Chapter 3
Economic Models of Communication Networks

Jean Walrand

Abstract
Standard performance evaluations of communication networks focus on the *technology layer* where protocols define precise rules of operations. Those studies assume a model of network utilization and of network characteristics and derive performance measures. However, performance affects how users utilize the network. Also, investments by network providers affect performance and consequently network utilization. We call the actions of users and network providers the "*economic layer*" of the network because their decisions depend largely on economic incentives. The economic and technology layers interact in a complex way and they should be studied together. This tutorial explores economic models of networks that combine the economic and technology layers.

3.1 Introduction

Why were QoS mechanisms for end users not implemented in the Internet? Should users have a choice of grade of service for different applications? Why is security of the Internet so poor? Should Internet service providers be allowed to charge content providers for transporting their traffic? Should cell phones be unlocked to work with multiple operators? Should municipalities deploy free Wi-Fi networks? How should different services of a WiMAX network be priced?

These questions that affect the future of communication networks go beyond technology. However, their answers certainly depend on technological features. Different protocols enable or prevent choices of users and influence the revenue of operators and, consequently, their investment incentives. The engineers who define network protocols typically focus on their performance characteristics but are largely unaware of the market consequences of their designs.

Jean Walrand
Department of EECS, University of California at Berkeley, e-mail: wlr@eecs.berkeley.edu

This tutorial explains some combined models of user and provider incentives and performance of the network. The objective is not to present a comprehensive survey of research in this area. Rather, it is to illustrate some key aspects of these studies.

3.1.1 General Issues

This tutorial is not concerned with the strategic behavior of users who manipulate protocols, such as cheating the backoff algorithm of a multiple access protocol or the window adjustment rules of TCP. For a review of the game-theoretic formulation of such behavior, see [3]. Instead, this tutorial explores the behavior of more typical users and providers that do not modify the basic hardware and software of the communication devices. The questions concern investments, pricing, and usage patterns. Users of a communication network decide to use services based on their utility for the services. The utility of a service for a given user depends on how much the user values the service and on its price. Services include access to content, to applications, and communication services. Users also add content through their web sites, possibly through a social network, or by making files available in a peer-to-peer network. The valuation of a service depends on the richness of the content, the quality of the transport service (bandwidth and delay), and the third parties with whom the user can communicate. Network operators choose their investments, prices, and the services they offer to maximize their profit or some other objective that depends on profits. Figure 3.1 illustrates the interactions among users and providers through the network. Thus, through the network, a user affects other users and providers and a provider affects other providers and users.

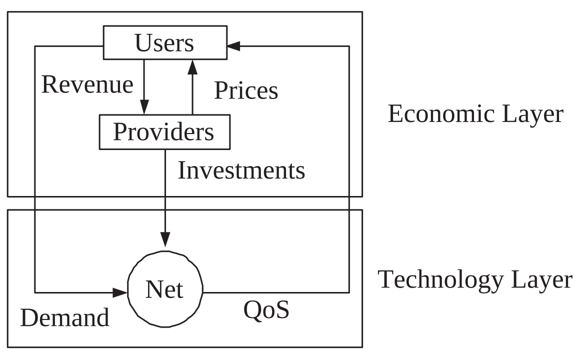

Fig. 3.1 Users and providers respond to economic incentives and affect the network.

When you drive on a congested highway, you increase the travel time of other vehicles and the air pollution. That is, your utilization of the highway has a negative impact on the welfare of other people. Similarly, in a network, the actions of users, content providers (such as Google and Yahoo!), and transport providers (such as AT&T and Comcast) affect the other users and providers. Economists call *externality* the impact on others of some action when prices do not reflect that impact. Externalities are present in most systems and they play a crucial role in the actual outcomes that economic agents face. We explain that networks exhibit positive and negative externalities, as summarized in Figure 3.2. The table also identifies the sections of the paper where we analyze the corresponding effects.

	Users	Content	Transport
Users	+ Contacts - Congestion 1.2, 2 + Security 3.3	+ Interest 3.2	+ Access 1, 2, 3
Content	+ Revenue 3.2	+ Appeal - Competition	+ Access 3.2
Transport	+ Revenue 1, 2, 3	+ Traffic 3.2	+ Access - Competition 2.4

Fig. 3.2 Externalities of an agent in a column on the agents in the rows.

The first column of the table shows the externalities of users on other users and on content and transport providers. The presence of users on the network has a positive externality on other users by adding one person they can contact through the network or by adding content in a peer-to-peer network, on a web site, or on a social network. However, the usage of the network by a user has a negative externality because it may increase congestion in the network and reduce the value of other services by slowing them down. The security investments of a user have a positive externality on other users because they reduce the chances of denial of service attacks, leak of confidential information, or the likelihood of virus infection of the computers of other users. The usage of the network by users increases the revenue of transport and content providers, a positive externality.

The second column shows that if a content provider invests more, it increases the interest that users have in the network and increases the utility they derive from the network, a positive externality. The column also shows that an increased investment in content by a provider may increase the revenue of other content providers indirectly by increasing the general appeal of the network and, consequently, the usage of the network. However, if a content provider is much more attractive, it may reduce the traffic of its competitors, a negative externality typical of competition. Improving content servers increases the traffic and the demand for transport provider services, a positive externality.

The third column shows that an increase in investment by a transport provider improves the access to content and other users, which improve the experience of users, a positive externality. Moreover, such an investment increase may generate more traffic on the content provider servers. Finally, if a transport provider improves its network, this may increase the traffic on the network of other transport providers, but it may decrease that of some competitors.

The standard view is that economic agents are selfish in the sense that they choose actions that maximize their own utility without regard for that of other agents. Because of externalities, it is generally the case that the selfish behavior of agents results in a less-than-maximum *social welfare*, defined as the total utility of all the agents minus the cost of providing the services. As an example, the negative externality of congestion results in a selfish demand by users that is larger than socially optimal. This effect is called the *tragedy of the commons* [16] as we discuss in Section 3.2.1. Also, the positive externality of investments typically results in selfish investments that are lower than socially optimal, an effect called *free-riding* [48] that we review in Section 3.3.1.

Another important effect in economic systems is the role of information in the efficiency of markets. If you suspect that a used car might be a lemon, you are not willing to pay much for it. For instance, if the dealer asks $10,000 for the car, you expect that he probably paid $5,000 for it, so that you might be willing to pay $6,000 for it, but not the asking price. However, the dealer may have a car that he paid $9,000 for that he would be willing to sell for $10,000 and that you might want to buy for that price. Consequently, the lack of information results in good used cars not being sold even though some buyers would be willing to purchase them at a price agreeable to the dealer, an effect called *missing market* by economists [2]. One could argue that the inability to select the quality of Internet connections results in missing markets. Users might be willing to pay more for dependable high-quality connections. See [43] for an interesting discussion of the importance of economic aspects of the Internet and [49] for a broader presentation of the economic aspects of the information technology.

Summing up, important issues in economic models of networks are the effects of externality of activity and of investments and the quality of information available to agents and when that information is available.

3.1.2 Paris Metro Pricing

This section illustrates the interdependence of the economic and technology layers on a simple example due to A. Odlyzko [39]. Imagine a metro whose otherwise identical cars are divided into expensive first class cars and inexpensive second class cars. The first class cars are less crowded because they are more expensive, which justifies their higher price. The general effect behind this example is that quality of service affects utilization (the number of users of the service) and utilization affects quality of service. Consequently, the closed-loop behavior of the system is more subtle than might be anticipated. We consider two models of this situation. The first model assumes that the utility of a network for each user depends on its utilization. The second model is more radical and considers applications that are incompatible.

Model 1: Utility Depends on Utilization

Imagine a network whose delays are acceptable for voice-over-IP as long as the utilization (number of users) is less than 200 and acceptable for web-browsing if the utilization is less than 800. Assume that the demand (potential utilization) for voice-over-IP is 100 as long as the price does not exceed 20 and that the demand for web-browsing is 400 as long as the price does not exceed 5 and that these demands vanish if the prices exceed those values, because users switch to a competitor's network. How much should the operator charge for the service?

If the operator charges 20, the web-browsing users do not connect. All the voice-over-IP users connect because their total demand (100) is small enough for the network delays to be acceptable for that application. The revenue of the operator is then the utilization (100) multiplied by the price (20), or 2,000. On the other hand, if the operator charges 5, then all the web-browsing users connect, the voice-over-IP users do not because the utilization is too large and the delays are not acceptable for them. The resulting revenue is now the utilization (400) multiplied by the price (5), or again 2,000.

Now consider the following strategy of the operator. He divides the network into two subnetworks, each with half of the capacity of the original network. The actual technology (e.g., time-division multiplexing or deficit-round-robin) used for this splitting of the network does not really matter. In each network, the delays are acceptable for voice-over-IP if the utilization is less than 100 (half of the previous acceptable utilization). Also, the delays are acceptable for web-browsing if the utilization is less than 400. The operator charges 20 for one network and 5 for the other. The voice-over-IP users connect to the first network and the web-browsing users connect to the second. The operator revenue is now 100×20 for the first network and 400×5 for the second, or a total of 4,000. The situation is illustrated in Figure 3.3 that shows that the quality of service (QoS) of the first network is automatically better than that of the second because it is more expensive.

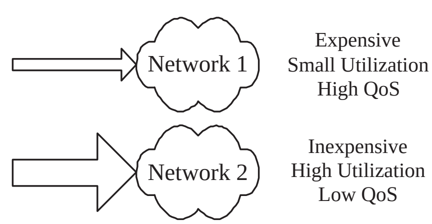

Fig. 3.3 Two identical networks with different prices have different QoS.

Obviously the example was designed specifically for the service differentiation to increase the revenue substantially. What happens in a more general situation? If the demand for voice-over-IP is of the form $A \times 1\{x \leq x_1, p \leq a\}$ and the demand for web browsing is $B \times 1\{x \leq x_2, p \leq b\}$ when the utilization is x and the price is p, then one can show that the revenue of the Paris Metro pricing cannot be more than twice the revenue of a single network.

As this example shows, even though the two networks are identical, the higher price of the first network results in a lower utilization and, consequently, a higher quality that justifies the higher price. Note that the ability to differentiate the service doubles the revenue of the operator by eliminating the negative congestion externality of the users of one class on the users of the other class. This differentiation requires the users to signal their service preference and the operator to charge differently for the two services.

Model 2: Incompatible Applications

The devices in a Wi-Fi network use a multiple access protocol to share one radio channel. The protocol specifies that a node must wait a random time after the channel is silent before it starts transmitting. If the transmission of one device happens to collide with another one, as the device notices because it does not get an acknowledgment of its transmission, the device increases the interval from which it picks a random waiting time before the next attempt. This protocol has the advantage of sharing the channel fairly among the different devices and also of stabilizing the network by avoiding many repeated collisions. However, because of this random waiting time, it may happen that the delay before transmission is occasionally quite long, especially if some packets are large and require a long transmission time. As a consequence, voice-over-IP and web browsing do not mix well in a Wi-Fi network. Experiments show that the delays become excessive for voice-over-IP as soon as more than four or five web browsing connections are active in the network [18].

Some technological solutions were developed to try to mitigate this problem. However, they tend to be complicated and incompatible with the widely available technology. A Paris Metro pricing approach would be to run two networks, on non-overlapping radio channels and with different prices. One network would be too expensive for web browsing and would only be used by voice-over-IP users. The other network would be used by web browsers. Wi-Fi networks have a few non-overlapping channels. Typically, the network manager allocates the channels to the access points to limit interference, using some spatial reuse that cannot be perfect. In a Paris metro scheme, one would have a set of access points on one channel with an expensive connection fee, and other access points using the remaining non-overlapping channels. This scheme would support voice-over-IP with a good quality because that application generates little traffic, so that the lack of spatial reuse would not matter. The other applications would suffer somewhat because their network has been deprived of one channel.

A similar situation arises for other applications. For instance, imagine an application that tolerates only very short delays but that does not generate a lot of traffic. One such application could be a class of quick-reflex networked games. Another could be some remote control applications. Such highly delay-sensitive applications are not compatible with web-browsing or with most other applications. A

Paris metro scheme would divide the network into two parallel networks, one being expensive per byte and having a relatively small capacity; the other being cheap and having most of the capacity of the original network. Because of the negligible amount of traffic that these delay-sensitive applications generate, they have essentially no impact on the other applications. Accordingly, this scheme generates additional revenues from the new applications with no loss of revenue from the others. The increase in revenue depends on the demand for the new applications.

Conclusions

The Paris metro pricing scheme provides quality of service by selecting the prices so that the applications automatically find a network with a utilization that results in suitable performance measures. A more expensive network has a lower utilization and, accordingly, lower delays and more throughput for each of its connections. It may be possible to adjust prices to guarantee that the performance measures remain adequate for the applications even if the demand increases. The advantage of the approach is that the technologies of the two parallel networks are identical. The users choose which network to use and the routers and switches do not differentiate packets. This approach contrasts with the a standard implementations of quality of service where routers and switches use different scheduling rules for different classes of packets.

A variation of the implementation of this scheme is to have users mark packets as type 1 or type 2 and charge them differently for the two types. In this variation, the routers can serve the packets of type 1 with high priority and those of type 2 with low priority. The technology for such differentiated service is deployed in the routers (it is called DiffServ). However, this technology is typically not used for end-users.

Why is a Paris metro scheme not implemented today? We have seen that it has the potential to increase operator revenues and to enable new applications for which there is probably a demand. One possible reason is that the implementation of such a scheme requires agreements between network operators about the meaning and pricing of the service types. Moreover, transport providers must agree on how to share the revenue. Finally, there is uncertainty about the additional revenue that the new service would generate.

3.2 Pricing of Services

WiMAX is a wireless cellular broadband access technology that enables the operator to define multiple services, from guaranteed bit rate to best effort. How much more should the operator charge for a guaranteed 1Mbps service than for a best effort service? In this section, we explain that pricing of services is complicated but has a substantial impact on user satisfaction and on operator revenue.

The interaction of a few effects make network pricing different and more complicated than pricing other goods and services:

(1) Customers vary in their valuation of services and providers may want to offer multiple services to capture that diversity;
(2) Congestion changes the quality of services as more people use them;
(3) Heterogeneity in valuations makes congestion externalities asymmetric;
(4) Service providers compete for customers.

Other products or services have some of the above effects, but networks have all of them interacting. For instance, there is a considerable literature on effect (1) that studies how to create versions of a product to ensure that high-end consumers buy the deluxe version and low-end consumers buy the basic version. We explore effect (2) in Section 3.2.1 where we demonstrate that, because of the negative externality of congestion, selfish users tend to over-consume. In Section 3.2.2 we show that congestion pricing can *internalize* the externality. That is, congestion pricing can make the selfish agent pay for the cost his usage imposes on other agents and force him to behave in a way that is socially optimal. Section 3.2.3 illustrates that effect when users can choose when they use the network. In Section 3.2.4 we explore the pricing of different services. When effects (1)-(2)-(3) are combined, the quality of the different services changes as a function of consumer choices, which also depend on the quality of the services. We explore that combination in our discussion of the Paris metro pricing scheme. We show that when effect (4) is included, the resulting pricing game may not have a pure-strategy Nash Equilibrium (a concept that we review in that section). Finally, in Section 3.2.5 we study bandwidth auctions where users compete for acquiring transmission capacity. The text [10] offers a comprehensive discussion of pricing of network services. See also [41] for a nice presentation of related recent results on pricing and competition of transport providers.

3.2.1 Tragedy of the Commons

We expect that agents who are not sensitive to the negative impact of their consumption on the utility of other agents tend to over-consume. We use a simple mathematical model to illustrate that effect called the *tragedy of the commons*.

The tragedy of the commons refers to the unfortunate inevitable result of people sharing a common good. Following Hardin [16], picture a pasture shared by many herdsmen that keep cattle in it. Each herdsman observes that he benefits more than he loses by adding one head of cattle. Indeed, he then gets to raise and sell one more animal whereas his herd suffers only a fraction of the loss due to the additional overgrazing. The inevitable outcome is then that the herdsmen add an excessive number of animals in the pasture. This story is another example of externality. Each herdsman imposes a total externality on the others that exceeds his own benefit increment.

Model

To cast the tragedy of the commons in a simple model, imagine that N identical users share a network and designate by $x_n \geq 0$ the level of activity of user n ($n = 1, \ldots, N$). Assume that user n derives a utility $U(x_n)$ from using the network but also faces a disutility $x_1 + \cdots + x_N$ due to the congestion in the network. (We could use more complicated model of the congestion disutility as a function of the total activity level, but this simple model suffices for our purpose.) That is, for $n \in \{1, \ldots, N\}$, user n has a net utility

$$U(x_n) - (x_1 + \cdots + x_N).$$

We assume that $U(\cdot)$ is an increasing and strictly concave function.

Say that user n chooses his activity level x_n to maximize his net utility. In that case, he chooses $x_n = \tilde{x}$ where $\tilde{x}$ is such that $U'(\tilde{x}) = 1$, as shown in Figure 3.4.

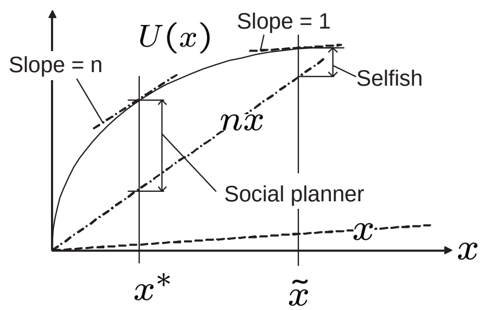

Fig. 3.4 Selfish users choose $\tilde{x}$ whereas the socially optimal choice is x^*.

The net utility of each user is then $R_A := U(\tilde{x}) - N\tilde{x}$.

Since all the users are symmetric, one expects the social optimal activity levels to be identical. If all the users agree to choose the same activity level x^* that maximizes $U(x) - Nx$, then $U'(x^*) = N$ and each user's net utility is $R_S := U(x^*) - Nx^*$ (see Figure 3.4). We see that *selfish users over-consume* because they are not penalized fully for the externality of their consumption.

The *price of anarchy* (a concept introduced in [24]; see also [45], [22] and [1]), defined as the ratio of the maximum social welfare over that achieved by selfish users, is

$$\pi = \frac{R_S}{R_A} = \frac{U(x^*) - Nx^*}{U(\tilde{x}) - N\tilde{x}}.$$

The price of anarchy is unbounded in this model. For a given N one can find a function U so that π is as large as one wants.

3.2.2 Congestion Pricing

Selfish users overconsume because they do not fully pay the cost that their consumption imposes on other users. To correct the situation, imagine that each user has to pay an explicit cost equal to $p = (N-1)x$ when his activity level is x.

The net utility of user n is now

$$U(x_n) - (x_1 + \cdots + x_N) - (N-1)x_n.$$

To maximize the net utility, user n selects an activity level $x_n = x^*$, which is socially optimal. The cost $(N-1)x_n$ is the congestion cost that user n imposes on the others by choosing an activity level x_n. Indeed, each of the $N-1$ other users sees an additional congestion cost equal to x_n because of the activity of user n. This congestion price *internalizes* the externality.

Note that the choice $x_n = x^*$ is socially optimal in that it maximizes the net utility $U(x_n) - (x_1 + \cdots + x_N)$ of each user n. This is the utility without subtracting the price p that every user pays.

For a discussion of congestion pricing in the Internet, see [23].

3.2.3 When to Use the Network?

Instead of choosing their activity level, assume that users may choose when to use the network, say during the day. Following [30], we model the problem as follows. There is a large population of users who can choose when to use the network among $T+1$ periods $\{0, 1, \ldots, T\}$ where period 0 means not using the network at all. The users belong to N different classes characterized by their dislike for time slots. The quantity g_t^n measures the disutility of slot t for users of class n. In particular, g_0^n is the disutility for a user of class n of not using the network.

Designate by x_t^n the number of users of class n that use the network during time slot t. The utility to some user of class n who uses the network in period t is

$$u = u_0 - [g_t^n + d(N_t)1\{t > 0\} + p_t], t = 0, 1, \ldots, T$$

where u_0 is the maximum utility for using the network, $N_t = \sum_n x_t^n$, $d(N_t)$ is a congestion delay of period t, and p_t is the price that users are charged for using period t. In this expression, $d(.)$ is a given increasing strictly convex differentiable function. Note that there is no congestion cost for time slot 0 as the term $1\{t > 0\}$ is equal to zero when $t = 0$ and to 1 when $t > 0$. By convention, $p_0 = 0$. We then have the following result.

Theorem 3.1. *Assume that each user is a negligible fraction of the population of users in each time slot that is used. Assume also that*

$$p_t = N_t d'(N_t), t = 1, \ldots, T.$$

Then selfish users choose the socially optimal time slots, i.e., the time slots that maximizes the total utility $Nu_0 - V(\mathbf{x})$ *where*

$$V(\mathbf{x}) := \sum_{t=0}^{T} [\sum_{n=1}^{N} x_t^n g_t^n + N_t d(N_t) 1\{t > 0\}].$$

Note that the price p_t *is the incremental total disutility per increase of one user in time slot* t *since each of the* N_t *users faces an increase in disutility equal to* $d'(N_t)$*. This price does not depend on the preferences* g_t^n*, which is fortunate for otherwise the scheme would not be implementable.*

Proof. Selfish users choose the time slots that offer the smallest disutility. Thus, at equilibrium, the disutility of all the slots is the same for a user of class n, thus reaching what is called a Wardrop equilibrium [52]. Designate by λ_n that disutility. Accordingly, we have

$$g_t^n + d(N_t) 1\{t > 0\} + p_t = \lambda_n \text{ for } t \in \{0, \ldots, T\} \text{ and } n \in \{1, \ldots, N\}. \tag{3.1}$$

The social optimization problem is as follows:

$$\text{Minimize } V(\mathbf{x})$$
$$\text{Subject to } \sum_{t=0}^{T} x_t^n = x^n \text{ for } n = 1, \ldots, N$$

where x^n is the total population of class n.

The problem is convex, so that the following KKT conditions characterize the solution. There are some constants λ_n such that the partial derivative with respect to x_t^n of

$$L(\mathbf{x}, \lambda) := V(\mathbf{x}) - \sum_{n=1}^{N} \lambda_n [\sum_{t=0}^{T} x_t^n - x^n]\}$$

is equal to zero for all n and t.

That condition states that the Lagrangian is stationary. To understand why the condition is necessary, note that, if the derivative of $V(\mathbf{x})$ with respect to x_t^n is smaller than that with respect to x_s^n, then one could decrease $V(\mathbf{x})$ by slightly increasing x_t^n and decreasing x_s^n by the same amount. Consequently, if x^n is optimal, the derivatives of $V(\mathbf{x})$ with respect to x_t^n must have the same value, say λ_n, for all t. It follows that the derivative of $L(\mathbf{x}_t, \lambda)$ must be equal to zero for some λ_n. The sufficiency of the conditions results from the convexity that implies the uniqueness of the optimal $\mathbf{x}$. (See e.g., [7] or [8] for a presentation of the theory of convex optimization.)

Now observe that the conditions (3.1) that characterize the selfish choices of the users imply that the KKT conditions are satisfied. To see this, note that

$$\frac{d}{dx_t^n} L(\mathbf{x}_t, \lambda) = g_t^n + d(N_t) + \sum_{n=1}^{N} x_t^n d'(N_t) - \lambda_n = g_t^n + d(N_t) + p_t - \lambda_n = 0,$$

by (3.1).

This result shows once again that *internalizing the externality* of each user's choice leads selfish users to make socially optimal choices.

For a study of pricing with asymmetric information between user and provider, see [35].

Congestion Game

In our example, users choose which network to use and the utility of a network depends on how many users also select it. This situation is a particular case of a *congestion game* [44]. In a congestion game, there is a set A of resources and a number N of users. Each user i selects a set $A_i \subset A$ of resources so that there are $n_j = \sum_{i=1}^{N} 1\{j \in A_i\}$ users of resource j, for $j \in A$. The utility of user i is then

$$u_i = \sum_{j \in A_i} g_j(n_j).$$

That is, user i derives some additive utility for each of the resources he uses and the utility of a resource is a function of the number of users of the resource. We assume that each $g_j(n)$ is positive, strictly convex and decreasing in n.

We explain that these games have a nice structural property: as each user tries to increase his utility, he increases a global function of the allocation. That function is called the *potential* of the game; it is concave in this case. Thus, as users act selfishly, they increase the potential that then converges to its maximum which corresponds to the unique Nash equilibrium for the game. By definition, a set of strategies for the users is a *Nash equilibrium* when no user benefits from changing his strategy unilaterally (see e.g., [14]).

Assume that user i changes his selection from A_i to A'_i and that the other users do not change their selection. Designate by u'_i the new utility of user i and by n'_j the number of users of resource j under the new selection. Let $B_i = A_i \setminus A'_i$ and $C_i = A'_i \setminus A_i$. Observe that $n_j = n'_j - 1$ for $j \in C_i$ and $n'_j = n_j - 1$ for $j \in B_i$. Also, $n_j = n'_j$ for $j \notin B_i \cup C_i$. Finally, let

$$\phi := \sum_j f_j(n_j) \text{ with } f_j(n) := \sum_{k=0}^{n} g_j(k)$$

and let ϕ' be the value that corresponds to the new selection. Then we find that

$$f_j(n_j) - f_j(n'_j) = \begin{cases} g_j(n_j), \text{ for } j \in B_i \\ -g_j(n'_j), \text{ for } j \in C_i \\ 0, \text{ for } j \notin B_i \cup C_i. \end{cases}$$

Consequently,

$$u_i - u_i' = \sum_{j \in A_i} g_j(n_j) - \sum_{j \in A_i'} g_j(n_j') = \sum_{j \in B_i} g_j(n_j) - \sum_{j \in C_i} g_j(n_j')$$
$$= \sum_j [f_j(n_j) - f_j(n_j')] = \phi - \phi'.$$

This calculation shows that if user i changes his selection to increase its utility u_i, he also increases the value of ϕ. Thus, a congestion game is a potential game in that it has a potential. The converse is also true: potential games are congestion games [34].

Thus, as the users selfishly modify their selection to increase their utility, the value of ϕ increases to its maximum. Once the selections A_i achieve the maximum value of ϕ, no change in selection by a user would increase his utility. Consequently, the *natural dynamics* of the game converge to the Nash equilibrium (which may not be the social optimum).

3.2.4 Service Differentiation

Coming back to the Paris metro model, assume that you have to choose between first and second class before you get to the metro and that you cannot change when you get there. This assumption is certainly contrived for a metro since you see the occupancy of the different classes when you get to the train. However, the assumption might be reasonable for selecting a service class in a communication network. Thinking that the second class might be very crowded, you are tempted to buy a first class ticket. However, if the price difference is small, it might be that most users buy a first class ticket and end up worse off than in second class. On a future trip, you learn from that mistake and buy a second class ticket. If all users behave like you, you are out of luck again. This is frustrating for the users and probably not in the interest of the provider. We examine this situation in this section. See [17] for a related study and a discussion of revenue sharing among network providers.

One Network

Consider a communication system with a large population of N users each characterized by a type θ that is an independent random variable uniformly distributed in $[0,1]$. A user of type θ finds the network connection acceptable if the number of users X using the network and the price p are such that

$$\frac{X}{2N} \leq 1 - \theta \text{ and } p \leq \theta.$$

In this expression, $2N$ is the capacity of the network. We use $2N$ because we will later divide the network into two networks, each with capacity N. The interpretation is that a user with a large value of θ is willing to pay quite a lot for the connection but he expects a low utilization for a high quality of service. Conversely, a user with a small value of θ does not want to pay much for his connection but is willing to tolerate high delays. For instance, we can think of users with a large θ as users of VoIP and those with a small θ as web browsers.

We solve the provider revenue maximization problem. That is, we find the price p that maximizes the product of the number of users of the network times the price p. Because the utility depends on the utilization and the utilization depends on the utility, we have to solve a fixed point problem to find the utilization that corresponds to a price p.

Assume that the network connection price is $p \in (0,1)$. If the number of users in the network is X, then a user of type θ connects if the inequalities above are satisfied, i.e., if $\theta \in [p, 1 - X/(2N)]$. Since θ is uniformly distributed in $[0,1]$, the probability that a random user connects is $(1 - X/(2N) - p)^+$. Accordingly, the number X of users that connect is binomial with mean $N \times (1 - X/(2N) - p)^+$, so that

$$\frac{X}{N} \approx (1 - \frac{X}{2N} - p)^+$$

by the law of large numbers, since N is large. Solving this expression we find that $x := X/N = (2-2p)/3$. The operator can maximize his revenues by choosing the value of p that maximizes $px = p(2-2p)/3$. The maximizing price is $p = 1/2$ and the corresponding value of px is $1/6$, which measures the revenue divided by N.

Paris Metro

Consider now a Paris metro situation, as discussed in the Introduction, where the operator divides the network into two subnetworks, each with half the capacity of the original network. That is, there are two networks: network 1 with price p_1 and capacity N and network 2 with price p_2 and capacity N. We expect the users to select one of the two networks, based on the prices and utilizations.

To be consistent with the previous model, each of the two networks is acceptable for a user of type θ if the utilization X of the network is such that $X/N \leq 1 - \theta$ and if the price p is such that $p \leq \theta$. Indeed, the capacity of each network is N, so that the ratio of the number of users to the capacity is X/N instead of $X/(2N)$ in the previous model. This ratio determines the quality of service in the network. A user joins an acceptable network if any. Moreover, he chooses the cheapest network if both are acceptable. Finally, if both networks are acceptable and have the same price, a user joins the one with the smallest utilization because it offers a marginally better quality of service.

We determine the prices p_1 and p_2 that maximize the revenue of the operator. We then compare the maximum revenue to the revenue in a single network. Our analysis shows that the Paris metro scheme increases the revenue of the operator by 35%. To perform the analysis, we consider separately the cases $p_2 < p_1$ and $p_1 = p_2$.

First assume that $p_2 < p_1$. If the numbers of users in the two networks are X_1 and X_2, respectively, then a user of type θ chooses network 2 if $X_2/N \leq 1 - \theta$ and $p_2 \leq$

θ. The probability that θ falls between p_2 and $1 - X_2/N$ is then $(1 - X_2/N - p_2)^+$. Arguing as in the case of a single network in the previous section, we conclude that $x_2 := X_2/N$ is given by

$$x_2 = \frac{1}{2}(1 - p_2). \tag{3.2}$$

A user of type θ selects network 1 if $X_1/N \le 1 - \theta, p_1 \le \theta$, and $X_2/N > 1 - \theta$. Arguing as before, we find that $x_1 := X_1/N$ is such that

$$x_1 = (1 - x_1 - \max\{p_1, 1 - x_2\})^+.$$

Substituting the value of x_2 and solving for x_1, we find

$$x_1 = \min\{\frac{1 - p_1}{2}, \frac{1 - p_2}{4}\}. \tag{3.3}$$

The revenue $R \times N$ of the operator is then such that

$$R = x_1 p_1 + x_2 p_2 = p_1 \min\{\frac{1 - p_1}{2}, \frac{1 - p_2}{4}\} + p_2 \frac{1}{2}(1 - p_2)$$

when $p_2 < p_1$.

Second, assume that $p_2 = p_1$, then a user of type θ with $\theta \ge p_1$ and $X_1/N \le 1 - \theta, X_2/N \le 1 - \theta$ selects the network with the smallest number of users. In that case, $x_1 = x_2 = x$ and we find that half of the users with $\theta \in [p_1, 1 - x]$ join network 1. Consequently, the number X_1 of users that join network 1 is such that

$$\frac{X_1}{N} = \frac{1}{2}(1 - \frac{X_1}{N} - p_1)^+,$$

and similarly for X_2. Consequently, $x_1 = (1 - x_1 - p_1)^+/2$ and one finds

$$x_1 = x_2 = \frac{1 - p_1}{3}. \tag{3.4}$$

The revenue $R \times N$ is then such that

$$R = 2p_1 \frac{1 - p_1}{3}$$

when $p_2 = p_1$.

Maximizing R over p_1 and p_2, we find that the maximum occurs for $p_1 = 7/10$ and $p_2 = 4/10$ and that the maximum is equal to $9/40$.

The example shows that the service differentiation with Paris metro pricing increases the revenue from $1/6$ to $9/40$, or by 35%.

Competition

Our previous example shows that two networks with the same capacity and different prices generate more revenue than a single network with twice the capacity. Now assume that the two networks belong to two competing operators. Will one operator settle on a network with high price to attract users of high-quality services and the other on a network with low price to attract users of lower-quality services? We know that this strategy would yield the maximum total revenue. However, this maximum revenue corresponds to the revenue $R_i := x_i p_i$ for network $i = 1, 2$ with $R_1 = 21/200$ and $R_2 = 12/100$. That is, the low-price network generates more revenue than the high-price network. We then suspect that both operators would compete to have the low-price network, which might lead to a price war. However, if the prices become too low, the operators might prefer to raise the price to serve the users of high-quality services. We explore this situation in more detail. In our model, we find that the operators try to segment the market but that there may not be prices that they find satisfactory (no pure Nash equilibrium).

To avoid technical complications that are not essential, assume that the prices p_1 and p_2 are restricted to multiples of $\varepsilon := 1/N$ for some large N. This makes the game finite and guarantees the existence of a best response.

Assume that p_2 is fixed. If $p_1 > p_2$, then using (3.3) we find that

$$R_1 = p_1 x_1 = p_1 \min\{\frac{1-p_1}{2}, \frac{1-p_2}{4}\}. \tag{3.5}$$

Also, if $p_1 = p_2$, from (3.4) we obtain

$$R_1 = p_1 \frac{1-p_1}{3}. \tag{3.6}$$

Finally, if $p_1 < p_2$, then

$$R_1 = p_1 \frac{1-p_1}{2}. \tag{3.7}$$

These expressions show that there is no pure-strategy Nash equilibrium. That is, there is no pair (p_1, p_2) from which no operator can deviate without decreasing its revenue. To see this, assume that (p_1, p_2) is a Nash equilibrium. We claim that $p_1 \neq p_2$. Indeed, if $p_1 = p_2$, then comparing (3.6) and (3.7) shows that operator 1 can increase R_1 by replacing p_1 by $p_2 - \varepsilon$. This is the price war that we anticipated. Assume now that $p_1 \neq p_2$. Comparing (3.5) and (3.7), together with some simple algebra, shows that the maximizing value of R_1 is $p_1 \approx (1 + p_2)/2$ if $p_2 \leq 1/3$. (More precisely, p_1 is the multiple of ε closest to $(1 + p_2)/2$.) If $p_2 > 1/2$, then R_1 is maximized by $p_1 \approx 1/2$. Finally, if $1/3 < p_2 \leq 1/2$, then the value of R_1 is maximized as $p_1 = p_2 - \varepsilon$.

Figure 3.5 illustrates the best response function $p_1(p_2)$ and the symmetric best response function $p_2(p_1)$ and shows that they do not intersect.

By trying to use the best response to a price of the competitor, an operator realizes that his price should be between $1/3$ and $1/2$. However, there is not pair of prices

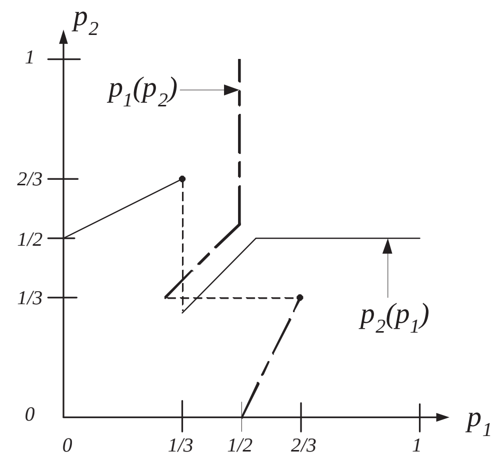

Fig. 3.5 The figure shows the best responses of the two providers: $p_2(p_1)$ and $p_1(p_2)$. Since the best responses do not intersect, the game has no pure-strategy Nash equilibrium.

from which the selfish operators have no incentive to deviate. See e.g. [14], [15], [40], [42], or [37] for an introduction to game theory.

The lesson of this example is that the pricing of services by competing operators can be quite complex and may not have a satisfactory solution. The situation would be simpler if the operators could collude and agree to split the total revenue they get by charging 4/10 and 7/10. The users would also be better served by such an arrangement.

See [1] and [24] for related results on the price of anarchy in pricing of competitive services.

3.2.5 Auctions

Auctions are an effective technique for eliciting the willingness to pay of potential buyers. Instead of having a set price for an item and hoping that one buyer will purchase the item at that price, the auction makes the potential buyers compete against one another. We review some standard results on auctions and then apply them to networks. For a presentation of the theory of auctions, see [25] and [38]. For related models, see [11], [31] and [32].

Vickrey Auction

Assume there is one item for sale and N potential buyers. Each buyer $i = 1, \ldots, N$ has a private valuation v_i for the item. The rules of the *Vickrey auction* are that the highest bidder gets the item and pays the second-highest bid. These rules are similar to those of ascending auctions where the last bidder gets the item and pays essentially what the second-highest bidder was willing to pay. The remarkable property of this auction is that the best strategy for each agent is to bid his true valuation, independently of the strategies of the other agents. To see this, consider the net utility $u_i(x_i, x_{-i})$ (valuation minus payment) of agent i when he bids x_i and the other agents' bids are represented by the vector $x_{-i} = \{x_j, j \neq i\}$. One has

$$u_i(x_i, x_{-i}) = [v_i - w_{-i}]1\{x_i > w_{-i}\}$$

where $w_{-i} := \max_{j \neq i} x_j$ is the highest bid of the other agents. Note that this function is maximized by $x_i = v_i$. Indeed, if $v_i > w_{-i}$, then the value for $x_i = v_i$ is the maximum $v_i - w_{-i}$. Also, if $v_i \leq w_{-i}$, then the maximum value is 0 and it is also achieved by $x_i = v_i$. We say that this auction is *incentive-compatible* because it is in the best interest of each agent to bid truthfully. Note that if the agents bid their true valuation, then the item goes to the agent who values it the most. In that case, the allocation of the item maximizes the social welfare, defined as the sum of the utilities that the agents derive by getting the item. Only one agent gets the item, so the social welfare is the valuation of the item by that agent. See [50].

Generalized Vickrey Auction

Consider the following generalization of the Vickrey auction [33]. There is a set A of items and N agents. Each agent i has a private valuation $v_i(S)$ for each subset S of A. For $i = 1, \ldots, N$, agent i announces a valuation $b_i(S)$ for every subset S of A. The auctioneer then allocates disjoint subsets A_i of A to the agents $i = 1, \ldots, N$ to maximize the sum $\sum_i b_i(A_i)$ of the declared valuations, over all possible choices of such disjoint subsets. Agent i has to pay a price p_i for his subset A_i. The price p_i is the reduction in the declared valuation of the other agents caused by agent i's participation in the auction. That is, if agent i did not bid, then agent j would receive a subset B^i_j and the total valuation of the agents other than i would be $\sum_{j \neq i} b_j(B^i_j)$ instead of $\sum_{j \neq i} b_j(A_j)$. The reduction in valuation is then

$$p_i := \sum_{j \neq i} b_j(B^i_j) - \sum_{j \neq i} b_j(A_j).$$

Thus, p_i is the externality of agent i on the other agents.

The claim is that each agent should bid his true valuations for the subsets. More precisely, that strategy dominates all other strategies, no matter what the other agents bid. To see this, first note that if agent i bids $v_i(.)$ his net payoff is

$$\alpha = v_i(A_i) - \sum_{j \neq i} [b_j(B_j^i) - b_j(A_j)].$$

Also, if agent i bids $b_i(.)$ the allocations are $\{A'_j\}$ and agent i's net payoff is

$$\beta = v_i(A'_i) - \sum_{j \neq i} [b_j(B_j^i) - b_j(A'_j)].$$

(Note that the B_j^i are the same in both cases since they do not involve i's bid.) The difference is

$$\alpha - \beta = [v_i(A_i) + \sum_{j \neq i} b_j(A_j)] - [v_i(A'_i) + \sum_{j \neq i} b_j(A'_j)] \geq 0$$

since the A_j's maximize the first sum.

Thus, this auction mechanism is incentive-compatible. Each user should bid truthfully. Consequently, the allocation maximizes the social welfare. The number of partitions of a set with M elements into $N < M$ subsets is exponential in M. Accordingly, the calculation of the optimal allocation and of the prices is generally numerically complex. However, if the problem has some additional structure, then the solution may in fact be quite tractable, as the next section illustrates.

Bidding for QoS

The mechanism we describe here is described in [47]. Consider a network that offers C classes of service characterized by a different bit rate, as could be implemented in WiMAX, for instance. Say that class c can accept n_c connections and offers a bit rate $r(c)$ with $r(1) > r(2) > \cdots > r(C) \geq 0$. There are N users that compete for access using an auction mechanism. Each user i has a valuation $v_i(r)$ for rate $r \in \{r(1), \ldots, r(C)\}$, where $v_i(r)$ is strictly increasing in r. User i declares a valuation $b_i(r)$ for the possible rates r. The network operator allocates the service classes to the users so as to maximize the sum of the declared valuations. This sum is

$$V := \sum_{i=1}^{N} b_i(c(i))$$

where $c(i)$ is the class of service of user i. If user i were not present, then user j would receive the service class $c^i(j)$ instead of $c(j)$. The price that user i has to pay is

$$p_i := \sum_{j \neq i} b_j(c^i(j)) - \sum_{j \neq i} b_j(c(j)).$$

In this level of generality, the problem is numerically hard. Let us assume that $v_i(r) = v_i f(r)$ and $b_i(r) = b_i f(r)$ where $f(r)$ is an increasing function. User i declares a coefficient value b_i that may not be correct. In this case the allocation that maximizes V is as follows. Assume without loss of generality that $b_1 \geq b_2 \geq \cdots \geq$

b_N. The operator allocates the first class to the first n_1 users, the second class to the next n_2 users, and so on. To see why this allocation maximizes V, consider an allocation where two users i and j with $i < j$ are such that $c(j) < c(i)$. Modify the allocation by interchanging the classes of the two users. Letting $f_i := f(r(c(i)))$, we see that the sum V then increases by $b_i f_j + b_j f_i - b_i f_i - b_j f_j \geq 0$. Note that if user i were not bidding, then some user j_1 would move from class $c(i) + 1 =: d + 1$ to class $c(i) = d$ and see his declared utility increase by $b_{j_1} f_d - b_{j_1} f_{d+1}$. Also, some user j_2 would move from class $d + 2$ to class $d + 1$, and so on. Consequently, the externality of user i is

$$p_i = \sum_{k \geq 1} b_{j_k} [f_{d+k-1} - f_{d+k}].$$

Bandwidth Auction

We describe a mechanism introduced in [54] that the authors call a *VCG-Kelly Mechanism*. See also [22], [31], [32] and [46] for related ideas. An operator has a link with capacity C that she wants to divide up among a set of N users. The goal of the operator is to maximize the sum of the utilities of the users. For instance, the link could be owned by a city that wants to use it to improve the welfare of its citizens. The difficulty is that the operator does not know the utility of the users and that they may declare incorrect utilities to try to bias the capacity allocation in their favor.

Each user i has utility $u_i(x_i)$ when he gets allocated the rate x_i. Here, $u_i(\cdot)$ is an increasing strictly convex function. Thus, the goal of the operator is to find the allocations x_i that solve the following social welfare maximization problem:

$$\begin{aligned} &\text{Maximize} \sum_i u_i(x_i) \\ &\text{over } x_1, \ldots, x_N \\ &\text{subject to } x_1 + \cdots + x_N \leq C. \end{aligned} \tag{3.8}$$

The auction rules are as follows. Each user i bids $b_i > 0$. The operator then implements a Vickrey mechanism assuming that the utility of user i is $b_i \log(x_i)$. The claim is that the unique Nash equilibrium solves problem (3.8). Thus, remarkably, the users do not have to reveal their actual utility function to the operator.

More precisely, the operator selects the allocations $x_i = x_i^*$ that solve the following problem:

$$\begin{aligned} &\text{maximize } \sum_{i=1}^{N} b_i \log(x_i) \\ &\text{over } x_1, \ldots, x_N \\ &\text{subject to } x_1 + \cdots + x_N \leq C. \end{aligned}$$

The problem being convex, we know that x^* is the solution of the first order KKT conditions:

$$b_i \frac{1}{x_i^*} = \lambda, i = 1, \ldots, N,$$

for some $\lambda > 0$. Consequently, since $\sum_i x_i^* = C$, one finds

$$x_i^* = C \frac{b_i}{B},$$

with $B = b_1 + \cdots + b_N$, so that the users get a rate proportional to their bid.

To calculate the price p_i, the operator performs the same optimization problem, assuming that i does not bid. Designate by $\{x_j^i, j \neq i\}$ the solution of that problem. Then the price p_i is given by

$$p_i = \sum_{j \neq i} b_j \log(x_j^i) - \sum_{j \neq i} b_j \log(x_j^*).$$

When user i is not bidding, the allocations are

$$x_j^i = C \frac{b_j}{B - b_i}.$$

Consequently, we find

$$\begin{aligned} p_i &= \sum_{j \neq i} b_j [\log(\frac{Cb_j}{B - b_i}) - \log(\frac{Cb_j}{B})] = \sum_{j \neq i} b_j \log(\frac{B}{B - b_i}) \\ &= (B - b_i) \log(\frac{B}{B - b_i}). \end{aligned}$$

Thus, the net utility of user i is

$$\begin{aligned} u_i(x_i^*) - p_i &= u_i(\frac{Cb_i}{B}) - (B - b_i) \log(\frac{B}{B - b_i}) \\ &= u_i(\frac{Cb_i}{B_i + b_i}) - B_i \log(\frac{B_i + b_i}{B_i}) \end{aligned}$$

where $B_i := B - b_i$ does not depend on b_i. Accordingly, to maximize his net utility, user i chooses b_i so that the derivative of the expression above with respect to b_i is zero. That is,

$$u_i'(\frac{Cb_i}{B})[\frac{C}{B} - \frac{Cb_i}{B^2}] - \frac{B - b_i}{B} = 0,$$

i.e.,

$$u_i'(\frac{Cb_i}{B}) = \frac{B}{C} =: \lambda.$$

This shows that the bids b_i and the resulting allocations $x_i^* = Cb_i/B$ satisfy the KKT conditions of problem (3.8).

For other mechanisms of bandwidth auction, see [27], [28], [26], [19], [20], and [21].

3.3 Investment Incentives

As we explained in the introduction, if investments of agents have a positive externality on the revenue of the other agents, then one may expect some *free-riding*. That is, each agent ends up investing less than socially optimal because he relies on the investments of the other agents. We explain that effect on simple models in section 3.3.1. We then illustrate the effect in a model of network neutrality in section 3.3.2. We conclude the section with a discussion of free-riding in security investment.

3.3.1 Free Riding

If your neighbor paints his house, your house value typically goes up as the neighborhood becomes generally more attractive. This effect may reduce your incentive to paint your own house. This situation is an example of free-riding. We start with a simple model borrowed from [48] that illustrates the free-riding effect. We then explore a model of joint investments by content and transport providers.

Illustrative Example

Assume that two agents jointly invest in a production. Agent 1 invests x and agent 2 invests y; the resulting revenue is $g(ax+by)$ for each of the agent where $g(\cdot)$ is a strictly concave increasing function and $0 < a < b$. The profits of agents 1 and 2 are

$$g(ax+by) - x \text{ and } g(ax+by) - y,$$

respectively. Acting selfishly, each agent tries to maximize his profit. Thus, given y, agent 1 chooses x so that the derivative of his profit with respect to x is equal to zero if that occurs for some $x > 0$ and chooses $x = 0$ otherwise. That is, agent 1 chooses x such that

$$ag'(ax+by) = 1,$$

or $ax+by = A$ where $g'(A) = 1/a$, if that x is positive. Thus, $x = [A/a - (b/a)y]^+$. Similarly, we find that, given x, agent 2 chooses $y = [B/b - (a/b)x]^+$ where $g'(B) = 1/b$. Figure 3.6 shows these best response functions. As the figure shows, the unique intersection of the best response functions is $x_N = 0, y_N = B/b$, which is then the unique pure-strategy Nash equilibrium. Note that this value of y maximizes

$$g(by) - y.$$

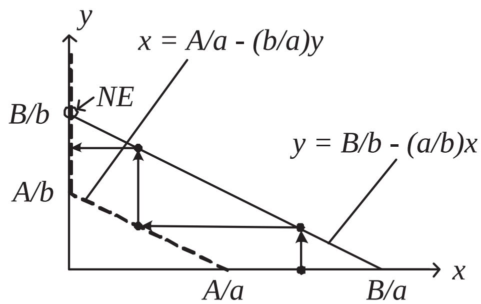

Fig. 3.6 The best response functions intersect at the unique pure-strategy Nash equilibrium $(0, B/b)$.

In this example, agent 1 free-rides on the investment of agent 2 by not investing at all and yet collecting a profit. As you probably suspect, these choices of the agents are not socially optimal. We explore that aspect of the game next.

Consider the sum of the profits of the two agents, the *social welfare* of this model. This sum is

$$W = 2g(ax + by) - x - y.$$

Imagine a *social optimizer* who chooses $x = x^*$ and $y = y^*$ to maximize W. To maximize $ax + by$ for a given value of $x + y$, one must choose $x = 0$ because $a < b$. Accordingly, the social optimizer must find the value of y that maximizes

$$2g(by) - y.$$

Thus, whereas the Nash equilibrium is the value of y that maximizes $g(by) - y$, the social optimal maximizes $2g(by) - y$. That is, in the social optimization, agent 2 knows that his investment contributes to the utility of agent 1 whereas in the Nash equilibrium agent 2 ignores that effect. Figure 3.7 illustrates the results. As the figure

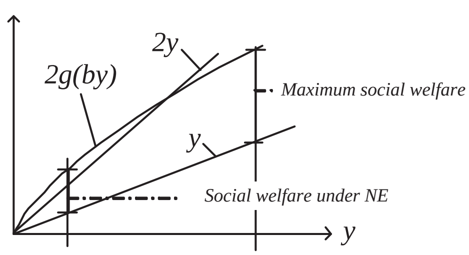

Fig. 3.7 The social welfare under the Nash equilibrium and its maximum value.

suggests, one can find a function $g(\cdot)$ for which the ratio between the maximum social welfare and that achieved by a Nash equilibrium is as large as one wishes.

Cobb-Douglas Example

Assume a content provider C invests c and a transport provider T invests t. As a result of these investments, there is demand for the network services and that demand generates revenues for C and T. Assume that

$$R_C = c^v t^w - c \text{ and } R_T = c^v t^w - t.$$

In these expressions, $v, w > 0$ and $v + w < 1$. The traffic in the network is proportional to $c^v t^w$, a function that is increasing and concave in the investments. That expression, called a Cobb-Douglas function, is commonly used to model the joint production by labor and capital investments [9]. In R_C, the first term is the revenue generated by the traffic on the web site and the term c is the *opportunity cost* that C loses by investing in the network instead of in some other productive activity. The term R_T admits an interpretation similar to that of R_C. One could make the model more general, but this would only complicate notation.

Being selfish, C chooses c to maximize R_C and T chooses t to maximize R_T. That is, for a given t, C chooses c so that the derivative of R_C with respect to c is equal to zero. This gives

$$vc^{v-1}t^w = 1, \text{ or } c = (vt^w)^{1/(1-v)}.$$

Similarly,

$$t = (wc^v)^{1/(1-w)}.$$

The Nash equilibrium is the intersection of these two best response functions. One finds

$$c = v^{(1-w)/\Delta} w^{w/\Delta} \text{ and } t = w^{(1-v)/\Delta} v^{v/\Delta}$$

where $\Delta = 1 - v - w$. The resulting revenues R_C and R_T are

$$R_C = (1-v)v^{v/\Delta} w^{w/\Delta} \text{ and } R_T = (1-w)w^{w/\Delta} v^{v/\Delta}.$$

In particular, the sum of the revenues is

$$R_C + R_T = [2 - v - w]v^{v/\Delta} w^{w/\Delta}.$$

Now assume that C and T are in fact the same operator that can choose c and t to maximize $R_T + R_C$, i.e., to maximize

$$2c^v t^w - c - t.$$

Setting the derivatives with respect to c and t equal to zero, we find

$$2vc^{v-1}t^w = 1 \text{ and } 2c^v wt^{w-1} = 1.$$

Solving these equations we get

$$c^* = 2^{1/\Delta} v^{(1-w)/\Delta} w^{w/\Delta} \text{ and } t^* = 2^{1/\Delta} w^{(1-v)/\Delta} v^{v/\Delta},$$

with

$$\Delta := 1 - v - w,$$

which corresponds to the sum of revenues

$$R_C^* + R_T^* = 2^{1/\Delta} \Delta v^{v/\Delta} w^{w/\Delta}.$$

Consequently, the price of anarchy of free-riding π is given by

$$\pi = \frac{R_C^* + R_T^*}{R_C + R_T} = \frac{2^{1/\Delta}\Delta}{[1+\Delta]}.$$

For instance, π is equal to 7.7 for $\Delta = 0.3$ and to 1.32 for $\Delta = 0.8$. We can also compare the investments of C and T under the socially optimal and the Nash equilibrium.

$$\frac{c^*}{c} = \frac{t^*}{t} = 2^{1/\Delta}.$$

For instance, this ratio is equal to 10 for $\Delta = 0.3$ and to 2.4 for $\Delta = 0.8$.

As we expected, free-riding reduces the investments and the social welfare.

3.3.2 Network Neutrality

In today's Internet, a user or content provider pays the transport provider to which he is directly attached. For instance, if a content provider C is attached to transport provider S, then C pays S a cost that depends on the rate of the traffic that C sends to S. However, to reach the end users of C's content, the traffic from C has to go through the Internet Service Provider (ISP) of those users. Should ISPs be allowed to charge C for transporting that traffic? The ISPs argue that they need to invest in their network to improve the delivery of the content and that C would benefit from those improvements. On the other hand, C argues that the additional charges would reduce his incentive to invest in new content, which would hurt the transport providers' revenue. Thus, the question is whether a *neutral network* where such charges are not allowed increases or reduces the revenue of content and transport providers and the demand for network services.

The question is important as neutrality regulation would have a substantial impact on investment incentives and on the future of Internet. Obviously, the answers depend on the assumptions made implicitly or explicitly in the model. It is certainly irresponsible to claim definite conclusions about such an important question based on cavalier models. That is not our pretention here. Rather, we want only to give a sense of how one can approach the question. We discuss a study developed in [36]. For a background on network neutrality, see [12], [13] and [55].

Model

Our model focuses on the impact on investments and revenues of additional charges paid by content providers to ISPs. Accordingly, the model, illustrated in Figure 3.8 includes a content provider C – that gets revenue from advertisers A – and users U attached to ISP T. The payments in the model are normalized per click on spon-

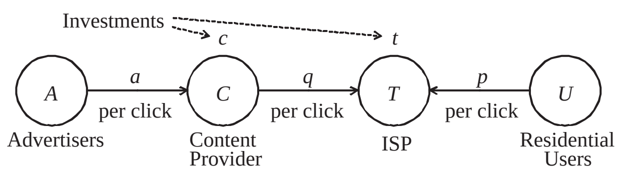

Fig. 3.8 Model to study the impact of neutrality.

sored advertisement. Advertisers pay a per click to the content provider. The content provider pays q and users pay p to the ISP, also per click on sponsored advertisement. Of course, C pays T based on the traffic rate, but that rate is roughly proportional to that of clicks on ads. Also, T gets revenue roughly in proportion to the number of users interested in content, and therefore roughly in proportion to the rate of clicks on ads.

In our model, C invests c and T invests t. The demand for network services increases with the richness of content and the ability of the network to transport it. Thus, the rate of clicks on sponsored ads increases with c and t and it decreases with p as fewer users are willing to pay a higher connection fee that corresponds to a higher value of the price p normalized per click on sponsored ads. We model the rate as

$$B = c^v t^w e^{-p}$$

where, as in the last example of section 3.3.1, $v, w > 0$ with $v + w < 1$. Accordingly, the revenue R_C of C ad the revenue R_T of T are given as follows:

$$R_C = (a - q)B - \alpha c \text{ and } R_T = (q + p)B - \beta t.$$

In these expressions, α and β are positive numbers that model the opportunity costs of the providers.

We consider that, because of different time scales of investment, T first selects (t, p, q) and C then chooses c. Moreover, in a neutral network $q = 0$ whereas in a non-neutral network, q can take any value. A positive value of q is a payment from C

to T. However, we allow q to be negative, which corresponds to a transfer of revenue from T to C.

The analysis then proceeds as follows. In the non-neutral network, assume that (t,p,q) are fixed by T. Then C finds the value $c(t,p,q)$ that maximizes R_C. Anticipating this best response by C, ISP T replaces c by $c(t,p,q)$ in the expression for R_T and then optimizes over (t,p,q). Designate by $(c_1,t_1,p_1,q_1,R_{C1},R_{T1},B_1)$ the resulting values for the non-neutral network. In the neutral network, the approach is identical, except that $q=0$. Designate by $(c_0,t_0,p_0,q_0,R_{C0},R_{T0},B_0)$ the resulting values for the neutral network.

After some algebra, one finds the following results:

$$p_0 = p_1 + q_1 = a(1-v)$$
$$q_1 = a - v$$
$$\frac{R_{C0}}{R_{C1}} = \frac{c_0}{c_1} = (\frac{a}{v})^{(1-w)/(1-v-w)} e^{(v-a)/(1-v-w)}$$
$$\frac{R_{T0}}{R_{T1}} = \frac{t_0}{t_1} = \frac{B_0}{B_1} = (\frac{a}{v})^{v/(1-v-w)} e^{(v-a)/(1-v-w)}.$$

Figure 3.9 illustrates those ratios when $v = 0.5$ and $w = 0.3$, as a function of a. As

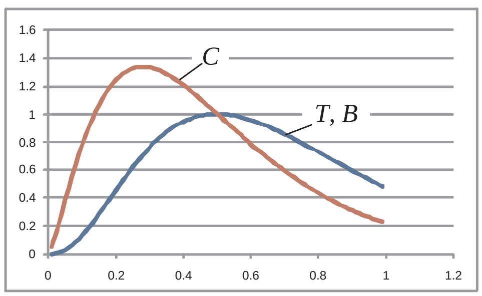

Fig. 3.9 Ratios of investments, revenues, and user demand (neutral/non-neutral) as a function of a.

this example shows, the neutral regime is never favorable for the ISP and it is favorable for the content provider only when the amount a that it can charge to the advertisers is neither large nor small. When a is small, the non-neutral regime is better for the content provider because it enables the ISP to pay him for generating content. Specifically, $q_1 = a - v < 0$ when $a < v$. When a is large, the non-neutral regime is preferable to C because he can provide revenues to T that uses them to improve the network. The results are similar when there are multiple content providers and ISPs [36].

3.3.3 Economics of Security

Many researchers agree that the main bottleneck to a secure Internet is not the absence of cryptographic tools or key distribution mechanisms. Rather, they point to the lack of proper incentives for users. For instance, most denial of service attacks come from many computers that were not properly patched to prevent intrusion. The users of those computers do not see the external cost of their lack of security and, accordingly, they do not bother to install the appropriate security software. Another example is the lack of encryption of private information in laptops or hard drives that can be easily stolen. The users of those devices do not understand the potential cost of their lack of basic precaution. See [5] for a survey of these issues. In this section, we explain one study of security investment and show that free-riding explains the sub-optimal investments. The discussion is borrowed from [29].

Model

Consider a set of N users of computers attached to the network. Designate by x_i the investment in security by user i. We model the utility of user i by $u_0 - u_i(\mathbf{x})$ where

$$u_i(\mathbf{x}) = g_i(\sum_j \alpha_{ji} x_j) + x_i$$

is the *security cost* of user i. In this expression, $\alpha_{ji} \geq 0$ measures the impact of user j's investment on user i's security and $g_i(\cdot)$ is a positive convex decreasing function. Thus, as user i invests more in security measures, such as purchasing software and configuring it on his system, he faces an increased direct cost but he reduces his probability of being vulnerable to an attack and also the probability that other users will be attacked. The impact of user j's investment on user i depends on the likelihood that a virus would go from user j's computer to user i's of the likelihood that user i's confidential information can be stolen from user j's computer.

Given the positive externality, one expects a free-riding effect. We study that effect. Designate by $\mathbf{x}^*$ the vector of efforts that minimizes the social cost $\sum_i u_i(\mathbf{x})$. Let also $\bar{\mathbf{x}}$ be a Nash equilibrium where each user minimizes his individual security cost. We are interested in characterizing the ratio ρ where

$$\rho := \frac{\sum_i u_i(\bar{\mathbf{x}})}{\sum_i u_i(\mathbf{x}^*)}.$$

This ratio quantifies the price of anarchy , which is the factor by which the cost to society increases because of the selfish behavior of users. (Note that this ratio differs from that of the utilities, which we designated previously by π.) One has the following result [29].

Theorem 3.2. *The ratio ρ satisfies*

$$\rho \leq \max_j \{1 + \sum_{i \neq j} \frac{\alpha_{ji}}{\alpha_{ii}}\}.$$

Moreover, the bound is tight.

As a first illustration, assume that $\alpha_{ij} = 1$ for all i, j. Assume also that $g_i(z) = [1 - (1-\varepsilon)z]^+$ for all i where $0 < \varepsilon \ll 1$. In this case, we find that $u_i(\mathbf{x}) = [1-(1-\varepsilon)(x_1 + \ldots + x_N)]^+ + x_i$, for all i. For any choice of $\{x_j, j \neq i\}$, the function $u_i(\mathbf{x})$ is increasing in x_i. Consequently, the unique Nash equilibrium is $\mathbf{x} = \mathbf{0}$, which is an extreme form of free-riding. The total security cost under that Nash equilibrium is $u_1(\mathbf{0}) + \cdots + n_N(\mathbf{0}) = N$. By symmetry, the value $\mathbf{x}^*$ of $\mathbf{x}$ that minimizes $u_1(\mathbf{x}) + \cdots + n_N(\mathbf{x})$ is such that $x_i^* = v/N$ where v minimizes $N[1-(1-\varepsilon)v]^+ + v$. That value of v is $v = 1$ and it corresponds to the total security cost approximately equal to 1. The price of anarchy in this example is $N/1 = N$ which is the value of the upper bound for ρ in the theorem.

As a second illustration, consider a network with $N+1$ identical nodes with N even and $g_i(v) = \delta e^{-\lambda v}$ for all i. We assume $\lambda\delta > 1$ to avoid cases where the optimal investment is zero. To simplify the algebra, picture the $N+1$ nodes arranged consecutively and regularly on a large circle and let $\alpha_{ij} = \beta^{d(i,j)}$ where $d(i,j)$ is the minimum number of hops between i and j. Then $\sum_{i \neq j} \alpha_{ji} = 2(\beta + \beta^2 + \cdots + \beta^{N/2}) = 2(\beta - \beta^{N/2+1})/(1-\beta)$. Assuming that $\beta^{N/2} \ll 1$, one finds that the upper bound is $\rho \leq (1+\beta)/(1-\beta) =: \gamma$. At the social optimal, by symmetry, one has $x_i^* = z$ where z minimizes

$$N\delta \exp\{-\lambda z(1 + 2\beta + 2\beta^2 + \cdots + 2\beta^{N/2})\} + Nz \approx N\delta \exp\{-\lambda\gamma z\} + Nz.$$

Hence, $z = (1/(\lambda\gamma)) \log(\lambda\gamma\delta)$. The resulting total security cost is

$$N(\lambda\gamma)^{-1}(1 + \log(\lambda\gamma\delta)).$$

On the other hand, at the Nash equilibrium, each user selects x_i to minimize

$$\delta \exp\{-\lambda x_i - \lambda \sum_{j \neq i} \alpha_{ji} x_j\} + x_i.$$

Consequently, $\overline{x_i} = u$ where $-\lambda\delta\exp\{-\lambda\gamma u\} + 1 = 0$. Hence, $\overline{x_i} = (1/(\lambda\gamma))\log(\lambda\delta)$. This result shows that free-riding reduces the investment in security from the socially optimal value. At the Nash equilibrium, the total security cost is

$$N\delta e^{-\lambda\gamma u} + Nu = \frac{N}{\lambda} + \frac{N}{\lambda\gamma}\log(\lambda\delta).$$

The ratio of the costs in this example is

$$\rho = \frac{\gamma + \log(\lambda\delta)}{1 + \log(\gamma) + \log(\lambda\delta)}.$$

For numerical values, say that $\lambda\delta \approx 1$. Then Figure 3.10 illustrates the price of anarchy as a function of β. As the graph confirms, the price of anarchy increases

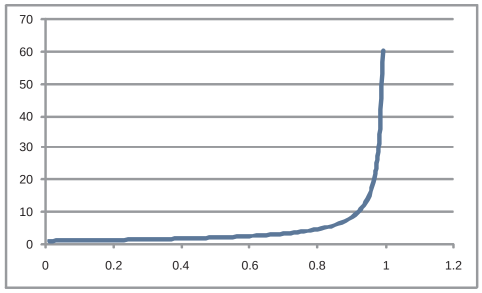

Fig. 3.10 The price of anarchy in security investments as a function of β.

with β because users have more impact on each other.

3.4 Conclusions

The main point of this tutorial is that the economic and technology layers of a communication network interact in a complex way. Accordingly, to understand their combined behavior, we must study these two layers jointly. The design of protocols affects not only the technology layer that determines the performance of the network under given operating conditions. It also affects the operating conditions by impacting the economic layer: the incentives for providers to invest in that technology and for users to use the network and the choices they make.

Our focus is the activity of users and the investments of providers. We stayed away from situations where users cheat with the rules of protocols to gain some strategic advantage. A recurrent theme is the externalities of actions of users and providers. These externalities result in selfish behavior that is not socially optimal. We characterized the social cost of selfish behavior (the price of anarchy) because of excessive usage or insufficient investments. We explored a number of schemes to entice selfish agents to align their behavior with the interest of society. For instance, we discussed congestion pricing and Vickrey auctions. In our analysis, we used simple ideas from game theory and convex optimization. More importantly, we were inspired by concepts from economics.

We used the Paris metro scheme to illustrate the inter-dependency of utilization and performance. In the introduction, we considered an example where the provider can double his revenue by using such a scheme. Remarkably, the Paris metro scheme does not require any QoS mechanism; only splitting the network into two identical networks, each with half the capacity of the original network and both with different prices. We discussed other situations where such a scheme could be employed, when applications are not compatible.

In Section 2, we explored the pricing of services. Our starting observation is that the price should reflect the congestion externality. Otherwise, all users tend to overconsume and they suffer from excessive congestion. We illustrated that point on a model where users select when to use the network. We observed that the pricing does not depend on the preferences of users for when to use the network; that pricing depends only on the congestion. We then revisited the Paris metro scheme using a more general model of user diversity. We examined whether two providers would end up splitting the market by having one choose a high price and the other a low price. That example illustrated the possibility of the absence of a pure-strategy Nash equilibrium. We concluded that section by a discussion of auctions applied to service differentiation.

Section 3 was devoted to investment incentives. After a discussion of the free-riding problem where positive externality leads to under-investments, we explored the delicate question of network neutrality. The conclusion of our simple example is that a good scheme should enable the sharing of revenue that results in the best investment incentives. If content providers are able to generate more revenue than ISPs, it might be beneficial for the content providers to share some of their revenue with the ISPs, and conversely. We concluded the section with a study of incentives for investments in network security. The model focused on externality and the resulting price of anarchy.

We hope that this brief tutorial will motivate you to explore the economic aspects of networks further. We believe that we all will benefit if future network designers have a deeper appreciation of these issues. These questions are the subject of an increasing number of papers in the traditional networking journals. Multiple sessions and special workshops are organized frequently to explore the economics of networks.

Acknowledgements The author thanks his collaborators Venkat Anantharam, Antonis Dimakis, Mark Felegyhazi, Assane Gueye, Linhai He, Rahul Jain, Libin Jiang, Jiwoong Lee, John Musacchio, Shyam Parekh, Galina Schwartz, Nikhil Shetty and Pravin Varaiya for multiple conversations and joint work on the topics of this tutorial and for comments about this material. He also thanks the National Science Foundation for supporting his work under Grant NeTS-FIND 0627161. Finally, the author is grateful to Zhen Liu, Cathy Xia and Sigmetrics for inviting him to present this tutorial and for arranging for its publication.

References

1. D. Acemoglu and A. Ozdaglar, "Competition and Efficiency in Congested Markets," *Mathematics of Operations Research*, vol. 32, no. 1, pp. 1-31, February 2007.
2. G. Akerlof, "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism," *Quarterly Journal of Economics*, 84 (3): 488−500, 1970.
3. E. Altman, T. Boulogne, R. El Azouzi and T. Jimenez, "A survey on networking games in telecommunications," *Computers & Operations Research*, 33, 286−311, 2006.

4. V. Anantharam, "On the Nash dynamics of congestion games with player-specific utility," *Proc. 2004 43rd IEEE Conf. on Decision and Control*, 2004.
5. R. Anderson and T. Moore, "The Economics of Information Security: A Survey and Open Questions," *Fourth bi-annual Conference on the Economics of the Software and Internet Industries*, Toulouse, 1/2007; http://www.cl.cam.ac.uk/ rja14/Papers/toulouse-summary.pdf
6. T. Basar and G.J. Olsder, *Dynamic Noncooperative Game Theory*, Second edition, SIAM, 1999.
7. D. Bertsekas, A. Nedic and A. Ozdaglar, *Convex Analysis and Optimization*, Athena Scientific, Cambridge, Massachusetts, 2003.
8. S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.
9. C.W. Cobb and P.H. Douglas "A Theory of Production," *American Economic Review*, 18 (Supplement), 139−165, 1928.
10. C. Courcoubetis and R. Weber, *The Pricing of Communications Services*, J. Wiley, 2002.
11. C. Courcoubetis, M.P. Dramitinos and G.D. Stamoulis, "An auction mechanism for bandwidth allocation over paths," *ITC-17*, Salvador da Bahia, Brazil, 2001.
12. D. Farber and M. Katz. "Hold Off On Net Neutrality," Washington Post, 1/22/2007; http://www.washingtonpost.com/wp-dyn/content/article/2007/01/18/AR2007011801508.html
13. E. Felten, "The Nuts and Bolts of Network Neutrality," *Princeton*, 2006; http://itpolicy.princeton.edu/pub/neutrality.pdf.
14. D. Fudenberg and J. Tirole, *Game Theory*, MIT Press, 1991.
15. R. Gibbons, *Game Theory for Applied Economists*, Princeton University Press, 1992.
16. G. Hardin, "The Tragedy of the Commons," *Science*, 162: 1243-1248, 1968.
17. L. He and J. Walrand, Pricing and Revenue Sharing Strategies for Internet Service Providers, *IEEE JSAC*, May 2006.
18. N. Hedge, A. Proutiere, and J. Roberts, "Evaluating the voice capacity of 802.11 WLAN under distributed control," *Proc. LANMAN*, 2005.
19. R. Jain, "An Efficient Nash-Implementation Mechanism for Allocating Arbitrary Bundles of Divisible Resources," *IEEE JSAC*, 2007.
20. R. Jain and P. Varaiya, "Combinatorial Exchange Mechanisms for Efficient Bandwidth Allocation," *Communications in Information and Systems*, 3, no. 4, 305−324, 2004.
21. R. Jain, A. Dimakis and J. Walrand, "Mechanisms for efficient allocation in divisible capacity networks," *Proc. Control and Decision Conference (CDC)*, December 2006.
22. R. Johari and J. N. Tsitsiklis, "Efficiency loss in a resource allocation game," *Mathematics of Operations Research*, 29(3): 407-435, 2004.
23. F.P.Kelly, "Charging and rate control for elastic traffic," *European Trans. on Telecommunications*, 8(1): 33-37, 1996.
24. E. Koutsoupias and C. Papadimitriou, "Worst-case equilibria," *Proc. Symp. on Theoretical Aspects of Computer Science*, 16: 404-413, 1999.
25. V.Krishna, *Auction Theory*, Academic Press, 2002.
26. R.J.La and V.Anantharam, "Network pricing using a game theoretic approach," *Proc. Conf. on Decision and Control*, 1999.
27. A. Lazar and N. Semret, "The progressive second price auction mechanism for network resource sharing," *Proc. Int. Symp. on Dynamic Games and Applications*, 1997.
28. A. Lazar and N. Semret, "Design and analysis of the progressive second price auction for network bandwidth sharing," *Telecommunication Systems - Special issue on Network Economics*, 1999.
29. L. Jiang, V. Anantharam, and J. Walrand, "Efficiency of selfish investment in network security," *preprint*, 2008; http://robotics.eecs.berkeley.edu/~wlr/Papers/libin-security.pdf
30. L. Jiang and J. Walrand, "Congestion Pricing of Network Access," *preprint*, 2008; http://robotics.eecs.berkeley.edu/~wlr/Papers/libin-pricing.pdf

31. R.MAHESWARAN AND T.BASAR, "Nash equilibrium and decentralized negotiation in auctioning divisible resources," *J. Group Decision and Negotiation* 12:361-395, 2003.
32. R.MAHESWARAN AND T.BASAR, "Social Welfare of Selfish Agents: Motivating Efficiency for Divisible Resources," *CDC*, 2004.
33. J. K. MACKIE-MASON AND H. R. VARIAN, "Generalized Vickrey Auctions," 1994; citeseer.ist.psu.edu/mackie-mason94generalized.html
34. D. MONDERER AND L.S. SHAPLEY, "Potential Games," *Games and Economic Behavior*, 14, 124143, 1996.
35. J. MUSACCHIO AND J. WALRAND, WiFi Access Point Pricing as a Dynamic Game, *IEEE/ACM Transactions on Networking*, vol.14, no.2, April 2006, pp. 289-301.
36. J. MUSACCHIO, G. SCHWARTZ, AND J. WALRAND, "A Two-Sided Market Analysis of Provider Investment Incentives with an Application to the Net-Neutrality Issue: Long Version," September 2007; http://robotics.eecs.berkeley.edu/~wlr/Papers/MSW-Long.pdf
37. R.B. MYERSON, *Game Theory: Analysis of Conflict*, Harvard University Press, 1997.
38. R. B. MYERSON AND M. A. SATTERTHWAITE, "Efficient mechanisms for bilateral trading," *J. of Economic Theory* 28: 265-281, 1983.
39. A. ODLYZKO, "Paris Metro Pricing for the Internet," *ACM Conference on Electronic Commerce*, 1998.
40. M.J. OSBOURNE AND A. RUBINSTEIN, *A course in game theory*, MIT Press, 1994.
41. A. OZDAGLAR AND R. SRIKANT, Incentives and Pricing in Communication Networks, chapter in *Algorithmic Game Theory*, Noam Nisan, Tim Roughgarden, Eva Tardos, and Vijay Vazirani (Editors), Cambridge University Press, 2007.
42. G. OWEN, *Game Theory*, Third edition, Academic Press, 1995.
43. C.H.PAPADIMITRIOU, "Algorithms, games, and the internet," *Proc. STOCS*, 2001.
44. R. W. ROSENTHAL, "A class of games possessing pure-strategy Nash equilibria," *International Journal of Game Theory*, 2, 6567, 1973.
45. T. ROUGHRADEN AND E. TARDOS, "How bad is selfish routing?" *Proceedings of the 41st Annual Symposium on Foundations of Computer Science*, 2000.
46. S.SANGHAVI AND B.HAJEK, "A new mechanism for the free-rider problem," *IEEE Transactions on Automatic Control*, to appear March 2008.
47. J. SHU AND P. VARAIYA, "Smart pay access control via incentive alignment," *IEEE JSAC*, 24(5): 1051-1060, May 2006.
48. H. VARIAN, "System Reliability and Free Riding," *Proceedings of ICEC 2003*, 355-366, ACM Press 2003.
49. H. VARIAN, J. FARRELL AND C. SHAPIRO, *The Economics of Information Technology: An Introduction*, Cambridge University Press, 2004.
50. W. VICKREY, "Counterspeculation, Auctions, and Competitive Sealed Tenders," *Journal of Finance*, XVI, 8-37, 1961.
51. J. VON NEUMANN AND O. MORGENSTERN, *Theory of Games and Economic Behavior*, Princeton University Press, 1953.
52. J. G. WARDROP, "Some theoretical aspects of road traffic research," *Proceedings of the Institute of Civil Engineers*, vol. 1, pp. 325-378, 1952.
53. H.YAICHE, R.R.MAZUMDAR AND C.ROSENBERG, "A game theoretic framework for bandwidth allocation and pricing in broadband networks," *IEEE/ACM Trans. on Networking* 8(5): 667-678, 2000.
54. S. YANG AND B. HAJEK, "VCG-Kelly mechanisms for allocation of divisible goods: Adapting VCG mechanisms to one-dimensional signals," *IEEE JSAC (Issue on noncooperative behavior in networks)* , vol. 25, pp. 1237-1243, 2007.
55. C.S. YOO, "Network Neutrality and the Economics of Congestion," *Georgetown Law Journal*, Vol. 94, June 2006.