_l + c_l),$$

which, for any c, gives

$$p_l(c,x) = \frac{1-x}{1+(1-x)xc}, \tag{3.40}$$

and we substitute this into (3.39):

$$\frac{1}{1+c_l} p_l(1-p_l) = \tau$$

to obtain the equation connecting c, x and τ:

$$(1-x)x = \tau \left[1 + (1-x)xc\right]^2. \tag{3.41}$$

Next, we use (3.40) and equate it with (3.17) for the equilibrium price in a general case (for any M), to obtain the relation between c and x:

$$p_l = 1/2 - \frac{c_l c_h}{2\left[(1+c_l)(1+c_h)c_l + c_h\right]} = \frac{1-x}{1+(1-x)xc},$$

and collecting all the terms we have:

$$(1-2x)(1+(1-x)c) - (1-x)^2 x^2 c^2 (1+(1-x)c) = 0.$$

We divide by $1+(1-x)c$ (which must be positive) and obtain the expression for c in terms of x:

$$(1-2x) - (1-x)^2 x^2 c^2 = 0,$$

$$c = \frac{\sqrt[2]{1-2x}}{x(1-x)}. \tag{3.42}$$

From (3.42), we see that $x \in (0, 1/2)$. For $x \in (0, 1/2)$, the numerator in the RHS of (3.42) strictly decreases while the denominator strictly increases. This implies that c is strictly decreasing for $x \in (0, 1/2)$ and takes all values between 0 and ∞. Hence, one can define an inverse function for x as a function of $c \in (0, \infty)$. Thus, we shown that, with perfect competition, for all $c > 0$, there exists a unique x in the equilibrium of the game $G(\mathbf{c})$ and

$$x(c, \infty) < 1/2. \tag{3.43}$$

Substituting (3.42) into (3.41), we get:

$$(1-x)x = \tau \left[1 + \sqrt[2]{1-2x}\right]^2,$$

which permits only a unique $x \in (0, 1/2)$ for any τ. We solve this equation numerically to obtain this unique x.

Proof of Theorem 3.2.7

Under the assumption of symmetry, we show that there exists a unique optimal $\tilde{x} = x(c, M)$ at which the ISPs reach maximum profit sustainable in the equilibrium of our game $G(c)$. $\tilde{x} \in (x(c, \infty), x(c, 1))$, where $x(c, \infty)$ and $x(c, 1)$ are the equilibrium x in the game $G(c)$ under perfect competition and a monopolist respectively. From (3.37) and (3.43), the interval $(x(c, \infty), x(c, 1))$ is non-empty, and from Theorems 3.2.5 and 3.2.6, the fractions $x(c, 1)$ and $x(c, \infty)$ are unique. Consider the subgame in which the ISPs have already sunk their investments and aggregate capacity is fixed at c.

To sum the proof, we evaluate $\frac{dR^m}{dx^m}$ at $x = x(c, \infty)$, and show it is positive:

$$\left.\frac{dR^m}{dx^m}\right|_{x(c,\infty)} > 0, \tag{3.44}$$

from which we will have $\tilde{x} > x(c, \infty)$. Next, we evaluate $\frac{dR^m}{dx^m}$ at $x = x(c, 1)$ and show it is negative:

$$\left.\frac{dR^m}{dx^m}\right|_{x(c,1)} < 0, \tag{3.45}$$

from which we have $\tilde{x} < x(c, 1)$.

$$R^m = (1 - x^m)c^m \frac{p_l(1-p_l)}{1+c_l} + x^m c^m \frac{p_h(1-p_h)}{1+c_h}.$$

We notice that for any $f(c_l, c_h)$

$$\frac{df(c_l, c_h)}{dx^m} = \frac{1}{M}\frac{df(c_l, c_h)}{dx},$$

from which, for $i = l, h$,

$$\frac{d}{dx^m}\frac{p_i(1-p_i)}{1+c_i} = \frac{1}{M}\frac{d}{dx}\left\{\frac{p_i(1-p_i)}{1+c_i}\right\}. \tag{3.46}$$

We differentiate R^m with respect to x^m to obtain:

$$\frac{dR^m}{dx^m} = \frac{c}{M}\left[fh - fl\right] + \frac{c}{M}\left[(1-x)\frac{d\left\{fl\right\}}{dx^m} + x\frac{d\left\{fh\right\}}{dx^m}\right], \tag{3.47}$$

$$\text{where } fh = \frac{p_h(1-p_h)}{1+c_h} \text{ and } fl = \frac{p_l(1-p_l)}{1+c_l},$$

and we use the fact that in a symmetric equilibrium $c^m = \frac{c}{M}$.

We use (3.46) to infer:

$$\frac{dR^m}{dx^m} = \frac{c}{M}\left(\left[fh - fl\right] + \frac{1}{M}\left[(1-x)\frac{d\left\{fl\right\}}{dx^m} + x\frac{d\left\{fh\right\}}{dx^m}\right]\right). \tag{3.48}$$

When $M = 1$, the FOC is:

$$\left.\frac{dR}{dx}\right|_{x(c,1)} = c\left[fh - fl\right] + c\left[(1-x)\frac{d\left\{fl\right\}}{dx^m} + x\frac{d\left\{fh\right\}}{dx^m}\right] = 0, \tag{3.49}$$

and we have proven in Theorem 3.2.5 that

$$\left.\frac{dR}{dx}\right|_{x<x(c,1)} > 0. \tag{3.50}$$

For perfect competition ($M \to \infty$), the term inside the second square bracket of (3.48) could be ignored, which gives the FOC:

$$\left.\frac{dR^m}{dx^m}\right|_{x(c,\infty)} = \left[fh - fl\right] = 0. \tag{3.51}$$

Since $x(c,\infty)$ is unique, for $M \to \infty$, for the m-th ISP, $\left.\frac{dR^m}{dx^m}\right|_{x>x(c,\infty)} < 0$. Thus,

$$\left.\frac{dR^m}{dx^m}\right|_{x>x(c,\infty)} = \left[fh - fl\right] < 0. \tag{3.52}$$

From (3.49), (3.50) and (3.51), at $x = x(c,\infty)$, we infer that

$$\left.\frac{dR}{dx}\right|_{x(c,\infty)} = \left[(1-x)\frac{d}{dx}\left\{fl\right\} + x\frac{d}{dx}\left\{fh\right\}\right] > 0.$$

Hence, the term inside the second square bracket in (3.47) is positive for any finite M too. Thus, for any finite $M > 1$, (3.44) is proven.

Next, from (3.49), we have $\left.\frac{dR}{dx}\right|_{x=x(c,1)} = 0$ for a monopolist. From (3.52), the first square bracket in (3.49) is negative, which implies that the second square bracket is positive, giving us:

$$\left.\left[(1-x)\frac{d\left\{fl\right\}}{dx^m} + x\frac{d\left\{fh\right\}}{dx^m}\right]\right|_{x(c,1)} = -\left.\left[fh - fl\right]\right|_{x(c,1)} > 0.$$

Thus, for any finite M, the terms inside the square brackets in (3.47) are equal and opposite in sign. However, the positive term inside the second square bracket is multiplied by $1/M$, making the positive component lower than the negative component for $M > 1$, which ends the proof for (3.45).

Therefore, from continuity of the underlying functions in the m-th ISP's FOC wrt x^m, for any finite $M > 1$, we infer

$$\left.\frac{dR^m}{dx^m}\right|_{x=\tilde{x}} = 0, \text{ where } \tilde{x} \in (x(c,\infty), x(c,1)).$$

From Theorem 3.2.5, aggregate ISP profit (and thus, due to symmetry, each ISP profit) increases with $\tilde{x}$. This gives us a unique $\tilde{x}$ at which profit is maximal for a fixed c. Thus, the symmetric Pareto efficient equilibrium in the game of M competing ISPs is unique, and Theorem 3.2.7 is proven.

Chapter 4

Missing Market - Security Insurance

Today, the Internet serves as the primary communication platform for both individuals and businesses. At present, due to the nearly universal connectivity, a huge amount of wealth is accessible online and the Internet has become a preferred destination for criminals. However, the Internet, which was originally conceived to be an academic network, has failed to address many of these security problems. Due to the ease of accessibility and programmability, unwary end users' computers are routinely infected with malware. These infected computers could be employed for future crimes, resulting in an interdependent security environment.

Technology-based defense and enforcement solutions are available, but a consensus among security researchers [7] is that the existing security problems cannot be solved by technological means alone. Indeed, these security problems primarily result from misaligned incentives of the networked parties with respect to their security. Users under-invest in security since they do not bear the true societal costs of their actions, which causes a negative externality.

Existing research [13, 34, 67, 91, 81, 15] indicates that *risk management* in general and cyber-insurance in particular are potentially valuable tools for security management. This chapter focuses on the effects of cyber-insurers on network security and user welfare, in a general setting with interdependent security and asymmetric information between users and insurers. We believe that these features of the environment induce socially suboptimal network security, and complicate the management of security risks.

In our model, all users are identical. Their wealth is identical and they suffer identical damage if cyber-attack on them is successful. The user's probability of being attacked depends on both the *user security level* and the *network security level*, which individual users take as given. Thus, there is an externality causing individually optimal user security level to be lower than the socially optimal one.

First, we investigate the effects of information asymmetry in the setting with interdependent security. Though our model allows to study both moral hazard (when insurers are not aware of user security levels) and adverse selection (when insurers cannot distinguish different user types), in this chapter, we address only moral hazard (see [83] for analysis of adverse selection). We find that cyber-insurance fails to improve the network security level though it may improve user utility, if an equilibrium exists. Second, we assume no information asymmetry between the insurers and the users. We demonstrate that user utility is higher with insurance, but surprisingly, even in this case, the network security level is not necessarily higher. On reverse, for a substantial range of parameters, network security worsens with insurers.

Our assumption of identical users is simplistic, and does not hold in the actual Internet. But, we argue that adding user and insurer heterogeneity to our setting only *increases* informational asymmetries. If insurers could separate users of different types, our results hold for every class of user types in such a heterogeneous environment. If insurers are unable to distinguish between users with different types, the problem of adverse selection arises due to which missing markets are likely, as [76] demonstrate. Finally, the presence of different insurer types also brings the "lemon problem" [5], another manifestation of adverse selection, which also lead to missing markets. Hence, our results will continue to hold in a heterogeneous environment as well.[1]

The chapter is organized as follows. In Section 4.1, we describe the related work. In Section 4.2, we propose and analyze the base model. In Section 4.3, we add competitive insurers to our base model, and consider two cases: with non-contractible individual security levels, and with required individual security level included into user's insurance contract. In Section 4.4, we discuss the intuition behind our results. In Section 4.5, summarize our findings and conclude. The technical details are relegated to Appendix.

4.1 Related Work

Cyber-insurance is complicated by the specific environment of the Internet, with correlation and interdependence being the two major factors precluding its existence. Effects of correlated risks on cyber-insurance were considered by [13, 12]. The presence of an OS monoculture in the Internet causes cyber-attacks which impose correlated damages on Internet systems. Such correlated damages result in higher insurance premiums, making cyber-insurance unviable for most institutions.

Bohme (2005) [13] derives the conditions under which the cyber coverage is viable despite the monoculture of installed platforms. Specifically, he finds that a potential market exists when clients are highly risk averse, and loss probability is huge. Bohme and Kataria (2006) [12] introduce two tiers of cyber risk correlations. First, the internal correlation, the correlation of cyber risks within a firm (i.e., a correlated failure of multiple systems on the internal network), and second, the global correlation – correlation in risk at a global level, that is correlation in the insurer's portfolio. They demonstrate that the conditions for existence of a cyber insurance market are high internal and low global risk correlation.

[1]See [83] where we extend our model to address moral hazard.

In this chapter, we focus on interdependent security. We build on the seminal ideas about the role of information in insurance markets [76],[2] which we combine with the ideas of interdependent security originated by [57, 35, 40].[3] Although security interdependence is present in other contexts (such as terrorist attacks [58]), network security is especially prone to these effects because many millions of users are interlinked. Previous work dealing with interdependent security include [72, 45, 63].

Ogut et.al. (2005) [72] provide a comprehensive analysis of effects of cyber insurance on incentives for IT security investment in the presence of interdependent risks, but assume no information asymmetry between the insurers and the insured. They demonstrate that with interdependent risks, security investments are lower than when no interdependence is present. They suggest that, with a high degree of interdependence, even with increased competition in the insurance market, prices may not necessarily fall – because firms use insurance rather than investment to manage security risks. The main difference of our study is that we address the question of information asymmetry between the insurers and insured and calculate the premiums for perfectly competitive insurance market in the presence of moral hazard.

Hofmann (2005) [45] considers interdependent security, with a continuum of heterogeneous agents, who differ in their cost of security investment (prevention cost), which affects the probability of loss. If an agent invests in prevention, he reduces own loss (direct loss), but the indirect loss caused by interdependence with other networked agents remains. In no-insurance equilibrium, the agents with low costs of prevention invest while those with high cost of prevention do not. Lelarge-Bolot (2009) [63] also use a model similar to [45]. This setting is markedly different from ours. Their user choice is binary – users either self-protect or not, but users differ in their costs of self-protection. In our model, though our users have an identical cost function, they choose the amount of investment and the degree of self-protection.

Hofmann's paper is somewhat different from other papers on cyber insurance, because (as far as we are aware) this study is unique by considering such a strong tool as compulsory cyber insurance. The chapter demonstrates that, even with compulsory insurance, competitive insurers cannot achieve socially optimal prevention level. Hofmann proves that social optimum can be achieved by a monopolist insurer who engages in premium discrimination and acts "like a social planner". However, unlike our study, she does not directly tackle moral hazard because she assumes that the insurer can perfectly observe whether a user invests in protection. Lelarge-Bolot (2009) consider only the contracts with full coverage. This makes it hard to address moral hazard, and they admit to leave the design of good incentives with moral hazard for future research. In our chapter, we address this question of moral hazard and derive the level of coverage offered in equilibrium.

[2]See [94] for the literature review.

[3]See also [34, 36, 97, 72, 10, 31, 47, 45, 63, 75]. This list is by no means exhaustive. See [14]

4.2 Model

We consider a network populated by N homogeneous (i.e., identical) users, each of whom possesses a wealth $W > 0$. In the absence of network security problems, user i utility U_i is:

$$U_i = f(W),$$

where the function f is increasing and concave ($f' > 0$, and $f'' \leq 0$), reflecting that user wealth W has a positive but decreasing marginal benefit for the user.

In the presence of network security problems, we assume that a user i incurs a monetary damage $D \in (0, W)$ when he is successfully attacked, and we let p_i denote the probability of successful attack. We assume that the probability p_i depends on two factors: the *security level* $s_i \in [0, 1]$ chosen by user i and the *network security level* $\bar{s} \in [0, 1]$, which depends on the security choices of all network users. We define the network security level $\bar{s}$ as the average security level in the network:

$$\bar{s} = \frac{1}{N} \sum_{i=1,\ldots N} s_i.$$

Further, we assume N to be large, i.e., each user has a negligible effect on $\bar{s}$ and takes the network security level as given. Then, we define the probability p_i of a successful attack on user i as

$$p_i = (1 - s_i)(1 - \bar{s}),$$

where the second term $(1 - \bar{s})$ can be viewed as the probability of an attack in the network and the first term $(1 - s_i)$ can be viewed as the probability of success of such an attack on user i.

We assume that, for any user i, achieving individual security s_i entails a cost $h(s_i)$. We let h be an increasing convex function ($h', h'' \geq 0$), with $h(0) = 0$ representing a completely insecure user and $h(1) = \infty$ characterizing the costs required to maintain a "perfectly secure" system. The intuition is that user security costs increase with security, and that improving security level imposes an increasing marginal cost on the user. Additionally, for expositional convenience, we impose $h'(0) = 0$, to ensure positive user investments $s_i > 0$. We assume that the user cannot modify his D by changing his investment h, i.e., users do not self-insure their damages. For e.g., users may backup their data to prevent loss of information. Such self-insurance does not have an externality effect on other users since the advantages of that investment are observed by the user alone [36]. One can also view our D as the residual damages after self-insurance.

Thus, the expected user utility can be expressed as

$$U_i = (1 - p_i) \cdot f(W) + p_i \cdot f(W - D) - h(s_i). \tag{4.1}$$

To simplify the exposition, we introduce the vulnerability of player i, $v_i = 1 - s_i$ and the network *vulnerability* level $\bar{v} = 1 - \bar{s}$. Then, the expected utility of user i is:

$$U_i = (1 - v_i\bar{v}) \cdot f(W) + v_i\bar{v} \cdot f(W - D) - g(v_i), \tag{4.2}$$

where $g(v_i) = h(1 - v_i)$. This gives us: $g' \leq 0$, $g'' \geq 0$, and $g(1) = 0$, $g'(1) = 0$ and $g(0) = \infty$.

4.2.1 Social Optimum

We define the social optimum as a level at which aggregate user utility is maximized. In Appendix, we show that the socially optimal security level is identical for all users: $v_i = v$. Then, $\bar{v} = v$, and from (4.2), we have:

$$U^{soc} = (1 - v^2) \cdot f(W) + v^2 \cdot f(W - D) - g(v). \tag{4.3}$$

In any social optimum, $\frac{\partial U^{soc}}{\partial v} = 0$, from which we have:

$$2v^{soc} \left[f(W) - f(W - D) \right] = -g'(v^{soc}), \tag{4.4}$$

and since $\frac{\partial^2 U^{soc}}{\partial v^2} < 0$, the socially optimal vulnerability level is unique.

4.2.2 Nash Equilibrium

We assume that all parameters are known to users. As discussed above, a user takes the network vulnerability $\bar{v}$ as given and chooses his vulnerability v_i to maximize his utility given by (4.2). Taking the partial derivative of (4.2) with respect to v_i, and equating to zero, $\frac{\partial U_i}{\partial v_i} = 0$, we obtain:

$$\bar{v} \left[f(W) - f(W - D) \right] = -g'(v_i). \tag{4.5}$$

From the properties of the function g, the solution of equation (4.5) is unique, from which vulnerability choice is identical for all users. Hence, in equilibrium, all users have identical security (vulnerability) level, which we denote by v^*. Then, $\bar{v} = v^*$ and the following holds:

$$v^* \left[f(W) - f(W - D) \right] = -g'(v^*). \tag{4.6}$$

As in the case of social optimum, equilibrium vulnerability level is unique. Thus, in equilibrium, user security investments are identical and positive. Optimal investment increases when the damage D increases relative to wealth W. From (4.2) and (4.6), the equilibrium expected utility is:

$$U^* = f(W) + v^* g'(v^*) - g(v^*). \tag{4.7}$$

Comparing (4.4) and (4.6), we observe that since the LHS in (4.4) grows twice as fast as in (4.6) (see Fig. 4.1), we must have the following proposition:

Proposition 4.2.1 *Individually optimal user security is strictly positive, and it is strictly lower than the socially optimal one ($v^{soc} < v^*$ or $s^{soc} > s^*$).*

The expected per user loss due to network insecurity is:

$$(v^*)^2 D,$$

which is higher than the expected per user loss in the social optimum: $(v^{soc})^2 D$. Thus, in our model, users under-invest in security relative to a socially optimal level and this negative externality results in higher losses to society. In the next section, we will add competitive cyber-insurers to our base model and study how the presence of cyber-insurer affects network security.

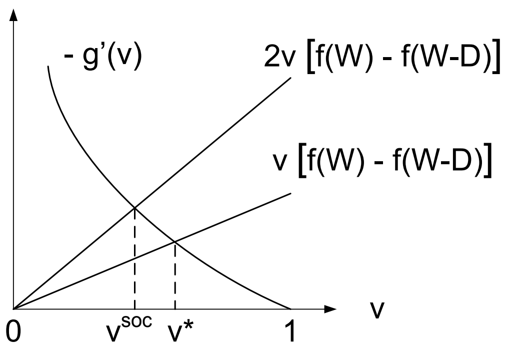


Figure 4.1. Nash Equilibrium Vs Social Optimum

4.3 Insurance

Equilibrium is defined in a way similar to [76], where insurance equilibrium is examined in the markets with adverse selection. Each insurer offers a single insurance contract in *a class of admissible contracts*, or does nothing. A Nash equilibrium is defined as a set of admissible contracts such that: i) all contracts offered at least break even; ii) taking as given the contracts offered by incumbent insurers (those offering contracts) there is no additional contract which an entrant-insurer (one not offering a contract) can offer and make a strictly positive profit; and iii) taking as given the set of contracts offered by other incumbent insurers, no incumbent can increase its profits by altering his offered contract.

The literature referred to such contracts as "competitive," because entry and exit are free, and because no barrier to entry or scale economies are present. Thus, we will consider insurance firms (insurers), who are risk neutral and compete with each other. In addition to these equilibrium conditions, we assume that individual insurers cannot affect the network vulnerability, and thus, take it as given.

Let ρ be the premium charged to a user and L be the amount of loss covered by the insurer. Let user vulnerability be v and network vulnerability be $\bar{v}$. The user pays the premium both when he is attacked and when he is not, but is covered a loss L when the attack occurs successfully. Thus, with probability $v\bar{v}$, the user is successfully attacked and receives utility $\big[f(W-D+L-\rho)-g(v)\big]$ and with probability $(1-v\bar{v})$, he obtains utility $\big[f(W-\rho)-g(v)\big]$. Denoting $U(v,\bar{v},\rho,L)$ as the corresponding expected user utility,

$$U(v,\bar{v},\rho,L) = (1-v\bar{v})\cdot f(\tilde{W}) + v\bar{v}\cdot f(\tilde{W}-\tilde{D}) - g(v), \tag{4.8}$$

where

$$\tilde{W} = W-\rho \text{ and } \tilde{D} = D-L.$$

The utility in (4.8) coincides with the no-insurance case if $\rho = 0, L = 0$. When v is identical for all users, we have $\bar{v} = v$, and:

$$U(v,v,\rho,L) = (1-v^2)\cdot f(\tilde{W}) + v^2\cdot f(\tilde{W}-\tilde{D}) - g(v). \tag{4.9}$$

4.3.1 Non-contractible User Security

In this subsection, we assume that insurers do not know, and have no control over user security level. This occurs when it is impossible (too costly) for the insurers to monitor the users' vulnerability v. Hence, the contract offered by an insurer will be of the form (ρ, L), i.e., the insurer sets the premium and the amount of coverage, and stipulates that no additional coverage can be purchased.

Note that the user is free to choose his required vulnerability here. Hence, users will choose the vulnerability level to maximize their utility, given the network security level. Thus, in the presence of competitive insurers, users choose which contract to buy, if any, and the corresponding vulnerability that maximizes their utility. In equilibrium, no user wishes to deviate from his equilibrium contract to any other contract or to not buying any insurance. We denote the equilibrium values in this non-contractible security case by the superscript †.

Social Planner

We assume that the social planner's objective is to maximize aggregate user utility with the constraint that the equilibrium contracts must not be loss-making. When social planner offers some contract(s), the users optimal choices could be described as if they play a game as in Section 4.2.2, but with wealth $\tilde{W} = W - \rho$ and damage $\tilde{D} = D - L$. In Appendix, we show that a social planner will offer a single contract (ρ, L) only. Then, user optimal choice is given by (4.6):

$$v^{\dagger soc}\left[f(\tilde{W}) - f(\tilde{W} - \tilde{D})\right] = -g'(v^{\dagger soc}). \qquad (4.10)$$

To maximize aggregate user utility, the social planner's contract must solve the following optimization problem:

$$\max_{\rho, L} U(v^{\dagger soc}, v^{\dagger soc}, \rho, L),$$

subject to (4.10) and budget constraint $\rho - (v^{\dagger soc})^2 L \geq 0$. In Appendix, we show that all users buy this insurance. No user deviates to not buying, when other users have bought the insurance.

With the insurance provided by a social planner, user utility is higher, but, the vulnerability $v^{\dagger soc}$ is also higher than in the no-insurance Nash equilibrium:

$$v^{\dagger soc} > v^*. \qquad (4.11)$$

Competitive Insurance

Insurers offer contracts (ρ, L), and users maximize their utility by choosing the contracts and a corresponding preferred security levels, given the network security. In Appendix, we show that in any equilibrium, the following proposition holds:

Proposition 4.3.1 *With competitive insurers present, and non-contractible user security, in equilibrium, the security is always worse than the security in the no-insurance Nash equilibrium ($\bar{v}^\dagger > v^*$ or $\bar{s}^\dagger < s^*$).*

Thus, we demonstrated that although insurers may allow users to reach a higher utility, the network security is strictly lower with insurers. The favorable effect of insurers on user utility is not free of cost for the society. The presence of insurers negatively impacts network security level, which increase the losses from network insecurity. Expected per user loss due to the insurers' presence increases relative to the no-insurance Nash equilibrium by $\Delta^\dagger$ given by:

$$\Delta^\dagger = \left[(\bar{v}^\dagger)^2 - (v^*)^2\right] D.$$

This is what one expects when insurers cannot monitor user security level. Since user risk is covered, users tend to under-invest in security. Next, we study the case where the insurer has perfect information about, and can perfectly enforce the security of his insured users.

4.3.2 Contractible User Security

Here, we assume that the insurers can monitor their insured users' vulnerability v at zero cost. Thus, we permit the contracts that specify user's required v. Let (v, ρ, L) be a contract that sets the premium ρ, the coverage L, and requires user vulnerability to be at most v. We denote the equilibrium values in this contractible security case by the superscript ‡.

Social Planner

We assume that the social planner's objective is to maximize aggregate user utility with the constraint that the equilibrium contracts must not be loss-making. In Appendix, we demonstrate that the social planner offers a single contract (v, ρ, L) only. Thus, $\bar{v} = v$, and to maximize total utility, the contract offered by the social planner must be a solution to the following optimization problem:

$$\max_{v,\rho,L} U(v, v, \rho, L), \text{ s.t. } v^2 L \leq \rho.$$

In Appendix, we show that the solution $(v^{\ddagger soc}, \rho^{\ddagger soc}, L^{\ddagger soc})$ is unique and satisfies:

$$\begin{aligned} (v^{\ddagger soc})^2 L^{\ddagger soc} &= \rho^{\ddagger soc} \\ L^{\ddagger soc} &= D \\ 2v^{\ddagger soc} D f'(W - (v^{\ddagger soc})^2 D) &= -g'(v^{\ddagger soc}). \end{aligned} \tag{4.12}$$

Thus, the optimal contract makes no profit and offers full coverage. The social planner choice $v^{\ddagger soc}$ is given by (4.12).

Competitive Insurance

In this case, insurers offer contracts (v, ρ, L). In equilibrium, if the network vulnerability is $\bar{v}$, then all equilibrium contracts must yield equal utility $U(v, \bar{v}, \rho, L)$ for the user. If there exists a contract (v, ρ, L) that an insurer can offer and improve this user utility, it is preferred by the users and users will deviate and buy that contract. Hence, in equilibrium, the contracts chosen by the insurers must maximize $U(v, \bar{v}, \rho, L)$. In Appendix, we prove the following proposition:

Proposition 4.3.2 *With competitive insurers present, and security level contractible, in equilibrium, profits are zero* ($v^{\ddagger 2} L^{\ddagger} = \rho^{\ddagger}$)*, and full coverage is offered* ($L^{\ddagger} = D$)*. The equilibrium contract is unique and in this equilibrium, the security is always worse than what will be chosen with a socially optimum insurance* ($v^{\ddagger} > v^{\ddagger soc}$)*. Also, compared to the no-insurance equilibrium, security is worse* ($v^{\ddagger} > v^{*}$) *except when the damage* D *is a small fraction of the wealth* W*. Users are strictly better off with insurers than when no insurers are present* ($U^{\ddagger} > U^{*}$)*.*

When security level is observable by the insurers, insurer presence allows to improve user welfare, but not necessarily the network security. Unless the damage is a small fraction of the wealth, with cyber-insurance, expected per user loss from network insecurity increases compared to the no-insurance Nash equilibrium by $\Delta^{\ddagger}$, where:

$$\Delta^{\ddagger} = \left[(\bar{v}^{\ddagger})^2 - (v^{*})^2\right] D.$$

Thus, for a significant range of parameters, the losses to society may increase when insurance is available.

4.4 Discussion

Our basic model captures the effects of interdependent security on network security. It highlights the incentive misalignment for end-user security investments, and captures a well known gap between individually and socially optimal user incentives for security (free-riding effect). We use this model to investigate the effects of the presence of competitive cyber-insurance on user utility and network security level.

We consider competitive cyber-insurance under two cases - one in which information asymmetry between the insurer and the user is present and the other in which it is absent. When this information asymmetry is present, the insurer is unable to observe and hence enforce any contract based on the user security level. Thus, we consider the two extreme cases (worst and best respectively) for cyber-insurance under the current network structure. We expect a real-life cyber-insurance to be intermediate between the two, reflecting the fact that cyber-insurers may have some information about user vulnerability but may not know/monitor user security level perfectly.

When information asymmetry is present, we show that the moral hazard problem becomes relevant. Here, users are insured and their risks are covered. In the contract, no conditions are imposed on user security, because even if they were, they cannot be enforced. In this case, users tend to further under-invest in security and the network security worsens in the presence of cyber-insurers, as is expected in a standard model [76]. Though the total user utility of the network has increased due to user risk being covered, the total damages (costs) to society have increased due to the lower security levels.

When no information asymmetry is present between users and insurers, we show that competitive cyber-insurers cover the entire user damage. However, even in this case, due to competitive pressures, in equilibrium, these insurers choose a security level that is lower than the socially optimum level. In this case, cyber-insurers free-ride on the security levels of their competitors which brings down the network security level. In fact, for a significant range of parameters, the network security level is lower than the security level without insurance. Thus, again, though user utility goes up, the total damages to society may increase.

Note that in the second case, there is no information asymmetry between the insurers and the users. However, there is still inefficiency of insurer action. Indeed, in our model, insurers cannot engage in an enforceable contract which would state the required security level for their offered contracts. With such a contract, insurers could have reached a better equilibrium, but this is unlikely because without an enforceable agreement, it is individually rational for each user to deviate and attract more users by offering a lax user contract.

Our assumption of identical users is simplistic, and does not hold in the actual Internet. But, adding user and insurer heterogeneity to our setting only increases informational asymmetries. When insurers could separate users of different types, our results should continue to hold for every user sub-type present in a heterogeneous environment. When insurers are unable to distinguish between users with different types, missing markets are likely, as [76] demonstrate. And, the presence of different insurer types brings adverse selection problems, such as the "lemon problem" [5], also leading to missing markets.

To conclude, we suggest that, for cyber-insurance to be an effective tool for improvement of network security, we need to solve two problems, not one. The traditional information asymmetry between insurers and users may be tackled by better monitoring and enforcing security best practices (via software certification techniques, for example). However, the free-riding between insurers also needs to be resolved. One way this may be achieved is via a mandate on the required user security in insurance contracts.

4.5 Conclusion

In this chapter, we investigate the effects of competitive cyber-insurers on network security and welfare. We highlight the impact of asymmetric information in the presence of network externalities and address the effects of interdependent security on the market for cyber-risks. The existing literature attributes cyber-insurance a significant role in cyber-risk management; it especially emphasizes positive effects of cyber-insurance market on security incentives. We find that, on reverse, the presence of competitive cyber-insurers, in general, weakens user incentives to improve security.

Though insurance improves the utility for risk-averse users, it does not serve as an incentive device for improving security practices. Indeed, insurance is a tool for risk management and redistribution, not necessarily a tool for risk reduction. To sum up, we argue that a combination of interdependent security and information asymmetries hinder cyber-insurance from performing the function of a catalyst for improvement of network security.

Appendix

Social Optimum with No Insurance

In the social optimum, the goal is to maximize aggregate user utility given by

$$U^{agg} = \sum_{i=1\dots N} [(1 - v_i\bar{v})f(W) + v_i\bar{v}f(W - D) - g(v_i)],$$

where $\bar{v} = \frac{\sum_{i=1\dots N} v_i}{N}$. To optimize this expression, we take the partial derivative w.r.t. v_j for some $j \in 1, \dots, N$ and equate to zero:

$$\begin{aligned} \frac{\partial U^{agg}}{\partial v_j} &= 0 \\ \sum_{i=1\dots N} \left[\frac{v_i}{N}\{f(W-D) - f(W)\}\right] + \bar{v}\{f(W-D) - f(W)\} - g'(v_j) &= 0 \\ 2\bar{v}\{f(W) - f(W-D)\} &= -g'(v_j). \quad (4.13) \end{aligned}$$

Since (4.13) is identical for all j, all users must be assigned an identical vulnerability to maximize the aggregate utility.

Proposition 4.3.1

Social Planner

First, we show that the social planner will offer a single contract in equilibrium only. Assume the reverse, and let there exist an equilibrium with network security $\bar{v}$, and at least two equilibrium contracts (ρ_1, L_1) and (ρ_2, L_2). Without loss of generality, let $v_1 > \bar{v} > v_2$.

Then, for any user with contract (ρ_1, L_1) optimal v_1 is the same as in the base model with $\tilde{W}_1 = W - \rho_1$ and $\tilde{D}_1 = D - L_1$, and thus v_1 is identical for all users with contract (ρ_1, L_1) and is given from (4.5):

$$\bar{v}\left[f(\tilde{W}_1) - f(\tilde{W}_1 - \tilde{D}_1)\right] = -g'(v_1). \tag{4.14}$$

Using (4.5), all these users' utility U_1 can be written as

$$U_1 = f(\tilde{W}_1) + v_1 g'(v_1) - g(v_1), \tag{4.15}$$

Similarly, for all users with contract (ρ_2, L_2) we have:

$$U_2 = f(\tilde{W}_2) + v_2 g'(v_2) - g(v_2).$$

Taking the derivative of $vg'(v) - g(v)$ w.r.t. v, we get

$$g'(v) - g'(v) + vg'' = vg'' \geq 0, \tag{4.16}$$

which implies that $vg'(v) - g(v)$ increases with v.

Now, consider instead a single contract (ρ, L_1) such that optimal user vulnerability in the base model with $\tilde{W} = W - \rho$ and $\tilde{D}_1 = D - L_1$ is v_1, i.e., from (4.6),

$$v_1\left[f(\tilde{W}) - f(\tilde{W} - \tilde{D}_1)\right] = -g'(v_1). \tag{4.17}$$

Comparing the LHS of (4.14) and (4.17), we infer that, since $v_1 > \bar{v}$, $\tilde{W} > \tilde{W}_1$ and hence $\rho < \rho_1$. Comparing the user utility with this single contract (ρ, L_1) with (4.15), we have

$$U = f(W - \rho) + v_1 g'(v_1) - g(v_1) > U_1 = U_2,$$

since $\rho < \rho_1$. Thus, this single contract (ρ, L_1) permits the social planner to achieve higher user utility and will be preferred to the two contracts (ρ_1, L_1) and (ρ_2, L_2). Hence, we have proven that only a single contract will be offered in the social planner optimum.

Second, we demonstrate that the network vulnerability with the optimal contract $(\rho^{\dagger soc}, L^{\dagger soc})$ is higher than in the no-insurance Nash equilibrium, i.e., $v^{\dagger soc} \geq v^*$. We know that $U^{\dagger soc}$ must be higher than U^* since U^* can always be reached by the planner offering the contract $(\rho, L) = (0, 0)$:

$$U^{\dagger soc} \geq U^*.$$

Next, for any contract (ρ, L) with optimal vulnerability v, similar to (4.15), the user's utility can be written as

$$U = f(W - \rho) + vg'(v) - g(v). \tag{4.18}$$

From $\rho > 0$, the monotonicity of $vg' - g$ from (4.16), (4.18) and (4.7), we infer that $U^{\dagger soc} \geq U^*$ holds only if

$$v^{\dagger soc} > v^*.$$

Last, we show that all users purchase this insurance. If a user i deviates to no-insurance, she obtains $U(v_i, v^{\dagger soc}, 0, 0)$, which is highest for $v_i = \tilde{v}$ determined from $\frac{\partial U(v_i, v^{\dagger soc}, 0, 0)}{\partial v_i} = 0$, which gives:

$$v^{\dagger soc}\left[f(W) - f(W - D)\right] = -g'(\tilde{v}). \tag{4.19}$$

Since $v^{\dagger soc} \geq v^*$, comparing the LHS of (4.19) and (4.6), we have

$$\tilde{v} \leq v^* \leq v^{\dagger soc}.$$

Next, we write $U(\tilde{v}, v^{\dagger soc}, 0, 0)$ using (4.19):

$$U(\tilde{v}, v^{\dagger soc}, 0, 0) = f(W) + \tilde{v}g'(\tilde{v}) - g(\tilde{v}). \tag{4.20}$$

Comparing with (4.7) using the monotonicity of $vg' - g$ derived in (4.16), we conclude that, since $\tilde{v} \leq v^*$, the user utility from deviation $U(\tilde{v}, v^{\dagger soc}, 0, 0)$ must be lower than U^*. Therefore,

$$U(\tilde{v}, v^{\dagger soc}, 0, 0) \leq U^* \leq U^{\dagger soc},$$

giving us the required result that no user will deviate and not buy insurance.

Competitive Insurers

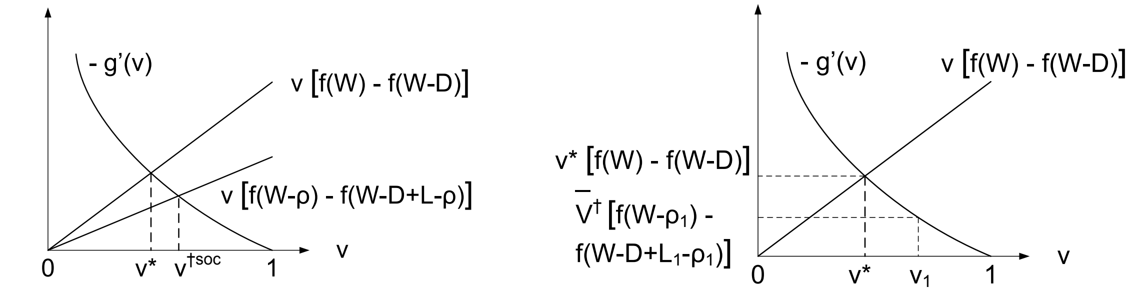
In the case of competing insurers, there may exist multiple contracts in equilibrium. However, the resulting network vulnerability $\bar{v}^\dagger$ will not be lower than the vulnerability in the Nash equilibrium v^*. Indeed, assume the reverse: $\bar{v}^\dagger < v^*$. Let (ρ_1, L_1) be some contract adopted by a non-zero fraction of users in this equilibrium.

From (4.5), replacing W by $\tilde{W}_1 = W - \rho_1$ and D by $\tilde{D}_1 = D - L_1$, we get an expression for the vulnerability v_1 chosen by users who adopt the contract (ρ_1, L_1):

$$\bar{v}^\dagger\left[f(\tilde{W}_1) - f(\tilde{W}_1 - \tilde{D}_1)\right] = -g'(v_1). \tag{4.21}$$

Note that $\rho_1 \leq L_1$, i.e., the premium must be lower than the coverage, else deviating from this contract to no-insurance gives higher utility to the users. Hence, $\left[f(\tilde{W}_1) - f(\tilde{W}_1 - \tilde{D}_1)\right] \leq [f(W) - f(W - D)]$.

From our assumption that $\bar{v}^\dagger \leq v^*$, we observe that the LHS of (4.21) is lesser than the LHS of (4.6) which implies that $v_1 \geq v^*$. (See Fig. 4.2(b).) However, the choice of (ρ_1, L_1) was arbitrary among all the contracts in equilibrium. Hence, the user adopting any contract in equilibrium will choose vulnerability not lower than v^*. This gives us $\bar{v}^\dagger \geq v^*$, which contradicts our assumption. Hence, $\bar{v}^\dagger \geq v^*$ in the competitive equilibrium as well.



(a) Nash Equilibrium Vs Social Optimum

(b) Deviation v_1

Figure 4.2. Competitive Non-contractible Insurance

Proposition 4.3.2

Social Planner

First, we show that a social planner will offer a single contract in equilibrium only. Indeed, assume the reverse. Let the social planner offer two contracts (The proof for more contracts is similar.) (v_1, ρ_1, L_1) and (v_2, ρ_2, L_2) to a fraction α and $(1-\alpha)$ of the population respectively. Thus,

$$\bar{v} = \alpha v_1 + (1-\alpha) v_2.$$

From the budget constraint, we have

$$\alpha\rho_1 + (1-\alpha)\rho_2 \geq \alpha v_1 \bar{v} L_1 + (1-\alpha) v_2 \bar{v} L_2.$$

From (4.8), we observe that the contracts with $L_1 = L_2 = D$ offer users a higher utility. Hence, we will only focus on the contracts (v_1, ρ_1, D) and (v_2, ρ_2, D) for the rest of the proof. In this case, the budget constraint becomes

$$\alpha\rho_1 + (1-\alpha)\rho_2 \geq \alpha v_1 \bar{v} D + (1-\alpha) v_2 \bar{v} D = \bar{v}^2 D. \tag{4.22}$$

From (4.8), the aggregate utility with the contracts (v_1, ρ_1, D) and (v_2, ρ_2, D) will be:

$$U^{agg} = \alpha[f(W-\rho_1) - g(v_1)] + (1-\alpha)[f(W-\rho_2) - g(v_2)].$$

Since both f and $(-g)$ are concave functions (and $f' > 0$),

$$\begin{aligned} U^{agg} &< f(W - \{\alpha\rho_1 + (1-\alpha)\rho_2\}) - g(\bar{v}) \\ &\leq f(W - \bar{v}^2 D) - g(\bar{v}), \end{aligned}$$

where the second inequality comes from (4.22).
$f(W - \bar{v}^2 D) - g(\bar{v})$ is the utility obtained from the contract $(\bar{v}, \bar{v}^2 D, D)$. Hence, there exists

the single contract $(\bar{v}, \bar{v}^2 D, D)$ which always provides a higher aggregate user utility than all contracts (v_1, ρ_1, L_1) and (v_2, ρ_2, L_2).

Thus, the social planner offers only a single contract. This contract must be a solution to the following optimization problem:

$$\max_{v,\rho,L} U(v, v, \rho, L)$$
$$\text{s.t. } v^2 L \leq \rho \text{ and } v \leq 1.$$

Next, we write the Lagrangian:

$$LAN = U(v, v, \rho, L) - \lambda_1(v^2 L - \rho) - \lambda_2(v - 1).$$

Let $\tilde{W} = W - \rho$ and $\tilde{D} = D - L$. Taking the derivatives of LAN w.r.t. v, L and ρ and equating to 0 gives us the following equations.

$$\frac{\partial LAN}{\partial v} = \frac{\partial U(v, v, \rho, L)}{\partial v} - 2\lambda_1 v L - \lambda_2 = 0,$$
$$\left[-2v(f(\tilde{W}) - f(\tilde{W} - \tilde{D}) - g'(v)\right] - 2\lambda_1 v L - \lambda_2 = 0. \tag{4.23}$$

$$\frac{\partial LAN}{\partial L} = \frac{\partial U(v, v, \rho, L)}{\partial L} - \lambda_1 v^2 = 0,$$
$$v^2 f'(\tilde{W} - \tilde{D}) - \lambda_1 v^2 = 0. \tag{4.24}$$

$$\frac{\partial LAN}{\partial \rho} = \frac{\partial U(v, v, \rho, L)}{\partial \rho} + \lambda_1 = 0,$$
$$-v^2 f'(\tilde{W} - \tilde{D}) - (1 - v^2) f'(\tilde{W}) + \lambda_1 = 0. \tag{4.25}$$

Further, from complementary slackness, we have

$$\lambda_1(v^2 L - \rho) = 0, \tag{4.26}$$
$$\text{and } \lambda_2(v - 1) = 0. \tag{4.27}$$

Note that $v \neq 0$, since that would require infinite security costs for the users. Hence, from (4.24), we conclude that $\lambda_1 = f'(\tilde{W} - \tilde{D}) > 0$ and thus, constraint (4.26) binds:

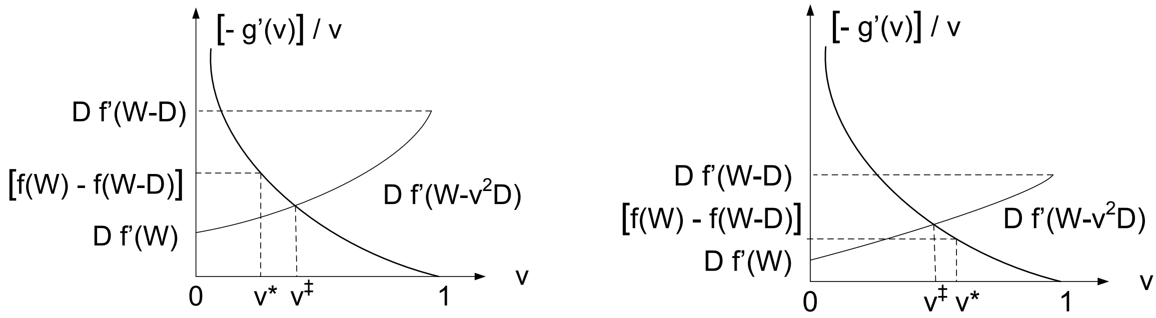
$$v^2 L = \rho. \tag{4.28}$$

Equating λ_1 from (4.24) and (4.25), we obtain:

$$f'(\tilde{W} - \tilde{D}) = v^2 f'(\tilde{W} - \tilde{D}) + (1 - v^2) f'(\tilde{W}). \tag{4.29}$$

Note that $v \neq 1$ because $g'(1) = 0$ and even a tiny decrease in v increases the expected benefit without increasing costs. Thus, from (4.29),

$$L = D. \tag{4.30}$$



(a) Large D

(b) Small D

Figure 4.3. Competitive Equilibrium with Contractible Security

Next, we can substitute (4.28) and (4.30) into (4.24) to get $\lambda_1 = f'(W - v^2 D)$. Also, $\lambda_2 = 0$, since $v < 1$. Substituting λ_1, λ_2, (4.28) and (4.30) into (4.23), we get

$$-g'(v) = 2vDf'(W - v^2 D). \tag{4.31}$$

Since the LHS is monotone decreasing while the RHS is monotone increasing, $v^{\ddagger soc}(< 1)$ is the unique solution to (4.31).

With $L = D$, and $\rho = v^2 D$ user utility at any v can be expressed as

$$U(v, v, v^2 D, D) = f(W - v^2 D) - g(v).$$

Maximum utility is reached at $\frac{dU(v)}{dv} = 0$:

$$-2vDf'(W - v^2 D) - g'(v) = 0, \tag{4.32}$$

and maximum is unique, because $\frac{d^2 U(v)}{dv^2} < 0$:

$$4v^2 D^2 f''(W - v^2 D) - 2Df'(W - v^2 D) - g''(v) < 0,$$

due to the properties of the functions f and g. Thus, there exists a unique social planner's optimum contract.

Competitive Insurers

First, we notice that in any equilibrium, due to competition, for any insurer, profit is zero, i.e. $\rho = v\bar{v}L$ for any equilibrium contract (v, ρ, L). If $\rho > v\bar{v}L$, some entrant insurer could offer a contract $(v, \tilde{\rho}, L)$ s.t. $\rho > \tilde{\rho} > v\bar{v}L$. From (4.8), the contract with a lower premium and same L, v and $\bar{v}$ improves user utility.

Second, full coverage, i.e. $L = D$ will be offered due to competition. Indeed, the contract $(v, v\bar{v}D, D)$ offers users the highest utility. To see this, consider the family of contracts

$(v, v\bar{v}L, L)$ for $L \leq D$. From (4.9), the utility $U(v, \bar{v}, v\bar{v}L, L)$ will be $(1-p)f(W-pL) + pf(W-pL-D+L) - g(v)$, where $p = v\bar{v}$. Differentiating w.r.t. L, and equating to 0, we get

$$p(1-p)f'(W-pL) = p(1-p)f(W-pL-D+L).$$

If $p \neq 0$ or 1, then $L = D$, which gives the required result. Henceforth, we restrict our analysis to contracts $(v, v\bar{v}D, D)$ only.

Third, in any equilibrium, user utility from deviation to no-insurance gives user a strictly lower utility. Indeed, assume the reverse. Consider a contract $(v_1, v_1\bar{v}D, D)$ that has a non-zero number of users adopting it in equilibrium. If a customer of this contract prefers to deviate to v_i with no insurance, then his utility without insurance must be greater than the utility with insurance contract, i.e., $U(v_i, \bar{v}, 0, 0) \geq U(v_1, \bar{v}, v_1\bar{v}D, D)$. Consider an entrant insurer who offers a contract $(v_i, v_i\bar{v}D, D)$ (full coverage at actuarially fair price). Adopting this contract improves user utility, which conflicts our equilibrium assumptions. Therefore, the utility from deviation to no-contract must be strictly lower than with a contract, and thus, all users strictly prefer to buy insurance.

Fourth, we prove that equilibrium contract is unique. Consider an equilibrium with network security $\bar{v}$. An entrant insurer could offer a contract $(\tilde{v}, \tilde{v}\bar{v}D, D)$ that maximizes $U(\tilde{v}, \bar{v}, \tilde{v}\bar{v}D, D) = f(W - \tilde{v}\bar{v}D) - g(\tilde{v})$. To determine $\tilde{v}$ at which user utility is the highest we differentiate

$$\frac{\partial}{\partial \tilde{v}} U(\tilde{v}, \bar{v}, \tilde{v}\bar{v}D, D) = \frac{\partial}{\partial \tilde{v}} \big(f(W - \tilde{v}\bar{v}D) - g(\tilde{v}) \big) = 0,$$
$$f'(W - \tilde{v}\bar{v}D)(-\bar{v}D) - g'(\tilde{v}) = 0,$$

and since the second derivative is always negative:

$$f''(W - \tilde{v}\bar{v}D)(\bar{v}D)^2 - g''(\tilde{v}) < 0,$$

user utility reaches its maximum at a single point $\tilde{v}$ only, which is exactly the contract offered in equilibrium. Thus, we have $\tilde{v} = \bar{v} = v^\dagger$, and in any equilibrium, all users buy an identical contract $(v^\ddagger, (v^\ddagger)^2 D, D)$, determined from

$$-g'(v^\ddagger) = v^\ddagger D f'(W - (v^\ddagger)^2 D). \tag{4.33}$$

The unique $v^\ddagger$ is strictly less than 1 since $f'(W-D) > 0$. (If $f'(W-D) = 0$, then $f(W) = f(W-D)$ and insurance does not improve user utility and is hence redundant.) Since the RHS of (4.31) is twice the RHS of (4.33), we conclude that $v^\ddagger \geq v^{\ddagger soc}$, i.e., the equilibrium security under competitive insurers is worse than under a social planner.

Next, we determine how $v^\ddagger$ compares to v^*. We rewrite (4.33) as

$$\frac{-g'(v^\ddagger)}{v^\ddagger} = Df'(W - (v^\ddagger)^2 D), \tag{4.34}$$

and compare with the Nash equilibrium by rewriting (4.6) as:

$$\frac{-g'(v^*)}{v^*} = f(W) - f(W-D). \tag{4.35}$$

Note that, from $f'' \leq 0$, we have $f'(W) \leq \frac{f(W)-f(W-D)}{D} \leq f'(W-D)$. Also, $f'(W-v^2D)$ is an increasing function of v, and $\frac{-g'(v)}{v}$ is decreasing. Hence, if $Df'(W-v^{*2}D) < f(W) - f(W-D)$ then $v^\ddagger > v^*$ else $v^\ddagger \leq v^*$. Thus, if the marginal benefit from full coverage offered at v^* is lower than the average loss of benefit per unit damage, insurance does not improve the security level.

Figure 4.3 depicts the solution of (4.34). From the figure, it is clear that only when D becomes small, the network security level in the equilibrium with insurers exceeds security level of no-insurance equilibrium. Note that when D is small, v^* is also large. Thus, competitive insurers improve network security only when equilibrium vulnerability in no-insurance equilibrium is high.

Bibliography

[1] Comcast Powerboost. [Online]. Available: http://www.comcast.net/powerboost/

[2] D. Acemoglu, K. Bimpikis, and A. Ozdaglar, "Price and capacity competition," *Games and Economic Behavior*, vol. In Press, Corrected Proof, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/B6WFW-4SWWT5G-2/2/f20546316e28b355815f539f0c1e9d3b

[3] D. Acemoglu and A. Ozdaglar, "Competition and efficiency in congested markets." *Math. Oper. Res.*, vol. 32, no. 1, pp. 1–31, 2007.

[4] D. Acemoglu, A. Ozdaglar, and R. Srikant, "The Marginal User Principle for Resource Allocation in Wireless Networks," in *Decision and Control, 2004. CDC. 43rd IEEE Conference on*, vol. 2, Dec 2004, pp. 1544–1549.

[5] G. A. Akerlof, "The market for 'lemons': Quality uncertainty and the market mechanism," *The Quarterly Journal of Economics*, vol. 84, no. 3, pp. 488–500, August 1970. [Online]. Available: http://ideas.repec.org/a/tpr/qjecon/v84y1970i3p488-500.html

[6] J. Altmann, B. Rupp, and P. Varaiya, "Effects of Pricing on Internet User Behavior," *Netnomics*, vol. 3, no. 1, pp. 67–84, 2001.

[7] R. Anderson, R. Böehme, R. Clayton, and T. Moore, "Security economics and european policy," in *Proceedings of WEIS'08*, Hanover, USA, Jun. 25-28 2008.

[8] T. Anderson, L. Peterson, S. Shenker, and J. Turner, "Overcoming the internet impasse through virtualization," *Computer*, vol. 38, no. 4, pp. 34–41, 2005.

[9] (1998) The economics of the Internet: Utility, utilization, pricing, and Quality of Service. AT&T Labs. [Online]. Available: http://www.dtc.umn.edu/~odlyzko/doc/internet.economics.pdf

[10] W. S. Baer and A. Parkinson, "Cyberinsurance in it security management," *IEEE Security and Privacy*, vol. 5, no. 3, pp. 50–56, 2007.

[11] J. P. Bailey, "The economics of internet interconnection agreements," pp. 155–168, 1997.

[12] R. Boehme and G. Kataria, "Models and measures for correlation in cyber-insurance," in *Fifth Workshop on the Economics of Information Security*, 2006.

[13] R. Böhme, "Cyber-insurance revisited," in *Proceedings of WEIS'05*, Cambridge, USA, 2005.

[14] R. Bohme and G. Schwartz, "Cyber Insurance: Hopes and Practices - Summary of the Literature," In preparation, Available March 2010, http://www1.inf.tu-dresden.de/~rb21/publications/BS2010_CyberInsuranceReview.pdf, 2010.

[15] J. Bolot and M. Lelarge, "A new perspective on internet security using insurance," *INFOCOM 2008. The 27th Conference on Computer Communications. IEEE*, pp. 1948–1956, April 2008.

[16] L. Brakmo and L. Peterson, "Tcp vegas: end to end congestion avoidance on a global internet," *Selected Areas in Communications, IEEE Journal on*, vol. 13, no. 8, pp. 1465 –1480, oct 1995.

[17] B. Briscoe, A. Jacquet, C. Di Cairano-Gilfedder, A. Salvatori, A. Soppera, and M. Koyabe, "Policing congestion response in an internetwork using re-feedback," *SIGCOMM Comput. Commun. Rev.*, vol. 35, no. 4, pp. 277–288, 2005.

[18] V. Chandrasekhar, J. Andrews, and A. Gatherer, "Femtocell networks: a survey," vol. 46, no. 9, pp. 59–67, Sept. 2008.

[19] V. Chandrasekhar and J. Andrews, "Uplink Capacity and Interference Avoidance for Two-Tier Cellular Networks," in *Global Telecommunications Conference, 2007. GLOBECOM '07. IEEE*, Nov. 2007, pp. 3322–3326.

[20] N. M. K. Chowdhury and R. Boutaba, "A survey of network virtualization," *Computer Networks*, vol. 54, no. 5, pp. 862 – 876, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/B6VRG-4XKBYY4-4/2/91ac30e6e1c4d24c53d95798a6148f03

[21] H. Claussen, "Performance of Macro- and Co-Channel Femtocells in a Hierarchical Cell Structure," in *Personal, Indoor and Mobile Radio Communications, 2007. PIMRC 2007. IEEE 18th International Symposium on*, Sept. 2007, pp. 1–5.

[22] H. Claussen, L. Ho, and L. Samuel, "Financial Analysis of a Pico-Cellular Home Network Deployment," in *Communications, 2007. ICC '07. IEEE International Conference on*, June 2007, pp. 5604–5609.

[23] R. H. Coase, "The problem of social cost," *Journal of Law and Economics*, vol. 3, pp. 1–44, 1960. [Online]. Available: http://www.jstor.org/stable/724810

[24] C. Courcoubetis and R. Weber, *Pricing Communication Networks: Economics, Technology and Modelling (Wiley Interscience Series in Systems and Optimization)*. John Wiley & Sons, 2003.

[25] J. Crowcroft, S. Hand, R. Mortier, T. Roscoe, and A. Warfield, "QoS's Downfall: At the Bottom, or Not at All!" in *RIPQoS '03: Proc. ACM SIGCOMM workshop on Revisiting IP QoS*. New York, NY, USA: ACM, 2003, pp. 109–114.

[26] B. Davie, "Deployment Experience with Differentiated Services," in *RIPQoS '03: Proc. ACM SIGCOMM workshop on Revisiting IP QoS*, 2003, pp. 131–136.

[27] M. D. M. de Montmarin, "A Result Similar to the Odlyzko's 'Paris Metro Pricing'," *Applied Economics*, vol. 38, pp. 1821–1824, 2006.

[28] J. Farrell and P. Klemperer, "Coordination and Lock-In: Competition with Switching Costs and Network Effects," *SSRN eLibrary*, 2006.

[29] J. Farrell and G. Saloner, "Installed base and compatibility: Innovation, product preannouncements, and predation," *The American Economic Review*, vol. 76, no. 5, pp. 940–955, 1986. [Online]. Available: http://www.jstor.org/stable/1816461

[30] E. Felten, "Nuts and Bolts of Network Neutrality," Working Paper, http://itpolicy.princeton.edu/pub/neutrality.pdf, Center for Intormation Technology Policy, 2006.

[31] M. Fisk, "Causes and remedies for social acceptance of network insecurity," in *Proceedings of WEIS'02*, Berkeley, USA, 2002.

[32] S. Floyd, M. Handley, J. Padhye, and J. Widmer, "Equation-based congestion control for unicast applications," in *SIGCOMM '00: Proceedings of the conference on Applications, Technologies, Architectures, and Protocols for Computer Communication*. New York, NY, USA: ACM, 2000, pp. 43–56.

[33] R. Gibbens, R. Mason, and R. Steinberg, "Internet Service Classes under Competition," vol. 18, no. 12, pp. 2490–2498, Dec. 2000.

[34] L. A. Gordon, M. Loeb, and T. Sohail, "A framework for using insurance for cyber-risk management," *Communications of the ACM*, vol. 46, no. 3, pp. 81–85, 2003.

[35] L. A. Gordon and M. P. Loeb, "The economics of information security investment," *ACM Trans. Inf. Syst. Secur.*, vol. 5, no. 4, pp. 438–457, 2002.

[36] J. Grossklags, N. Christin, and J. Chuang, "Secure or insure?: a game-theoretic analysis of information security games," in *WWW '08: Proceeding of the 17th international conference on World Wide Web*. New York, NY, USA: ACM, 2008, pp. 209–218.

[37] A. Gupta, B. Jukic, M. Li, D. O. Stahl, and A. B. Whinston, "Estimating Internet Users' Demand Characteristics," *Computational Economics*, vol. 17, no. 2-3, pp. 203–218, 2001.

[38] A. Gupta, D. O. Stahl, and A. B. Whinston, "Priority Pricing of Integrated Services Networks," *Internet economics*, pp. 323–352, 1997.

[39] ——, "The economics of network management," *Commun. ACM*, vol. 42, no. 9, pp. 57–63, 1999.

[40] K. Hausken, "Returns to information security investment: The effect of alternative information security breach functions on optimal investment and sensitivity to vulnerability," *Information Systems Frontiers*, vol. 8, no. 5, pp. 338–349, 2006.

[41] L. He and J. Walrand, "Pricing Differentiated Internet Services," *Proc. IEEE INFOCOM 2005*, vol. 1, pp. 195–204, Mar. 2005.

[42] B. E. Hermalin and M. L. Katz, "The Economics of Product-Line Restrictions With an Application to the Network Neutrality Debate," July 2006.

[43] L. Ho and H. Claussen, "Effects of User-Deployed, Co-Channel Femtocells on the Call Drop Probability in a Residential Scenario," in *Personal, Indoor and Mobile Radio Communications, 2007. PIMRC 2007. IEEE 18th International Symposium on*, Sept. 2007, pp. 1–5.

[44] J. D. Hobby and H. Claussen, "Deployment options for femtocells and their impact on existing macrocellular networks," *Bell Labs Technical Journal*, vol. 13, no. 4, pp. 145–160, 2009.

[45] A. Hofmann, "Internalizing externalities of loss prevention through insurance monopoly: an analysis of interdependent risks," *Geneva Risk and Insurance Review*, vol. 32, no. 1, pp. 91–111, 2007.

[46] L. G. S. Holger Claussen, Lester T. W. Ho, "An overview of the femtocell concept," *Bell Labs Technical Journal*, vol. 13, no. 1, pp. 221–245, 2008.

[47] P. Honeyman, G. Schwartz, and A. V. Assche, "Interdependence of reliability and security," in *Proceedings of WEIS'07*, Pittsburg, PA, 2007.

[48] J. W. J. Musacchio and S. Wu, "A game theoretic model for network upgrade decisions," in *44th Annual Allerton Conference on Communication and Control*, 2006, pp. 191–200.

[49] L. Jiang, V. Anantharam, and J. Walrand, "Efficiency of selfish investments in network security," in *NetEcon '08: Proceedings of the 3rd international workshop on Economics of networked systems.* New York, NY, USA: ACM, 2008, pp. 31–36.

[50] Y. Jin, S. Sen, R. Guérin, K. Hosanagar, and Z.-L. Zhang, "Dynamics of competition between incumbent and emerging network technologies," in *NetEcon '08: Proceedings of the 3rd international workshop on Economics of networked systems.* New York, NY, USA: ACM, 2008, pp. 49–54.

[51] R. Johari and J. N. Tsitsiklis, "Efficiency loss in a network resource allocation game," *Mathematics of Operations Research*, vol. 29, no. 3, pp. 407–435, 2004. [Online]. Available: http://www.jstor.org/stable/30035659

[52] R. Johari and G. Y. Weintraub, "Competition in Prices and Service Level Guarantees," *SSRN eLibrary*, 2007.

[53] D. Joseph, N. Shetty, J. Chuang, and I. Stoica, "Modeling the adoption of new network architectures," in *CoNEXT '07: Proceedings of the 2007 ACM CoNEXT conference*. New York, NY, USA: ACM, 2007, pp. 1–12.

[54] F. P. Kelly, A. K. Maulloo, and D. K. H. Tan, "Rate control for communication networks: Shadow prices, proportional fairness and stability," *The Journal of the Operational Research Society*, vol. 49, no. 3, pp. 237–252, 1998. [Online]. Available: http://www.jstor.org/stable/3010473

[55] E. Koutsoupias and C. Papadimitriou, "Worst-case equilibria," *Computer Science Review*, vol. 3, no. 2, pp. 65 – 69, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/B8JDG-4W9S30K-1/2/ebc5bf18b0ab9dda637a62d9b8a9c61f

[56] D. M. Kreps and J. A. Scheinkman, "Quantity precommitment and bertrand competition yield cournot outcomes," *Bell Journal of Economics*, vol. 14, no. 2, pp. 326–337, Autumn 1983. [Online]. Available: http://ideas.repec.org/a/rje/bellje/v14y1983iautumnp326-337.html

[57] H. Kunreuther and G. Heal, "Interdependent security," *Journal of Risk and Uncertainty*, vol. 26, no. 2-3, pp. 231–49, March-May 2003. [Online]. Available: http://ideas.repec.org/a/kap/jrisku/v26y2003i2-3p231-49.html

[58] H. C. Kunreuther and E. O. Michel-Kerjan, "Evaluating the effectiveness of terrorism risk financing solutions," National Bureau of Economic Research, Inc, NBER Working Papers 13359, Oct. 2007. [Online]. Available: http://ideas.repec.org/p/nbr/nberwo/13359.html

[59] C. Labovitz, D. McPherson, and S. Iekel-Johnson. 2009 Internet Observatory Report. Arbor Networks. [Online]. Available: http://www.nanog.org/meetings/nanog47/abstracts.php?pt=MTQ1MyZuYW5vZzQ3&nm=nanog47

[60] P. Laskowski and J. Chuang, "Network monitors and contracting systems: competition and innovation," *SIGCOMM Comput. Commun. Rev.*, vol. 36, no. 4, pp. 183–194, 2006.

[61] ——, "A Leap of Faith? From Large-Scale Testbed to the Global Internet," *37th Telecommunications Policy Research Conference*, 2009.

[62] M. Lelarge and J. Bolot, "Network externalities and the deployment of security features and protocols in the internet," *SIGMETRICS Perform. Eval. Rev.*, vol. 36, no. 1, pp. 37–48, 2008.

[63] ——, "Economic incentives to increase security in the internet: The case for insurance," in *INFOCOM 2009, IEEE*, April 2009, pp. 1494–1502.

[64] R. Levitan and M. Shubik, "Duopoly with price and quantity as strategic variables," *International Journal of Game Theory*, vol. 7, no. 1, pp. 1–11, Mar. 1978. [Online]. Available: http://www.springerlink.com/content/q14gp60741540455

[65] D. Lopez-Perez, A. Valcarce, G. De La Roche, E. Liu, and J. Zhang, "Access methods to WiMAX femtocells: A downlink system-level case study," in *Communication Systems, 2008. ICCS 2008. 11th IEEE Singapore International Conference on*, Nov. 2008, pp. 1657–1662.

[66] J. K. MacKie-Mason, S. Shenker, and H. R. Varian, "Network Architecture and Content Provision: An Economic Analysis," *THE INTERNET AND TELECOMMUNICATIONS POLICY, G. Brock and G. Rosston, eds., Lawrence Erlbaum: 1996.*

[67] R. P. Majuca, W. Yurcik, and J. P. Kesan, "The evolution of cyberinsurance," ACM Computing Research Repository, Tech. Rep. CR/0601020, 2006.

[68] N. McKeown, T. Anderson, H. Balakrishnan, G. Parulkar, L. Peterson, J. Rexford, S. Shenker, and J. Turner, "Openflow: enabling innovation in campus networks," *SIGCOMM Comput. Commun. Rev.*, vol. 38, no. 2, pp. 69–74, 2008.

[69] D. C. North, "Economic performance through time," *The American Economic Review*, vol. 84, no. 3, pp. 359–368, 1994. [Online]. Available: http://www.jstor.org/stable/2118057

[70] A. Odlyzko, "Paris Metro Pricing for the Internet," in *EC '99: Proc. of the 1st ACM conference on Electronic commerce.* ACM, 1999, pp. 140–147.

[71] ——, "The Economics of the Internet: Utility, Utilization, Pricing, and Quality of Service," Tech. Rep., 1999.

[72] H. Ogut, N. Menon, and S. Raghunathan, "Cyber insurance and it security investment: Impact of interdependent risk," in *Proceedings of WEIS'05*, Cambridge, USA, 2005.

[73] picoChip Designs Ltd, "The Case for Home Base Stations," White Paper, Apr. 2007. [Online]. Available: http://www.femtoforum.org/femto/Files/File/picoChipFemtocellWhitePaper1.1.pdf

[74] S. ping Yeh, S. Talwar, S. choon Lee, and H. Kim, "WiMAX femtocells: a perspective on network architecture, capacity, and coverage," vol. 46, no. 10, pp. 58–65, Oct. 2008.

[75] S. Radosavac, J. Kempf, and U. Kozat, "Using insurance to increase internet security," in *Proceedings of NetEcon'08*, Seattle, USA, Aug. 22 2008.

[76] M. Rothschild and J. E. Stiglitz, "Equilibrium in competitive insurance markets: An essay on the economics of imperfect information," *The Quarterly Journal of Economics*, vol. 90, no. 4, pp. 630–49, November 1976. [Online]. Available: http://ideas.repec.org/a/tpr/qjecon/v90y1976i4p630-49.html

[77] ——, "Competition and insurance twenty years later," *The Geneva Risk and Insurance Review*, vol. 22, no. 2, pp. 73–79, December 1997. [Online]. Available: http://ideas.repec.org/a/pal/genrir/v22y1997i2p73-79.html

[78] T. Roughgarden and E. Tardos, "How bad is selfish routing?" *J. ACM*, vol. 49, no. 2, pp. 236–259, 2002.

[79] ——, "How bad is selfish routing?" *J. ACM*, vol. 49, no. 2, pp. 236–259, 2002.

[80] C. Saraydar, N. Mandayam, and D. Goodman, "Pricing and power control in a multicell wireless data network," vol. 19, no. 10, pp. 1883–1892, Oct. 2001.

[81] S. E. Schechter, "Computer security strength and risk: a quantitative approach," Ph.D. dissertation, Cambridge, MA, USA, 2004, adviser-Smith,, Michael D.

[82] H. Schulzrinne and J. Rosenberg, "Internet telephony: architecture and protocols—an ietf perspective," *Comput. Netw.*, vol. 31, no. 3, pp. 237–255, 1999.

[83] G. Schwartz, N. Shetty, and J. Walrand, "Cyber Insurance with Interdepedent Security and Aysmmetric Information," In preparation, www.eecs.berkeley.edu/~nikhils/EconSec.pdf, 2010.

[84] ——, "Impact of QoS on Internet User Welfare," in *4th International Workshop On Internet And Network Economics (WINE)*, 2008. [Online]. Available: http://www.eecs.berkeley.edu/~nikhils/PublishedPapers/SSW-WINE08.pdf

[85] ——, "Network Neutrality: Avoiding the Extremes," in *Forty-sixth Annual Allerton Conference*, 2008. [Online]. Available: http://www.eecs.berkeley.edu/~nikhils/PublishedPapers/AllertonSSW-NNATE08.pdf

[86] ——, "Net Neutrality and ISPs Property Rights: Why Internet Lacks Differentiated Services?" Working Paper, www.eecs.berkeley.edu/~nikhils/SSW-Econ.pdf, 2010.

[87] M. Schwartz and P. J. Weiser, "Introduction to a special issue on network neutrality," *Review of Network Economics*, vol. 8, no. 1, pp. 1–12, Mar. 2009. [Online]. Available: http://ideas.repec.org/a/rne/rneart/v8y2009i1p1-12.html

[88] S. Shakkottai, R. Srikant, A. Ozdaglar, and D. Acemoglu, "The Price of Simplicity," *Selected Areas in Communications, IEEE Journal on*, vol. 26, no. 7, pp. 1269–1276, Sep 2008.

[89] S. Shenker, D. Clark, D. Estrin, and S. Herzog, "Pricing in computer networks: reshaping the research agenda," *SIGCOMM Comput. Commun. Rev.*, vol. 26, no. 2, pp. 19–43, 1996.

[90] J. G. Sidak, "A consumer-welfare approach to network neutrality regulation of the Internet," *Journal of Competition Law and Economics*, vol. 2, no. 3, pp. 349–474, 2006. [Online]. Available: http://jcle.oxfordjournals.org/cgi/content/abstract/2/3/349

[91] K. Soohoo, "How much is enough? a risk-management approach to computer security," Ph.D. dissertation, Stanford University.

[92] P. Srinagesh, "Internet cost structures and interconnection agreements," pp. 121–154, 1997.

[93] D. O. Stahl, R. Dai, and A. B. Whinston, "An Economic Analysis of Multiple Internet QoS Channels," 2003.

[94] J. E. Stiglitz, "Information and the change in the paradigm in economics," *American Economic Review*, vol. 92, no. 3, pp. 460–501, June 2002. [Online]. Available: http://ideas.repec.org/a/aea/aecrev/v92y2002i3p460-501.html

[95] L. Subramanian, S. Agarwal, J. Rexford, and R. Katz, "Characterizing the internet hierarchy from multiple vantage points," in *INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 2, 2002, pp. 618 – 627 vol.2.

[96] B. van Schewick and D. Farber, "Point/counterpoint network neutrality nuances," *Commun. ACM*, vol. 52, no. 2, pp. 31–37, 2009.

[97] H. Varian, "System reliability and free riding," in *Workshop on the Economics of Information Security, WEIS 2002*, Cambridge, USA, 2002.

[98] H. R. Varian, "The Demand for Bandwidth: Evidence from the INDEX Project," http://www.sims.berkeley.edu/~hal/Papers/brookings.pdf, 2001.

[99] J. Walrand, *Economic Models of Communication Networks.* Springer, 2008, ch. 3, pp. 57–87. [Online]. Available: http://www.springer.com/computer/hardware/book/978-0-387-79360-3

[100] Z. Wang, *Internet QoS: Architectures and Mechanisms for Quality of Service.* San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2001.

[101] D. X. Wei, C. Jin, S. H. Low, and S. Hegde, "Fast tcp: motivation, architecture, algorithms, performance," *IEEE/ACM Trans. Netw.*, vol. 14, no. 6, pp. 1246–1259, 2006.

[102] G. Y. Weintraub, R. Johari, and B. Van Roy, "Investment and Market Structure in Industries with Congestion," *SSRN eLibrary*, 2007.

[103] O. E. Williamson, "The economics of organization: The transaction cost approach," *The American Journal of Sociology*, vol. 87, no. 3, pp. 548–577, 1981. [Online]. Available: http://www.jstor.org/stable/2778934

[104] ——, "Transaction Cost Economics: How it Works; Where it is Headed," *SSRN eLibrary*, 1998.

[105] C. S. Yoo, "Network Neutrality and the Economics of Congestion," *Georgetown Law Journal*, vol. 94, June 2006. [Online]. Available: http://ssrn.com/paper=825669