R&D N 86/2000

Bjørn Hansen and Eric Nævdal

Competition in the Internet and Dynamic Pricing by ECN Marks

Title

Competition in the Internet and dynamic pricing by ECN marks

R&D Notat	86/2000
ISBN	
ISSN	**0809 - 1021**
Project no	**DN02M3**
Program	
Security gr.	**Open**
No. of pages	**33**
Date	**00.12.20**

Author(s)

Bjørn Hansen and Eric Nævdal

Subject headings

Dynamic pricing, Internet, Competition

Abstract

In a network where users adjust their bit rate optimally according to their preferences as well as dynamic pricing signals, users can obtain the wanted class of service. E.g. users can obtain a constant bit rate by accepting a fluctuating price, or users can obtain a constant payment per minute by accepting a fluctuating bit rate. In this paper we consider a market for information transfer where ECN (explicit congestion notification) marks are used as dynamic pricing signals. We describe the welfare maximising allocation of network capacity in such a network, and then we proceed to consider the market equilibrium under

Title (Norwegian)

Konkurranse i Internett under dynamisk prising med ECN merker

© Telenor AS 20.12.00

All rights reserved. No part of this publication reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from the publisher.

Contents

Preface

This work is partly financed by Telenor R&D project DN02M3 and partly RTD Project No IST-1999-11429 under the European Union's Fifth Framework Programme.

The authors are grateful for comments from members of the M3I consortium, participants at the M3I QofIS workshop on Internet Charging in Berlin September 2000. Corresponding author: Bjørn Hansen, Telenor Research and Development, P.O. Box 83, N-2027 Kjeller, Norway. E-mail: bjorn.hansen@telenor.com

1 Introduction

Users of the current Internet are charged either on a flat rate basis, by the minute or by volume. All these three pricing regimes will typically lead to inefficient usage patterns. This is due to the fact that capacity in network resources in some periods is superfluous and in other periods is a scarce resource. In periods where capacity in network resources is superfluous, flat rate pricing has nice properties since usage is encouraged, whereas time based and usage based tariffing is leading to inefficiently low usage. This is opposed to periods where capacity in network resources is scarce. In such periods flat rate charging is using a very inefficient allocation mechanism since usage is prioritised according to a “first in, first out scheme” and there is no mechanism to discriminate between traffic with high and low value to the users. There is accordingly a potential for introducing pricing schemes providing incentives for more efficient usage.

Dynamic pricing designed in a proper way might provide such incentives. One possible implementation of dynamic pricing is to use Explicit Congestion Notification (ECN) marks as pricing signals as proposed in a series of papers by Frank Kelly and others (Kelly, 1999, Gibbens and Kelly 1999, Kelly, Maulloo and Tan, 1998). In the present paper we will generalise the model presented in Kelly 1999. Although our results are derived with the ECN scheme in mind, our results will typically be applicable to other dynamic pricing schemes.

The results in Kelly, 1999 are derived under a particular set of assumptions. It is assumed that utility functions are concave and the supplier is some sort of benevolent monopoly. Neither of these assumptions is fulfilled in the current Internet in general. Many applications results in utility being nonconcave in bandwidth. Furthermore, the firms currently providing Internet connectivity is maximising profits and is operating under (imperfect) competition. In this paper we will show that ECN marks can be implemented, also when utility is nonconcave, and the providers are maximising profits.

There is a growing body of literature discussing the pricing of access to- and usage of -Internet resources. A recent overview is provided in Dolan (2000). Among the suggestions are McKie-Mason and Varian (1992) smart market scheme. The basic idea is to let users pay for priority in congested network resources by attaching bids to each packet they send. If the packets traverse the network without passing any congested resources, the price of the packet transfer is zero. If a resource in the network is congested, it is performed a second price auction such that an equilibrium price is established. The network is then placing priority on the packets with the highest bids. Since the auction is of the second price type, it is a dominant strategy for the user to bid his true valuation. Alternatively one can introduce priority classes. Odlyzko (1997) has suggested a scheme called Paris Metro pricing where network capacity is divided into two logically separate networks. The price of access to the two networks is differentiated. The idea is to let the price differentiation lead to product differentiation. Users will expect the high priced network to have higher quality. Suppose users are divided among the networks such that the quality of the two networks is identical. Then some users will switch from the high to the low priced network. This will result in increased quality in the high priced network and reduced quality in the low priced network due to increased congestion. In equilibrium the quality differentiation and thus difference in priority will reflect the pricing differential.

The present paper is organized as follows: In the next section we will reformulate the ECN pricing model such that standard results from economic theory can be deployed. In the subsequent sections we will consider perfect competition, dynamics, monopoly and duopoly. In the final two sections of the paper we consider inelastic users and congestion externalities respectively.

2 The basic model

In this section we will outline the model of pricing by ECN marks as presented in Kelly 1999. After that we will modify it in such a way that the structure and dynamics of the model is unchanged. The modified model will enable a discussion of the topics of this paper, competition stability and nonconcave utility.

2.1 Production technology, costs and ECN marks

A network can be considered as a set J of resources. Users are connected to the network. Any communication via the network is carried out by using capacity in a subset of the resources along a route r such that information (bits) is transported from one user to another user. Let R denote the set of all possible routes. Let x_r denote the rate (i.e. bits/s or packets/s) along route r. Consider a resource j in the network. The traffic load of this resource is:

$$y_j = \Sigma_{s\,:\,j \in s}\, x_s\,(t)$$

As a resource in the network becomes more heavily loaded, an increasing cost is incurred. The differentiable function $C_j(y)$ is the rate at which total costs is incurred at resource j at load y. We assume that $C' > 0$ and $C'' > 0$. In Kelly 1999 it is assumed that a unit flow through resource j is priced at marginal cost:

(1.) $$\frac{d}{dy}C_j(y) = p_j(y)$$

This is the pricing function, i.e. the function is determining the unit price of traversing resource j at load y. It is accordingly postulated that price equal marginal cost. Since the total load of the resource is the sum of rates for all routes traversing the particular resource the unit price of traversing the resource μ_j is:

(2.) $$\mu_j(t) = p_j\left(\sum_{s:j\in s} x_s(t)\right)$$

This pricing is implemented by ECN marks. Thus the rate at which packets are marked in resource j; ρ_j is chosen such that $\mu_j(t) \equiv \rho_j(t)\cdot\pi$, where π is the price per marked packet being charged form the customer. Without loss of generality we will set the price per mark to unity such that marking rate and price are identical.

2.2 User behaviour

A user derives utility from the rate at which she communicates over route *r*, i.e. $U_r(x_r)$. The utility function is assumed to be an increasing strictly concave and continuously differentiable function with $\lim_{x_r \to 0} U' = \infty$ and $\lim_{x_r \to \infty} U' = 0$. The user is maximising utility, thus she solves $\max U_r(x_r) - w_r$. The first order condition for the problem of the user is:

$$(3.) \qquad w_r(t) = x_r(t) U_r'(x_r(t))$$

Where *w* is the "weight" or willingness to pay for the user (will be described more closely later in the paper). Finally the user is assumed to adjust her rate at any point in time according to the following differential equation:

$$(4.) \qquad \frac{d}{dt} x_r(t) = \kappa_r \left(w_r(t) - x_r(t) \sum_{j \in r} \mu_j(t) \right)$$

We are now in a position to formulate one of the results as presented in Kelly, 1999:

Kelly's theorem 2.2:

The strictly concave function:

$$(5.) \qquad W(x) = \sum_{r \in R} U_r(x_r) - \sum_{j \in J} C_j \left(\sum_{s: j \in s} x_s \right)$$

is a Lyapunov function for the system(1.), (2.), (3.) and (4.). The unique value maximising $W(x)$ is accordingly a stable point of the system.

2.3 A reinterpretation

In line with Kelly's notation, let λ_r denote the unit price of traversing route *r* defined by:

$$(6.) \qquad \lambda_r \equiv \frac{w_r}{x_r}$$

Then we can write the optimisation problem of the user as:

$$(7.) \qquad \max\left(U_r(x_r) - \lambda_r x_r\right)$$

Reformulated in this way we have the standard consumer optimisation problem. The consumer is buying a number of units *x* of a good at unit price λ. The problem is to determine the number of units such that the difference between utility and the total outlay for buying the good is maximised. In our context the "good" is a unit flow through a network (e.g. measured in bits/s). The first order condition is $U_r'(x_r) = \lambda$, which is equivalent to equation (3.) $w_r = x_r U_r'(x_r)$ above. This first order condition is implicitly defining a demand curve, since for any price λ there exist an optimal flow *x*. This demand curve is furthermore downward sloping since *U* is strictly concave. A demand curve is illustrated below:

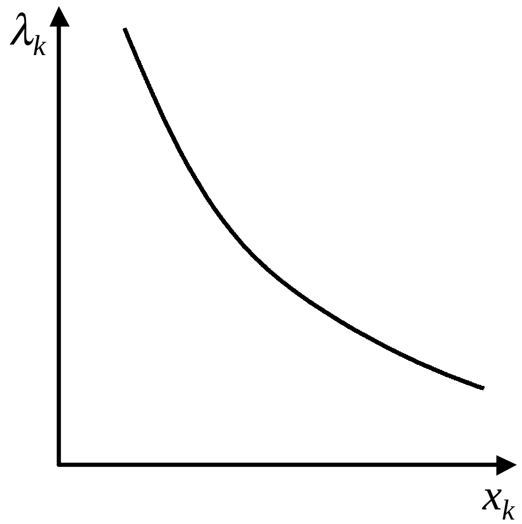

Figure 1, the demand curve

The rate given by the demand function is the optimal rate for the user at a given unit price and not necessarily the actual rate since prices fluctuate and there is a lag in observing the actual price at a point in time due to round trip times etc. In order to stress that it is the optimal rate we write it *x**. The demand function is:

$$(8.) \qquad x^*_r(\lambda_r) = \arg\max_{x_r}\left(U_r(x_r) - \lambda_r x_r\right)$$

We have defined λ_r as the unit price of traversing route *r*. This unit price is given, from the network, as the number of feedback signals (ECN marks) along route *r* times the price attached to these feedback signals (combining (2.) and the leftmost sum in equation (4.):

(9.) $$\lambda_r = \sum_{j \in r} \mu_j(t) = \sum_{j \in r} \left[p_j \left(\sum_{s: j \in s} x_s(t) \right) \right]$$

Then we can replace *w* in Kellys system since: $w_r = \lambda_r x^*{}_r$. The rate control differential equation (4.) can accordingly be written:

(10.) $$\frac{d}{dt} x_r(t) = \kappa_r \lambda_r \left(x_r^*(t) - x_r(t) \right)$$

We are now in position to reformulate Kelly's theorem:

Proposition 1

The function

(11.) $$W(x) = -\sum_{r \in R} U_r(x_r) + \sum_{j \in J} C_j \left(\sum_{s: j \in s} x_s \right)$$

is a Liapunov function for the system consisting of (1.), (8.), (9.) and (10.).

The proof of proposition 1 is provided in the appendix.

The unique value minimising the Liapunov function is globally stable, furthermore as shown in the appendix this solution is Pareto efficient (maximising social welfare). This is hardly a surprise, since prices, by assumption equals marginal cost, users maximise utility and there is no market imperfections. This is a special case of the first theorem of Welfare economics (Se e.g. Varian, 1992, p. 326).

By now *w* is now longer part of the system. Instead we have included a new concept, the optimal rate. The user will in equilibrium generate exactly this flow. Outside equilibrium the user may generate a flow different form the optimal flow, and then the flow will be adjusted according to the rule (10.). By this little exercise we have demonstrated that we don't need to explicitly consider *w* when we model the market interaction and consider dynamic aspects of the system.[1] The interaction and communication between the network and the user is captured in the two variables, price and flow. For a given flow the network determine a price and for a given price the user determine a flow.

[1] The weight *w* may however be of interest if one is focusing on fairness and related topics.

2.4 The supply side

Until now we have assumed that prices equal marginal cost without considering what market conditions that may result in such a pricing rule. One possibility is of course that the network is owned and managed by a benevolent social planner. Alternatively we can assume perfect competition.

Perfect competition in this context implies that there are sufficiently many suppliers providing services on any route such that the price charged by the other suppliers are unaffected by decisions made by a single supplier. Furthermore, if the supplier set his price (marking rate) above the market price, no routes will traverse his resource since users can buy connections cheaper elsewhere. If the supplier tries the opposite, to determine his price below the market price he will obviously experience reduced revenues since he can increase revenues by increasing his price. Thus the suppliers are price takers.[2] A set of assumptions that yields the pricing rule in equation (1.) is then the following:

Assume suppliers are price takers. Assume furthermore that pricing in a single resource only can be contingent upon the state of that particular resource and not depend directly upon the state of the neighbouring resources. Thus we assume that the pricing rule has to be decentralised.[3]

Profit maximising on a particular resource is then implying that the supplier finds the optimal aggregated rate y solving $\max_y[\mu y - C(y)]$ where μ is the market marking rate and $C()$ is the convex cost function. The supplier is accordingly observing the market price μ and on this basis maximise the difference between revenues and cost by finding the optimal y. The unique solution of this problem is:

$$(12.) \qquad \frac{d}{dy}C_j(y) = \mu_j$$

---

[2] In the next section of the paper we will argue that this assumption hardly can be fulfilled in a market with ECN pricing. It is possible to argue that the assumption is inconsistent with the idea of pricing by ECN marks.

[3] It may be the case that we can relax this assumption, without changing the result. The assumption is however simplifying the exposition and it is not unreasonable since the time scale of changing prices are a critical factor in obtaining a stable equilibrium.

This first order condition gives a functional relationship between marking rate (price) and the rate that is passing the resource. Notice that this relationship is identical to the pricing rule postulated in (1.). The supply function along a particular route is then given by summing up the price at each resource along the route as carried out in (9.):

$$\lambda_r = \sum_{j \in r} \mu_j(t) = \sum_{j \in r} \left[p_j \left(\sum_{s:j \in s} x_s(t) \right) \right] \qquad \text{where} \qquad p_j(y) = \frac{d}{dy} C(y)$$

This is the market supply function on a particular route. We have already argued that (8.) is a demand function. We can accordingly illustrate the market equilibrium as the intersection of the downward sloping demand function and the upward sloping supply function:

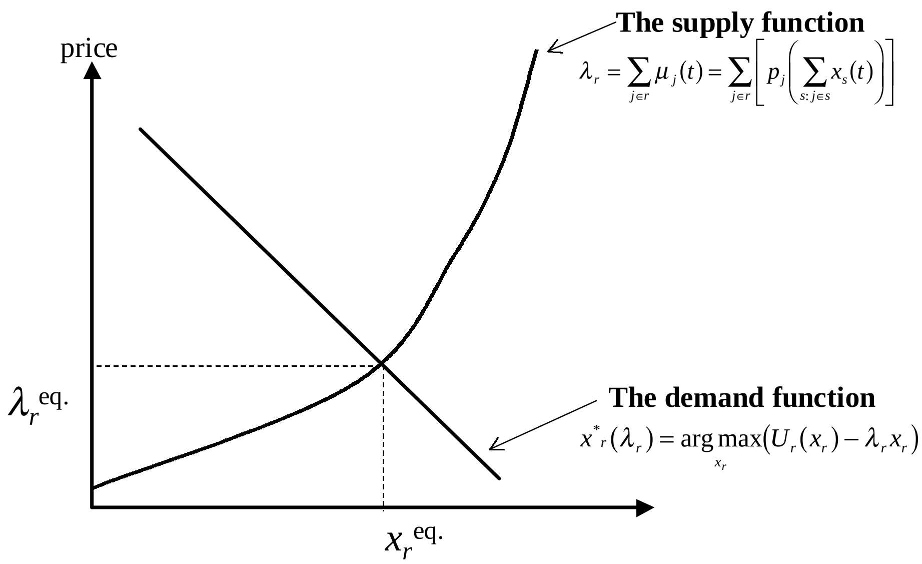

Figure 2, Market equilibrium

The system is in equilibrium in the intersection point between these two curves. Notice that a resource in general is used on more than one route, thus there will be “across route effects” in the sense that when the load on one route increase (decrease), the supply function for all related routes will shift upwards (downwards).

As already pointed out, the assumed price taker assumption on the supply side can hardly be expected to be fulfilled in the market place. First of all, in the presence of economies of scale, there will be a finite number of suppliers. Secondly, the price taker assumption may be problematic. On a very short timescale, taken literally, the assumption of suppliers being price takers is clearly violated. Suppliers can change pricing schedules in the middle of a session. The market response will be delayed due to the round trip time in the network. In a repeated game between supplier and user we may however find that the equilibrium is to never change the pricing schedule in the middle of sessions because users “punish” this type of behaviour in such a way that it not is worthwhile for the supplier.

It can also be argued that the price taker assumption hardly can be fulfilled on a longer time scale. The price taker assumption implies that the supplier is observing the market price and based on this price determine the rate that he would like to pass trough the resource. We can see this directly from the optimisation problem we solved above where the aggregated rate and not price is the decision variable; $\max_y\left[\mu y - C(y)\right]$. This is opposed to how pricing by ECN marks are described in the literature. E.g. in Kelly (2000), p. 5 we read: “Suppose that as a resource become more heavily loaded it generates feedback signals intended to indicate congestion ()”. Thus it seems like the decision variable is (by construction) the feedback signal (price) and not the aggregated rate traversing the resource. If the suppliers are deploying a three- stage procedure, they may however act as if they are price takers. At the first stage suppliers observe the market price, at the second stage, optimal aggregated rate given this price is calculated and then finally at the third stage the suppliers determine a pricing schedule consistent with the optimal aggregated rate.

As indicated in the paragraphs above it is possible to construct cases where price taker behaviour is possible. A necessary condition in any case is however that it is very many suppliers (infinitely many) on any route. This requirement can hardly be meet in the current Internet. Thus, we have to relax the assumption of perfect competition.

The straight- forward way of relaxing the price taker assumption is to introduce imperfect competition where suppliers indeed set their own price. Before looking into imperfect competition we will however take a brief look at dynamic properties of the model because dynamics under imperfect competition will have similarities to dynamics under perfect competition since the demand side will remain unchanged.

3 Dynamics and stability

We have already proved that the system is globally stable (see the appendix). It may however be of interest to discuss the dynamic properties of the system. The major difference between the market considered here and a standard economics textbook market is that the user will set his actual rate $x(t)$ at a point in time and not observe the actual price that he will pay for the flow until a later point in time $t + T$.[4] Since prices fluctuate dynamically in the network, the user will, on a very short time scale, not be able to adjust his rate optimally. There is a lag. Then the rate $x(t)$ can be said to be determined on the basis of the *expected* prices at time t. A possible interpretation of the rate control differential equation is accordingly that it implicitly describes how the user is updating his price expectations.

As long as the Liuapunov function is strictly convex the system is stable. The curvature of this function is determined by the slope of the demand and supply functions. In the standard case with a downward sloping demand function and an upward sloping supply function the convexity assumption is fulfilled. Furthermore, as long as $\frac{\partial}{\partial\lambda}\left[D(\lambda)-S(\lambda)\right]<0$ where D and S is the demand and supply function respectively we will typically have a stable system (see e.g. Varian 1992 p. 399). Thus there can exist stable solutions where both the demand and supply curves are downward sloping and similarly we can have a stable system where the curves slopes upwards. A necessary condition is however that the demand function crosses the supply function from “below” such that there is excess demand for low prices and excess supply for high prices.

These stability properties may however be compromised if we take explicitly into account the time lag between changes in price signals and the changes in the sending of ECMs. If we let Δt denote the lag, then one can write equation (10) as:

$$\dot{x}_r(t) = \kappa_r \lambda_r (t - \Delta t)\left(x_r^*(t - \Delta t) - x(t)\right)$$

The stability properties of this system are not obvious. Work on similar equations from the natural resource literature (see e.g. Wangersky and Cunningham, 1957) indicates that such systems are notoriously cyclical. One would expect that for small values of Δt, the cycles would be of marginal importance. However this would

---

[4] This delay is due to round trip times

depend on a number of factors. For instance, it may well be the case that the size of κ plays a role in promoting the cyclical nature of the solution. The larger the value of κ, the faster may *x* zoom by it's equilibrium value. This is an interesting contrast with the non-lagged case, where the larger the size of κ, the faster will the system approach its equilibrium. The properties of such lagged differential equations within the context of internet pricing is very much a subject for further research.

Modeling rate control in the Internet as a continuous differential process is an appropriate approximation in resources where traffic is aggregated over many users. The rate control carried out by a single user is however more like a discrete process. Taking this fact into consideration as well as the time lag considered above we can reformulate the rate control differential equation into a fairly simple discrete dynamic equation:

$$x_r(t+1) - x_r(t) = \kappa_r \lambda_r(t)\left[x_r^*(\lambda(t)) - x_r(t)\right]$$

With such a rate control equation the system can be stable, cyclical or unstable depending upon the slope of the demand function, the supply function and the value of the parameter κ. Below we will provide numerical illustrations where we for simplicity assume linear supply and demand functions. We will consider the following simplified system:

Demand function:	$x^* = A - b\lambda$
Supply function:	$\lambda = mc \cdot x$
Rate control:	$x_r(t+1) - x_r(t) = \kappa_r \lambda_r(t)\left[x_r^*(\lambda(t)) - x_r(t)\right]$

Below we have illustrated this system for parameter values: A = 5, b = 1, mc = 0.7, κ =0.5 and with an initially expected price= 4.5

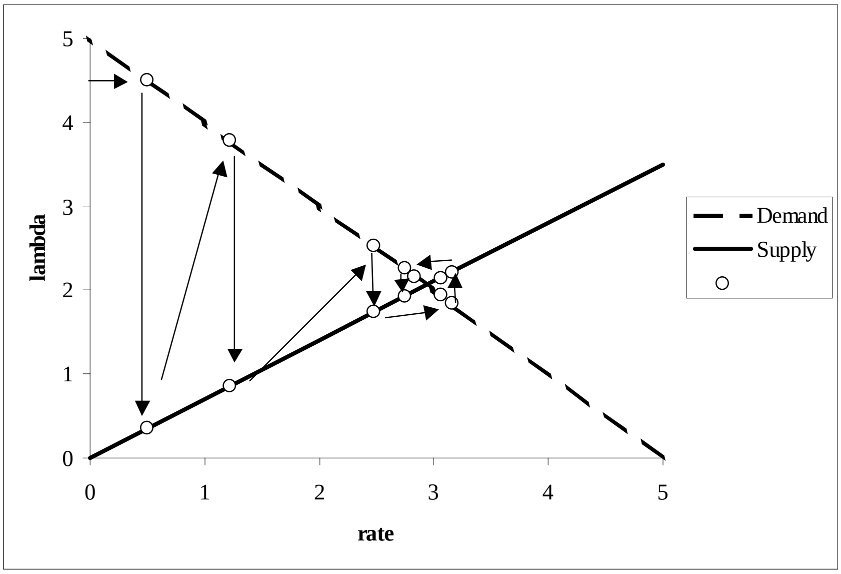

Figure 3 a stable system

In figure 3, the user is assumed to initially expect the price to be 4.5 (based on earlier experience etc). With this expected price it is optimal for the user to send at a rate of 0.5. Since this is a very low rate the network responds by determining a price of 0.35. Given this low price the user employ his rate control function to update his price expectation to 3.77 and then it is optimal to adjust the sending rate to 1.22 this goes on and the system is converging to the equilibrium value of price = 2.05 and rate = 2.94.

By small parameter changes, the behaviour of this system change radically, for instance if the parameter κ is increased from 0.5 to 0.7 and we keep other parameters at the same values as in figure 3 (A = 5, b = 1, mc = 0.7) the system becomes cyclical:

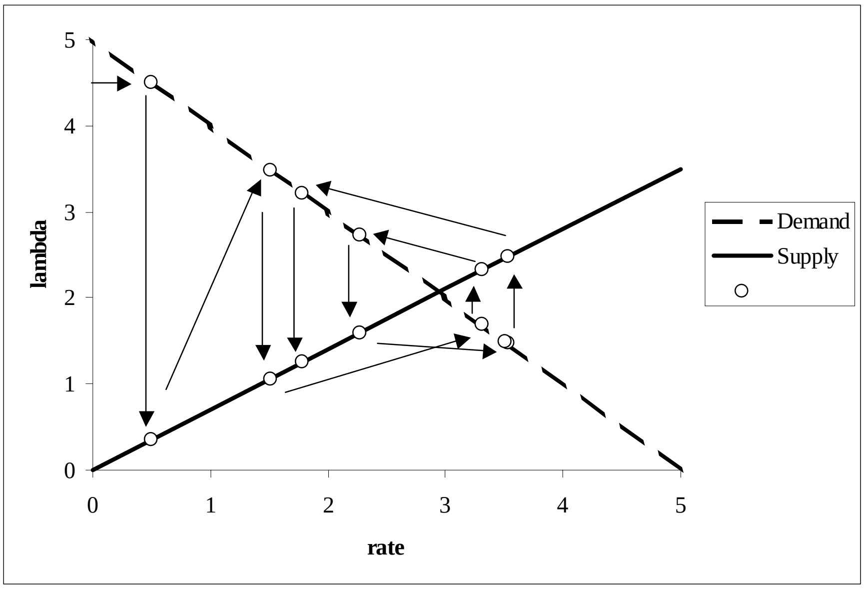

Figure 4, a cyclical version

If we increase the parameter κ even more, to e.g. 0.9 the system explodes as illustrated below (for the same parameter values as earlier):

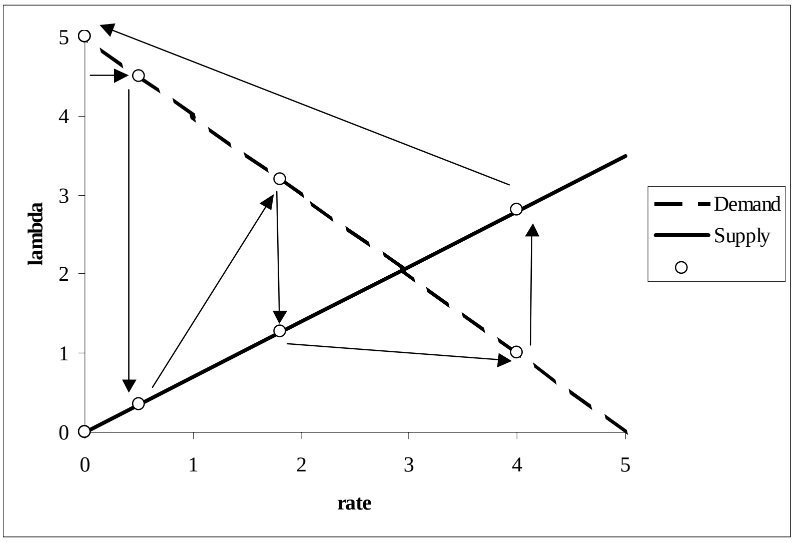

Figure 5 an exploding system

Notice however that the simplified linear demand and supply function used in these illustrations are only approximations to the demand and supply functions resulting from the assumed utility and cost function used in the previous sections. This approximation is in particular leading to errors when values approach zero.

Below we illustrate the system for different slopes of the demand function and obtain a stable, cyclical and an exploding system respectively:[5]

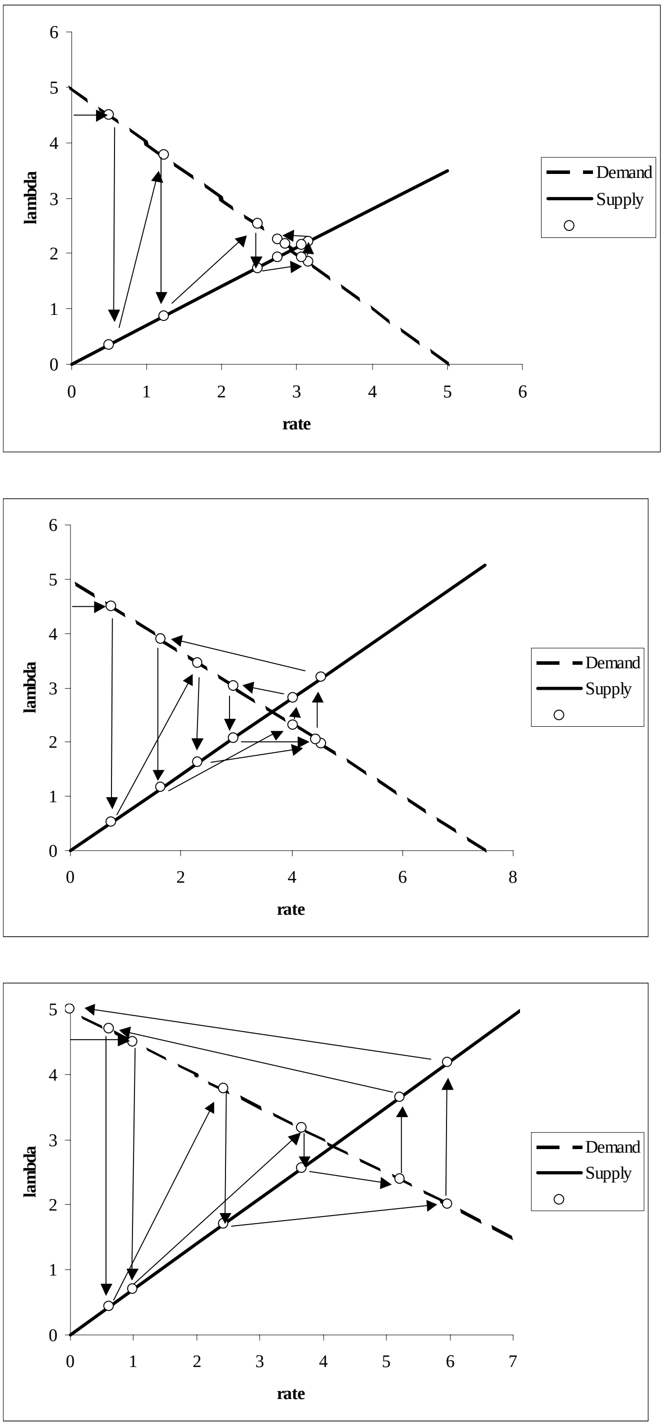

Figure 6 Systems with variations in the demand function

---

[5] In figure 6 we have used, parameter values of mc = 0.7, κ = 0.5, and initial expected price = 4.5. The three different dynamic patterns are obtained by variations in the demand function, in the left most figure the we use A = 5 and b = 1, in the centre figure we use A = 7.5 and b = 1.5 and finally to the right we use A = 10, b = 2.

All of the demand functions illustrated in figure 6 are plausible but they result in radically different dynamic patterns.

We obtain similar results by playing around with the slope of the supply function:[6]

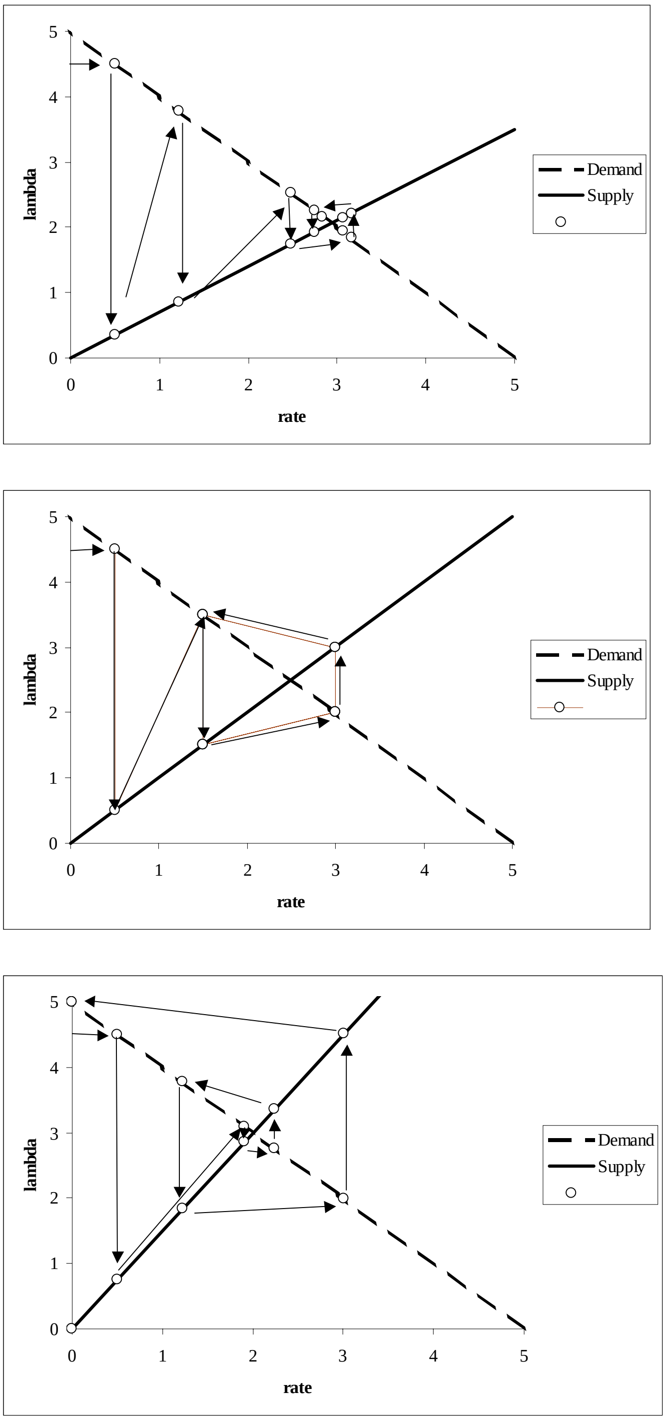

Figure 7, Systems with variations in the supply function

As already stated, the linear functions used here are only approximations to the functions resulting from the assumed utility and cost function postulated in earlier sections of this paper. Thus the behaviour of the system may be quite different when it is far off the equilibrium values. Systems like the one to the right of figure 7 may then not explode but they will be extremely cyclical.

[6] In figure 7 we have used parameter values A = 5, b = 1, κ = 0.5, and initial expected price = 4.5. The three different dynamic patterns are obtained by variations in the supply function, in the left most figure we use mc = 0.7, in the centre we use of mc = 1, and to the right we use mc = 1.5.

4 Monopoly

As argued above, the price taker assumption may be inconsistent with the principles of pricing by ECN marks. We will in this section consider a monopoly on a particular route and let the network, or the supplier determine the pricing policy in order to maximise profits. In this section we will not take into account the possible dynamic problems discussed in the previous section. We will assume the continuous version of the rate control function. Thus, system dynamics are well behaved and we can focus on monopoly.

Under monopoly the supplier is maximising the difference between revenues and costs on the route under the constraint that demand is a downward sloping function of the price. The optimisation problem is accordingly:

$$\pi_r = \max_{\lambda_r}\left[\lambda_r x_r - \sum_{j\in r} C_j\left(\sum_{r:j\in r} x_r\right)\right] \qquad \text{s.t.} \qquad x_r = x_r(\lambda_r)$$

In this problem the price is the decision variable and not quantity. By inserting the demand function and differentiating with respect to price we obtain the following first order condition (with our assumptions with respect to utility and cost functions second order conditions are fulfilled):

$$x_r(\lambda_r) + \lambda_r \tfrac{\partial x_r}{\partial \lambda_r} - \sum_{j\in r}\frac{\partial C_j(\cdot)}{\partial x_r}\tfrac{\partial x_r}{\partial \lambda_r} = 0 \Leftrightarrow \sum_{j\in r}\frac{\partial C_j(\cdot)}{\partial x_r} = \lambda_r + x_r(\lambda_r)\frac{1}{\frac{\partial x_r}{\partial \lambda_r}}$$

This is a standard monopoly solution. The optimal solution is to set marginal cost equal to marginal revenue. Marginal cost consist of the sum of marginal costs along the path and marginal revenue is the revenue from the marginal unit plus the negative effect of receiving a lower price for all the infra marginal units. The monopoly solution is illustrated below:

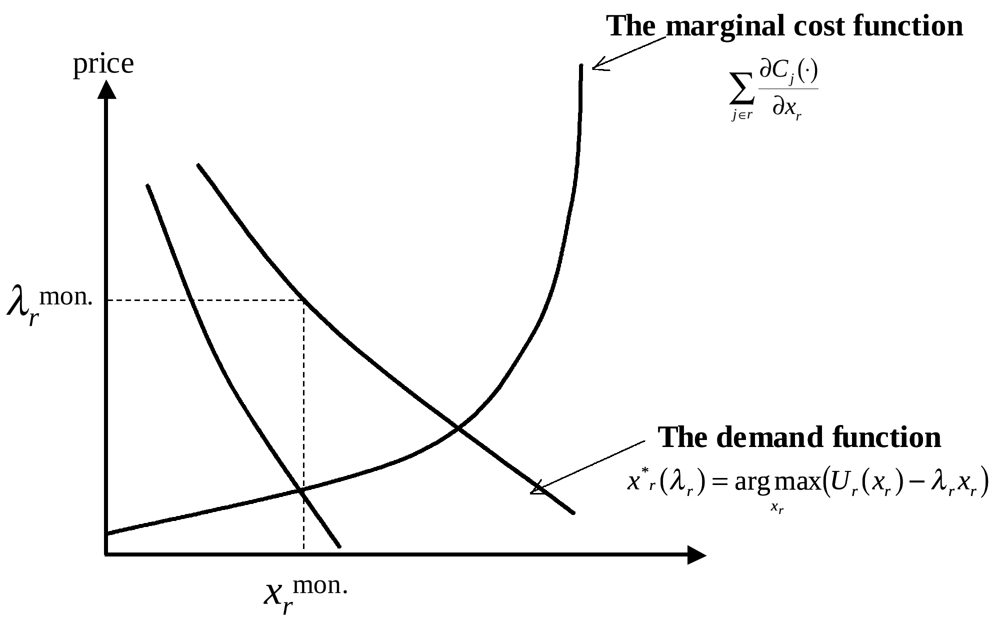

Figure 8 Monopoly

This is a standard textbook result. The optimal solution for a monopoly is where marginal revenue equals marginal cost. Notice in particular that the price is higher than the marginal cost. Thus this solution is not Pareto optimal.

5 Duopoly

Consider a market with two competing network providers that both deploy dynamic pricing in the network. For simplicity we will assume that the service provided by these two suppliers are perfect substitutes in the eyes of the consumers. We will assume that there always is users comparing the effective price charged by the two network and route their traffic to the network with the lower price such that an equilibrium is characterised by a distribution of traffic such that the price in the two networks are identical.

We are basically searching for equilibrium in price schedules. As a first attempt to describe such a game we will consider an extremely simplified model. We will assume that market demand is given by the linear function $x = D - d\lambda$ (where the parameters $D > 0$ and $d > 0$). We will assume that the two firms (firm *A* and *B*) set their price functions sequentially. At the first stage firm *B* determine its price function and in stage 2 firm *A* set its price function. For simplicity we will assume that both firms are restricted in determining pricing rules. The price of traversing a particular network is proportional to total load in that network.[7] Thus firm *i* has pricing rule: $\lambda_i = \gamma_i x_i$. The decision variable for the firms is accordingly to determine the slope of the pricing function; γ. Finally we will assume that the two firms have identical cost structures and marginal costs are assumed constant and is normalised to zero. The game will be solved by backward induction.

Assume firm *B* is the first mover. Firm *A* will then at stage 2 maximise profits subject to the demand function constraint and the constraint coming from the pricing decision that already has been made by firm *B*. Let $\lambda_B = bx_B$ be the price function determined by firm *B* at stage 1. We will analyse the decision at stage 2 of the game by assuming that firm *A* choose one single optimal equilibrium price. At the end of this section we will demonstrate that this equilibrium price can be implemented by a permissible pricing function. At stage 2, firm *A* is accordingly maximising profits subject to the residual demand function he is facing. At any price he sets λ, firm *B*

---

[7] Both the assumption of a sequential game and the restriction on the pricing rule is chosen in order to simplify the game. It is evidently a subject for further studies to consider a simultaneous move game where more general pricing functions are allowed.

will capture a fraction of total demand: $x_B = \frac{\lambda}{b}$ thus firm *A* is facing the residual demand function:

$$x_A(\lambda) = D(\lambda) - x_B(\lambda) = D - d\lambda - \frac{\lambda}{b}$$

Supplier *A* is solving the following optimisation problem at stage 2:

$$\pi_A = \max_{\lambda}\left[x_A(\lambda)\cdot\lambda\right] = \max_{\lambda}\left[\lambda\left(D - \lambda\frac{db+1}{b}\right)\right]$$

$$FoC: \quad D - 2\lambda\frac{db+1}{b} = 0 \Leftrightarrow \lambda = \frac{bD}{2(db+1)}$$

At stage 1 supplier *B* knows what supplier *A* is going to do at stage 2, supplier 2 will thus maximise profits by determining an optimal price function $\lambda_B = bx_B$. For a given *b* Supplier *B* knows that the number of units he will sell in the market place is:

$$\lambda = bx_b \Leftrightarrow x_B = \frac{\lambda}{b} = \frac{D}{2(db+1)}$$

He will then maximise profits by determining optimal slope for his price function: *b:*

$$\pi_B = \max_{b}\left[x_B\lambda\right] = \max_{b}\left[\frac{bD^2}{4(db+1)^2}\right]$$

$$FoC: \quad 0 = \frac{1-db}{4(db+1)^3}$$

The denominator will never be zero, thus making the numerator equal to zero is the only candidate optimal slopes for the pricing function, thus:

$$b = \frac{1}{d}$$

By inserting optimal values we can now describe equilibrium in the model:

$$b^* = \frac{1}{d}, \quad \lambda^* = \frac{D}{4d}, \quad x^*{}_B = \frac{D}{4}, \quad x^*{}_A = \frac{D}{2}$$

Finally we can demonstrate our claim that firm *A* (the second mover) can indeed implement the solution above by determining a constant slope α for his price function satisfying:

$$\lambda^* = \alpha x_A^* \Leftrightarrow \alpha = \frac{\lambda^*}{x_A^*} = \frac{1}{2d}$$

Below we provide a numerical illustration of the equilibrium. Parameter values are D = 5, and d = 1:

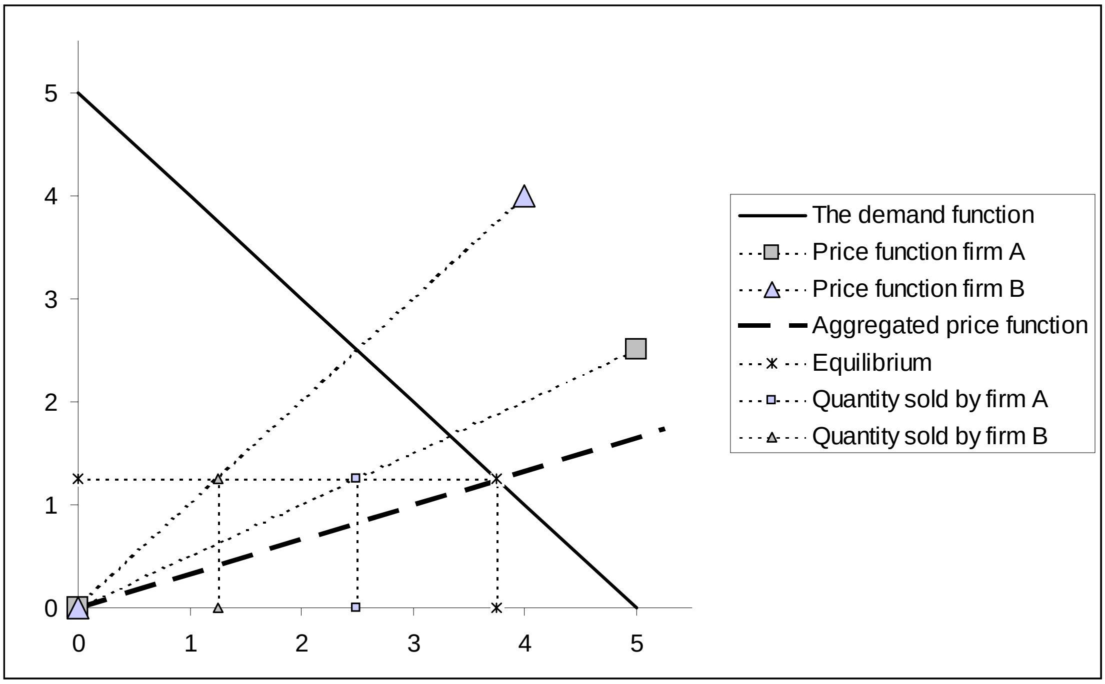

Figure 9, the duopoly solution in a sequential move game

The equilibrium in this simple sequential duopoly model is accordingly characterised by an increasing aggregated price function. If, for some reason, the market is outside equilibrium, the rate control dynamics outlined under perfect competition will be deployed and thus this market will have similar dynamic properties as the perfect competition model.

It is evident from the equilibrium solution that the second mover sells a higher quantity in equilibrium. Since the prices charged by the firms are identical and marginal costs are identical (and normalised to zero), the second mover derives twice as high profits as the first mover. Thus if timing of announcing price rules is endogenous both firms would seek to postpone this decision as long as possible. This is sort of dynamic game is outside the scope of this paper.

6 Inelastic traffic

Until now we have made quite restrictive assumptions with respect to the utility function. We have assumed that utility is a continuous, increasing strictly concave function of the rate. Following Schenker (1995) such traffic is called elastic traffic. The shape of a utility function where the only argument is rate will evidently depend upon the application that the user(s) is running. File transfers may be of an elastic kind whereas the utility function for a user running a real time voice application will typically have a radically different shape. If the rate is below a given threshold the utility is near zero since the quality of the transmitted voice is extremely low. If the rate is increased above the threshold utility will not be further increased since the application not can take advantage of the increased band-with.

Consider a user with a fixed-rate application (in Schenker's terminology a hard real time application). The utility function is given by:

$$U(x) = \begin{cases} 0 & \text{for} \quad x < \tilde{x} \\ \upsilon & \text{for} \quad x \geq \tilde{x} \end{cases}$$

The user is maximising utility minus outlay. The demand function is accordingly:

$$x^*(\lambda) = \begin{cases} \tilde{x} & \text{for} \quad \lambda\tilde{x} \leq \upsilon \\ 0 & \text{for} \quad \lambda\tilde{x} > \upsilon \end{cases}$$

Thus the demand function is like:

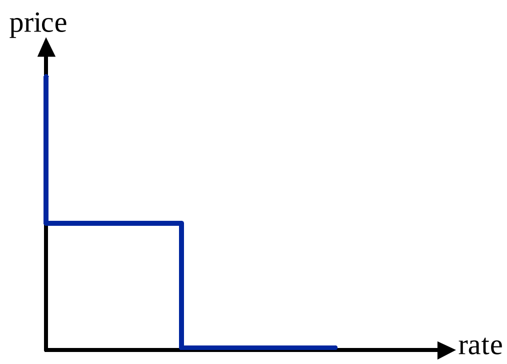

Figure 10, Demand from a user with a hard real time application

The rate generated by this user has a jump at threshold υ, and thus the horizontal line segment is indicating a discontinuity in demand since $\lim_{p \to \upsilon^-} D(p) = x_0$, $\lim_{p \to \upsilon^+} D(p) = 0$. In a network with inelastic users there may accordingly be problems with existence of equilibrium if the supply function crosses the demand function at a horizontal segment. We can however add realism to the model by relaxing the assumption of identical users. The threshold υ will typically be a function of disposable income for the user, type of real time application and in what stage of a real time session the user is (the threshold will probably be different in the beginning as compared to the end of e.g. a movie). By aggregating demand functions over many heterogeneous inelastic users we obtain a function that is almost nice and downward sloping:

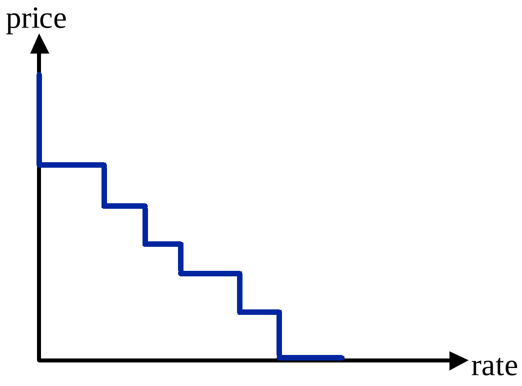

Figure 11, demand aggregated over many users with hard real time applications

If the supply function crosses at a vertical segment, equilibrium will exist whereas if it crosses at a horizontal segment the problem of non-existence persists. According to Schenker (1995) real time applications like audio and video are being developed in the direction of becoming delay or rate adaptive. In such cases, utility as a function of rate is *S* shaped as illustrated below:

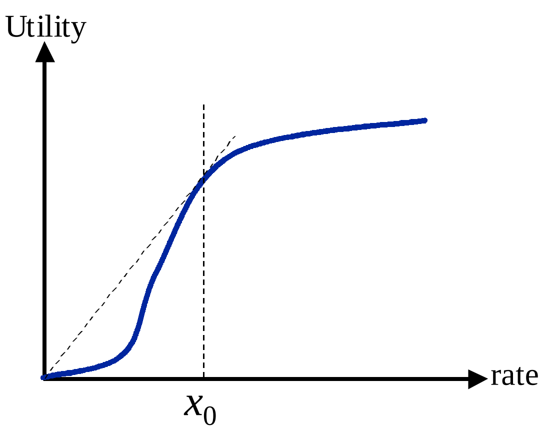

Figure 12, utility for a rate adaptive application

For a sufficiently high price such a user will not generate traffic. At a given threshold price, illustrated by the dotted tangent in the figure above, demand generated by the user will have a discontinuous jump to x_0. As the price is further decreased, the optimal rate will move along the concave segment of the utility function. Thus we obtain a demand function like the one illustrated below:

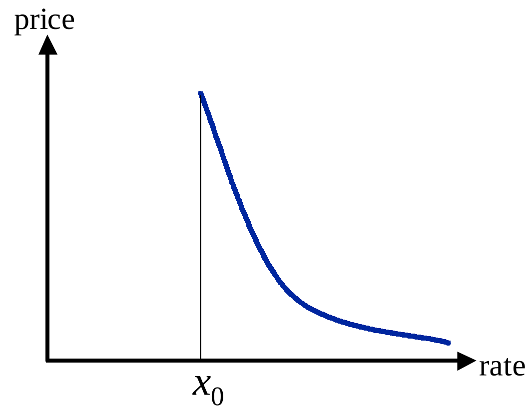

Figure 13, demand from a user with a rate adaptive application

Aggregation over many heterogeneous consumers, some running elastic, some inelastic and some rate (delay) adaptive applications yields a downward sloping demand function with some horizontal segments:

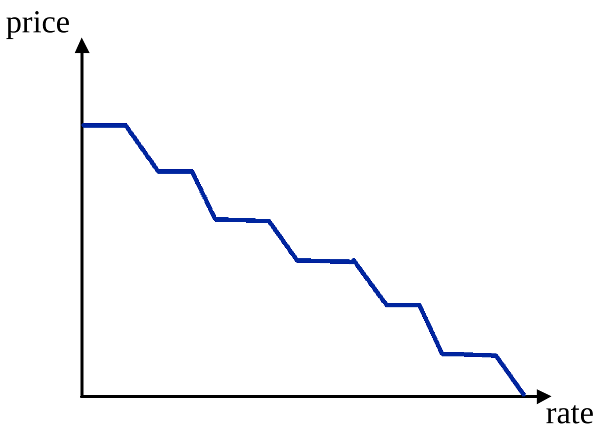

Figure 14, aggregated demand

The discontinuities on the demand side in the model presented here is very similar to the discontinuities we have in any economy where consumers have unit demand (as is the case for most goods sold in discrete units like cars, computers or books). Furthermore it is similar to the supply side in an industry where average cost curves are U shaped. As stated by Varian (1992) p. 394: *“the discontinuities will be irrelevant if the scale of the market is sufficiently large”*. Thus when the (marginal) discontinuities are small compared to the total load in each resource, there will generally be a price vector that results in demand being close to supply.

7 Externalities

In the models discussed in this paper we deploy a convex cost functions. According to Gibbens and Kelly (1999), p. 5 the derivative of the cost function is equal to shadow price at the resource. This cost stems from the (potential) loss of packets as resources become overloaded. Packet loss is partly a cost that is carried by the network since packet has to be resent. There will be an opportunity cost associated with the required capacity required to resend packets. Packet loss, will however also be a cost to users because it typically results in degradation of service due to delays jitter etc. This is a classic externality. Network owners will in general not fully take into account this externality. We will in the following make a minor change to the model by including a negative externality in the utility function and demonstrate that resources in equilibrium will be over congested. This will be done in a straight forward textbook way. Let total utility for user *i* be:

$$U_i(x_i, X) - x_i\lambda \quad \text{where} \quad X = \sum_j x_j$$

$$\frac{\partial U_i}{\partial x_i} > 0, \frac{\partial^2 U_i}{\partial {x^2}_i} < 0 \text{ etc,} \quad \frac{\partial U_i}{\partial X} < 0$$

This is gross utility minus outlay for network usage in the same way as in the previous sections. In the utility function a second term is however included. This is the inconvenience due to packet loss. Packet loss is a function of total traffic on the network. As packet loss increase, utility decrease. User optimisation yields:

$$x^*_i(\lambda) = \arg\max_{x_i}\left(U_i(x_i, X) - \lambda x_i\right), \quad \text{FoC:} \qquad \frac{\partial U_i}{\partial x_i} + \frac{\partial U_i}{\partial X} = \lambda$$

For simplicity we will assume that the network consist of one single resource and the cost taken into consideration is only the packet loss costs carried by the network. The pricing rule is thus price equal to marginal cost;[8] $\lambda = C'(X)$. Market equilibrium is accordingly similar to what we have studied in earlier sections of this paper, marginal individual willingness to pay equals marginal cost in optimum. The socially optimal solution has however changed, maximising welfare yields:

---

[8] We are thus, for simplicity, assuming that suppliers are price takers.

$$W = \sum_i U_i\left(x_i, \sum_j x_j\right) - C\left(\sum_j x_j\right)$$

Foc: $$\frac{\partial U_i}{\partial x_i} + \frac{\partial U_i}{\partial X} + \sum_{j \neq i} \frac{\partial U_j}{\partial X} = C'$$

The socially optimal solution is to equate marginal cost in the network with the total marginal social willingness to pay. The first two terms in this willingness to pay is identical to the individual willingness to pay considered above. The third term captures the negative externality. The marginal user imposes a negative externality on all the other users by increasing their waiting time, jitter etc. Thus social marginal willingness to pay is below individual willingness to pay. The solutions are illustrated below:

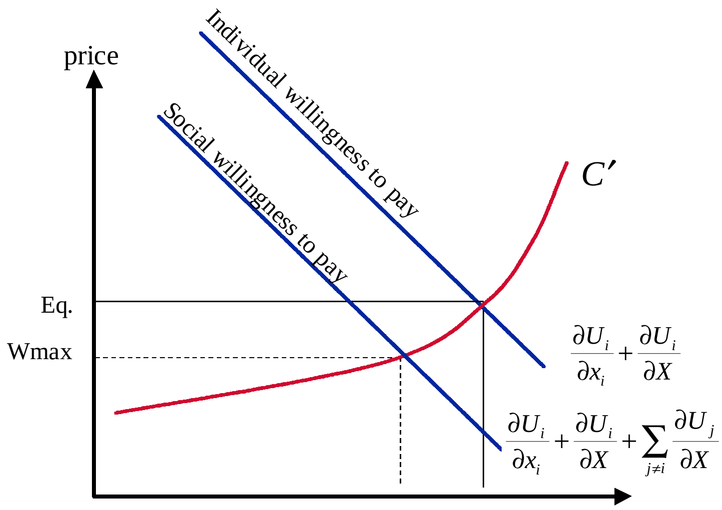

Figure 15, A market with congestion externalities

Market equilibrium is indicated by *eq* in the figure and the socially optimal solutions is indicated by *Wmax*. The market solution results in an over utilisation of network resources. This is also a standard textbook result, (see e.g. Varian, 1992, pp 432 – 439) The welfare loss can be corrected. A welfare maximising benevolent monopoly can increase the price of traversing the network or in a competition scenari or a regulator can introduce a tax on usage of network resources. In any case the price paid by consumers will be equal to: $\lambda = C'(X) + t$. Where the tax, t, is identical to the value of the adverse effect on all other users, i.e.:

$$t = -\sum_{j \neq i} \frac{\partial U_j}{\partial X}$$

Then the individual first order conditions yields a solution identical to the first best solution:

$$x^*_i(\lambda) = \arg\max_{x_i}\left(U_i(x_i, X) - (c'+t)x_i\right), \quad \text{FoC:} \quad \frac{\partial U_i}{\partial x_i} + \frac{\partial U_i}{\partial X} = c' + t$$

This solution is illustrated below:

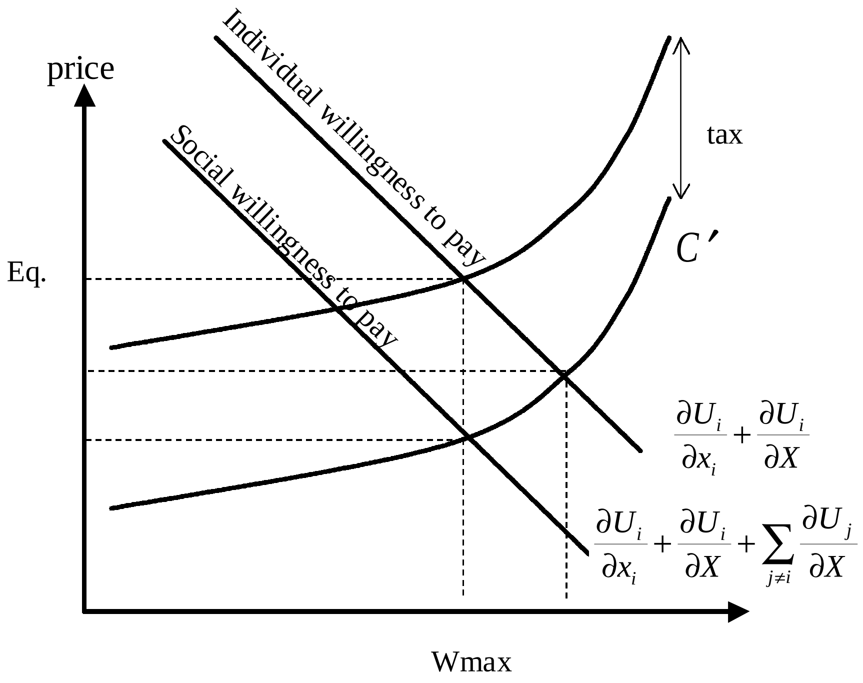

Figure 16 Externalities and a tax

In figure 16 we have illustrated the case where a tax added to the price (that is equal to marginal cost) yields the optimal solution. Each single user is not directly taking into account the negative externality, but since the tax, by construction, exactly equals the negative externality, the user act as if they considered the negative externality.

8 Conclusions

In this paper we have demonstrated that the ECN pricing scheme can be reformulated such that we are in a position to take advantage of standard results from the economics literature.

We have found, on the one hand, that the demand side as outlined in e.g. Kelly 1999 fits nicely with standard economic theory. On the other hand the necessary assumptions required to make the ECN set-up directly consistent with economic theory on the supply side is unlikely to be fulfilled in the market place. The presence of economies of scale will violate the assumption of (infinitely) many suppliers and the price taker assumption may thus be inconsistent with the design of ECN marking. In the paper we have relaxed the strict assumptions by considering imperfect competition. The equilibrium pricing rules will then however deviate from the pricing rule in the ECN literature. Equilibrium usage of network resources is then typically deviating from socially optimal usage.

We have furthermore discussed stability properties of the system. In the ideal model the system is globally stable, but as realism (?) is added to the model by considering discreteness and timelags, the dynamic properties of the model is changing and one may e.g. experience cyclical behaviour.

Furthermore, by taking advantage of standard aggregation results from the economics literature we have demonstrated that the presence of inelastic users resulting in a discontinuous demand function not necessarily will result in problems with existence and stability. It is basically a question of scaling. When (marginal) discontinuities are small compared to the total load in each resource, there will generally be a price vector that results in demand being very close to supply. Finally we have demonstrated that when the cost of congestion not exclusively is carried by the network, the presence of congestion results in a classic externality and network resources will accordingly be over-utilised in equilibrium.

Appendix A, proof of proposition 1

In order to investigate the properties of the system (1.), (8.), (9.) and (10.) we first examine whether equilibria exists and then investigate stability. The system is:

$$(*)\qquad \begin{array}{ll} (1.) & \frac{d}{dy}C_j(y) = p_j(y) \\ (10) & \frac{d}{dt}x_r(t) = \kappa_r\lambda_r(t)\left[x_r^*(t) - x_r(t)\right] \\ (9) & \lambda_r(t) = \sum_{j\in r}\left[p_j\left(\sum_{s:j\in s} x_s(t)\right)\right] \\ (8) & x^*{}_r(\lambda_r(t)) = \arg\max_{x_r}\left(U_r(x_r) - \lambda_r(t)x_r\right) \end{array}$$

From (8.) it follows that x_r must be defined by the equation $U_r' = \lambda_r$. It follows that if $x_r^*(t) = x_r(t)$ and $U_r' = \lambda_r$ holds for all *r*, then $\dot{p}_s$ and $\dot{\lambda}_r$ is zero. Note that in this equilibrium utility is maximised and (by assumption) prices are equal to marginal cost. Thus this equilibrium is optimal in the sense that social welfare is maximised. This observation turns out to be important in our discussion of stability. Let us refer to the equilibrium values of x_r as $\mathbf{\bar{x}}$.

In order to discuss stability it is beneficial to rewrite the system (*) into:

$$\frac{d}{dt}x_r(t) = \kappa_r\sum_{j\in r}\left[p_j\left(\sum_{s:j\in s} x_s(t)\right)\right]\left[x_r^*(t) - x_r(t)\right]$$

This system is autonomous and the stability theory is a bit less involved than if we worked with the original system. The function:

$$W(x) = -\sum_{r\in R} U_r(x_r) + \sum_{j\in J} C_j\left(\sum_{s:j\in s} x_s\right)$$

is a Liapunov function for the system (*) if the following conditions hold, Varian (1992):

$\mathbf{\bar{x}}$ minimise W(*x*).

$\dot{W}(x) = \mathbf{D}W(x)\dot{x} < 0 \quad \forall\, x \neq \mathbf{\bar{x}}$

The first part is easily checked. If $\mathbf{\bar{x}}$ minimises W(*x*), then $\mathbf{\bar{x}}$ maximises -W(*x*). But if $\mathbf{\bar{x}}$ maximise -W(*x*), this implies that $\mathbf{\bar{x}}$ maximises social welfare, and this is exactly what we found would occur in equilibrium. It follows the assumptions made about the U_r and the cost functions that there is only one value of x that maximise social welfare. Thus the equilibrium $\bar{\mathbf{x}}$ is the same as the $\bar{\mathbf{x}}$ that minimise $W(\cdot)$.

In order to check the second part we calculate $\dot{W}(x)$. This expression is given by:

$$\dot{W}(x) = -\left(\sum_{r\in R} U_r'\dot{x}_r - \sum_{j\in J} C_j'\dot{x}_r\right) = -\left(\sum_{r\in R}\left(U_r' - \sum_{j\in J} C_j'\right)\dot{x}_r'\right)$$

If we examine the last expression we see that it consists of a sum with a term for every r. Each element in the sum is the marginal benefit of an additional unit of x_r multiplied by the differential equation for x_r. If x_r is below the social welfare maximising level of $\bar{x}_r$, then marginal benefit is positive and $\dot{x}_r'$ is positive. If x_r is above the social welfare maximising level of $\bar{x}_r$, then marginal benefit is negative and $\dot{x}_r'$ is negative. (This follows from the strict concavity of the Liapunov function.[9]) Either way the product is strictly positive. If x_r is exactly at $\bar{x}_r$, then the product of marginal benefit and $\dot{x}_r$ is zero. Thus $\dot{W}(x)$ is strictly negative for all x except at $x = \bar{\mathbf{x}}$ and the proof is completed.

[9] In fact it may hold for more general utility and cost functions. This needs to be checked.

References

Dolan, T. (2000), Internet Pricing – Is the end of the World Wide Wait in view?, Communications strategies, no 37, 1st quarter 2000, 15 - 46

Gibbens, RJ. And F.P Kelly (1999), Resource pricing and the evolution of congestion control, Automatica 35

Kelly, F. 1999, Mathematical modelling of the Internet, paper available at http://www.statslab.cam.ac.uk/~frank

Kelly, F. 2000, Models for a self managed Internet, Philosophical Transactions of the Royal Society A 358, 2335 - 2348

Kelly, F. AK, Maulloo and DKH Tan, 1998, Rate control in communication networks: shadow prices, proportional fairness, and stability, Journal of the operational research society, no 49, 237 - 252

Kelly, K. 1998, New Rules for the New Economy : 10 Radical Strategies for a Connected World, (the Penguin group, New York, New York)

MacKie-Mason, J. And H. Varian (1995), Pricing Congestible Network Resources, IEEE J. Selected areas in Commun. 13, 1141 – 1149

Shapiro, C. and H. Varian 1998, Information Rules: A Strategic Guide to the Network Economy, (Harvard Business School Press, Boston, Massachusetts)

Shenker, S. (1995), Fundamental design issues for the future internet, IEEE J. Selected areas in Commun. 13, 1176 – 1188

Varian, H.R. (1992), Microeconomic Analysis third edition, W.W. Norton Company, New York

Wangersky, P.J., and W. J. Cunningham (1957), Time lag in predator prey population models, Ecology, 38 (1957), 136 - 139.