Diskussionspapierreihe
Working Paper Series

HELMUT SCHMIDT
UNIVERSITÄT
Universität der Bundeswehr Hamburg

PRIORITY AND INTERNET QUALITY

JÖRN KRUSE

Nr./ No. 96
August 2009

Department of Economics
Fächergruppe Volkswirtschaftslehre

Autoren / Authors

Jörn Kruse
Helmut Schmidt Universität Hamburg / Helmut Schmidt University Hamburg
Institut für Wirtschaftspolitik / Institute of Economic Policy
Holstenhofweg 85
22043 Hamburg
Germany
joern.kruse@hsu-hh.de

Redaktion / Editors

Helmut Schmidt Universität Hamburg / Helmut Schmidt University Hamburg
Fächergruppe Volkswirtschaftslehre / Department of Economics

Eine elektronische Version des Diskussionspapiers ist auf folgender Internetseite zu finden/
An electronic version of the paper may be downloaded from the homepage:
http://fgvwl.hsu-hh.de/wp-vwl

Koordinator / Coordinator

Kai Hielscher
wp-vwl@hsu-hh.de

Helmut Schmidt Universität Hamburg / Helmut Schmidt University Hamburg
Fächergruppe Volkswirtschaftslehre / Department of Economics

Diskussionspapier Nr. 96
Working Paper No. 96

PRIORITY AND INTERNET QUALITY

JÖRN KRUSE

Zusammenfassung/ Abstract

The significant increase of internet traffic is to a large extent caused by more high-data-rate applications like file-sharing etc. Although network operators constantly increase router and line capacities, overload occurs from time to time, causing delays, jitter and packet-losses at the data packet level. At the service level this may significantly reduce the quality of certain applications. Among those are interactive services like VoIP, some business applications, online gaming etc. and other on-time-services like internet-television.

This will result in systematic inefficiencies which can mainly be attributed to two reasons. The first one is the widespread use of internet-flatrates and the second one is a much too strict interpretation of the network neutrality principle. The latter describes the fact that every single data packet will be handled strictly equal at every router, no matter what application it belongs to and what the technical and economic consequences of delayed or dropped packets will be.

A congestion model shows that flatrates which users' marginal outlays cause to be zero are inefficient as soon as positive marginal overload externalities exist. It comes up with a general pricing solution which will be questioned later on. In this framework, optimal internet capacity is identified and the networks' overprovisioning policy is found to be inefficient.

The key issue for internet congestion problem is the fact that, although services are homogenous at the data packet level, they are very heterogenous at the service level. They are very different with respect to data rate, quality-sensitivity, and economic value. Under strict network neutrality rule it can be demonstrated that certain valuable, quality-sensitive services will be significantly harmed (and potentially be crowded out altogether) by non-quality-sensitive, high-data-rate services which may have low economic value.

Giving priority to certain services in overload situations looks like the adequate solution to the problem. However, this always bears the risk of discriminating some service providers and applications and will be heavily debated. A more appropriate solution is provided by priority pricing, whereby users express their willingness to pay for priority treatment in case of an overload. Customers have an ex ante choice between different qualities of service. The choice of a service provider to pay for high priority (high quality of service) will depend mainly on two factors, the quality-sensitivity and the end-users' willingness to pay for such services. Only providers of quality-sensitive services will have any reason whatsoever to pay

for traffic prioritization. Providers of non-quality-sensitive services (file sharing, e-mailing, webbrowsing) will be adequately served by best effort traffic and will thus obtain it cheaply.

Priority pricing (quality of service) results in an economically efficient use of scarce router capacity according to the economic congestion effects of the specific service. It avoids the crowding-out problem. It allows to generate more economic value out of a given internet capacity.

JEL-Klassifikation / JEL-Classification: L86, L96

Schlagworte / Keywords: Internet, Quality of Service, Priority Pricing, Overprovisioning, Filesharing

1. INTRODUCTION

Internet traffic is increasing significantly. This is mostly caused by an increasing number of users and, especially, by more high-data-rate applications like file-sharing etc. (Schulze/Mochalski, 2009). Although network operators constantly increase router and line capacities, overload occurs at several times leading to quality deterioration. When the number of data packets exceeds router capacity, additional packets will be intermediately stored and, with more traffic coming in, will finally be dropped altogether (Ganley/Allgrove, 2006; Marcus, 2006). Overload is leading to increased delay, jitter and packet-loss which may significantly reduce the quality of certain applications. Among those are interactive services like VoIP, online gaming etc. and other on-time-services like internet-television. In this paper it will be demonstrated that this will result in systematic inefficiencies and a general solution will be presented.

The overload problems are mainly due to two factors which, at the same time, deserve some credit for the internet success in general. The first one is the widespread usage of internet-flat-rates and the second one is the network neutrality principle. If flatrates are the internet users' typical pricing schedule, their marginal costs with respect to data packets are zero. It is discussed in section 2 that this is inefficient as soon as overload occurs deteriorating the quality of at least one service. Additionally, any conventional congestion pricing would not be optimal because of some specific internet characteristics (section 5).

Network neutrality in a moderate sense describes the fact that every single data packet is treated equal at every router, no matter who was sending the data packets, which destination they go to, and what application they belong to. An interpretation that goes much further includes, in addition, the strictly equal treatment of all packets no matter what the technical and economic consequences of blocked or dropped packets would be and regardless of what the willingness to pay of the users would be.

In the United States, there has been an intensive debate on wether or not (and in what sense) network neutrality should be regulated by law (Sidak, 2006; Hahn/Wallsten, 2006; van Schewick, 2007; Litan/Singer, 2007; Economides, 2007). The controversy was fostered by the intention of some network operators to charge content providers like Google or Amazon or services like voice over IP differently from others or to treat their data packets differently depending on some criteria and presumably depending on the effect of these services on their own businesses (Cheng/

Bandyopadhyay/Guo, 2007). Any network operators intervention into the internet content flow is seen as a discrimination and as an offence against the free internet.

On the other hand, neglecting the economic characteristics and values of individual data packets at the router level turns out to be inefficient because different services and applications at the end user level are reacting very differently to overload situations. This means that quality-sensitivity is very different among services. One consequence of such a situation is a tendency that certain valuable, quality-sensitive services will be significantly harmed by high-data-rate services which may have low economic value. These issues on internet congestion, priority pricing and service qualities are at the core of the paper.

In the next section a congestion model will be applied to internet overload to resolve this problem, including a general congestion pricing solution. In section 3, one of the potential network strategies to avoid internet congestion, the overprovisioning of internet capacity, will be discussed and shown to be inefficient. Section 4 considers that the internet services are very different with respect to data rate, quality-sensitivity, and economic value. It will be shown that economically valuable, quality-sensitive services may be crowded out by high-data-rate, non-quality-sensitive services that may be of low economic value. In section 5 the concept of priority pricing is introduced as a solution to the beforementioned problems. Section 6 summarizes and concludes.

2. INTERNET OVERLOAD, PRICING AND WELFARE

In economic terminology, reduction in quality due to overload (synonymous with congestion) is caused by a "partial rivalry". This partial rivalry is defined in particular by the fact that although serving an additional user, does not exclude other users, it affects all users by reducing the quality of their service utilization.

This phenomenon can be expressed theoretically in the form of a "quality-adjusted demand function" DQ_C (F 2.1), where X is the quantity of data packets per time slot. The "congestion-free" demand function D* (ZX_9) shows the demand and the willingness to pay for the use of the internet infrastructure. It is assumed that non-rivalry prevails up to quantity X_3 which means that the routers forward all packets without any loss or delay.

Beyond X_3, the quality of some services is reduced due to packet delays and/or losses. The users' willingness to pay for these services decreases. The

resulting overall quality-adjusted demand function DQ_C is depicted by the curve ZJ_cX_7. Since the price is equal to zero, the relevant quantity is X_7.

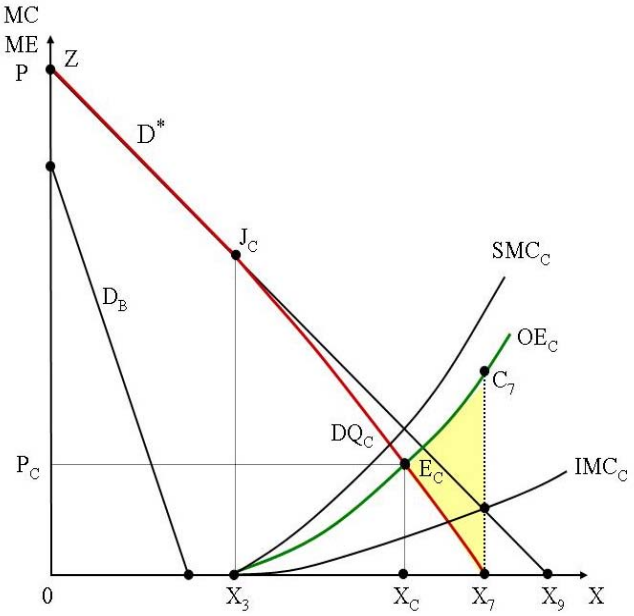

Figure 2. 1: Partial rivalry and overload inefficiency in the internet

The vertical distance between the overload-free demand function D* and the quality-adjusted demand function DQ_C for each quantity of data packet is shown by the individual marginal costs of congestion IMC_C. The social marginal costs of all users are SMC_C. The difference between IMC_C and SMC_C is OE_C. OE_C are negative external effects on third parties and will be referred to as overload externalities (synonymous with congestion externalities).

According to conventional congestion models[1] the welfare maximising data volume would be X_C, determined by the point of intersection E_C of the quality-adjusted demand function DQ_C and the marginal overload externalities curve OE_C. The actual quantity, however, is X_7, causing allocative

[1] Congestion models have been developped mostly with applications to road traffic. See for discussions on congestion models Gomez-Ibanez/Small, 1994; Kruse/Berger, 1995; Santos, 2004; Button, 2004; Knieps, 2007.

inefficiency with a welfare loss equal to the triangle $E_C X_7 C_7$. Congestion models suggest that this problem can be solved by implementing a usage price P_C, which is derived by the optimum quantity X_C and the relevant demand function DQ_C. A theoretically equivalent rationing of the quantity to X_C is neither practicable nor efficient in the internet.

The price P_C per data packet reflects the opportunity costs (reduction in quality) and differentiates (and rations) basically between high value and low value services in an efficient manner. But the optimum price P_C only applies for the specific demand function N* which would only exist for short periods since usuage fluctuates. If, at other times of the week, demand were lower (higher), so would be the optimal congestion price. If, for example, the demand function were D_B, the optimum price would be equal to zero. This would apply for internet routers for most time periods, since overload only appears at specific peak usage times (neglecting the case of infrastructure breakdown due to natural or other catastrophies). If it were possible to implement such a peakload pricing schedule in the internet, prices would vary significantly over time.

However, this is not available for the internet. Peakload pricing requires that the demand functions in future periods can be anticipated. But internet demand is, to a certain extent, stochastic and therefore not possible to forecast. Overload often occurs so suddenly that, even if prices could adequately reflect that, users would not be able to effectively adjust usage to prices.

But even if these problems could be resolved, there is an extra specific internet problem. The congestion model assumes homogeneity among users with respect to quality reduction. Since the internet services are affected very differently by temporary overload and their quality-sensitivity with respect to delay, loss and jitter at the data packet level is totally different (see section 4), a specific congestion price can only be regarded as a first approximation, but is not the optimal solution and is inferior to priority pricing (see section 5).

3. INTERNET CAPACITY AND OVERPROVISIONING

In a long-term view it would principally be possible to avoid most of the overload problems if investments were made into larger infrastructure capacities (routers, transmission line etc.). Capacity is defined as the maximum number of data packets that can be conveyed per period (in a very short time slot) without any overload (X_3 in F 2.1). As an extreme one might consider very large capacities such that all packets can always be forwarded

immediately. This requires to expand capacity from X_3 to X_9 , if D* were the highest demand that could ever be expected. Nevertheless, overload could also occur due to unexpected network failures as a result of disasters, which will be neglected in the following.

Sizing capacities to a potential maximum peak load is referred to as "overprovisioning". It requires high reserve capacities and incurs correspondingly high costs for network operators. This raises the question if overprovisioning would be economically efficient. What is the optimum capacity taking into account overload-induced reductions in quality and reduced utility?

The occurrence of overload and its quantitative effects on service quality reduction certainly depends on the capacity of the internet infrastructure. The smaller capacity is, the more likely is it that impairments will occur at peak times and the more severe they will be for a given set of demand functions (varying over time).

With capacity X_3 in F 2.1 total utility is $0X_3E_CJ_CZ$. If capacity is varied we derive the long-term total utility function LN(Y) in F 3.1. The utility increases with growing capacity and reaches its maximum at Y_M, where no overload occurs with the relevant demand function N* which is assumed to be the maximum demand. Non-rivalry in infrastructure use prevails throughout. If the capacity is further increased, LU(Y) remains constant. Overprovisioning exists, if $Y \geq Y_M$. Differentiation of the utility function LU(Y) to capacity gives the long-term marginal utility curve LMU(Y). It therefore shows the additional utility of an extra capacity unit. This utility is positive (although decreasing) until capacity Y_M is reached where it is equal to zero.

Increasing infrastructure capacity also incurs additional costs. For the sake of simplicity, the long-term marginal costs LMC(Y) of an additional capacity unit are assumed to be constant (although this is not essential), so the curve runs horizontal throughout.

The point of intersection of the marginal utility curve LMU(Y) and the marginal costs curve LMC(Y) determines the optimum capacity Y_{opt}. Up to this point the costs of an additional capacity unit are lower than the additional utility. To the right of this point, the additional consumption of resources is higher than the additional utility.

Since it can be assumed that the costs of expanding capacity are consistently positive, optimum capacity for the economy as a whole is generally smaller than the capacity that results in maximum utility for infrastructure users, i.e. complete freedom from overload. Thus, in a state of optimized welfare, utilization rivalries and overload externalities at certain peak times

(and eventually in cases of network failures) do exist. Since $Y_M > Y_{opt}$ overprovisioning is inefficient.

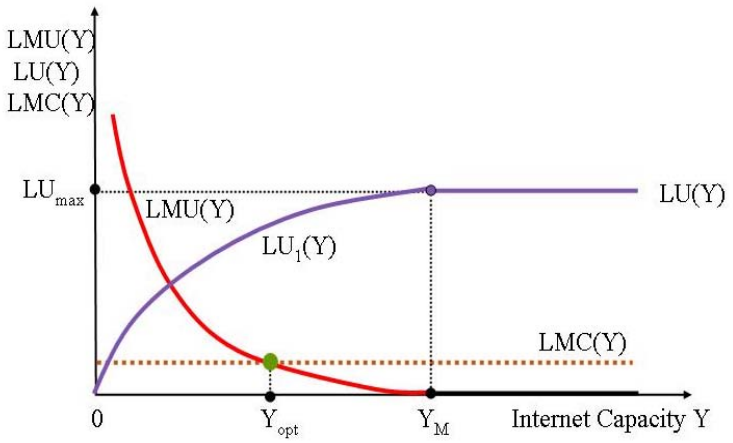

Figure 3.1: Optimum internet capacity

4. HIGH-VALUE AND QUALITY-SENSITIVE SERVICES AND THE CROWDING OUT EFFECT

Although data packets are homogenous at the transmission and switching level, they are not homogenous at all at the service level. The whole set of services and applications that share the intenet as a common conveyance infrastructure differ significantly with respect to several characteristics, namely (1) data rate, (2) quality-sensitivity, and (3) economic value.

(1) Data rate. Individual services have very different data rates which can be measured either by the number of packets per time unit or per typical unit of consumption. While some services like e-mail, webbrowsing etc. have comparatively small data rates, other services produce high workloads for the networks in off-peak as well as in overload times. Certain services thus play a particularly large role in the reduction in quality of all services. These include in particular downloads, especially downloads via file sharing plat-

forms.[2] A large percentage of contents consist of videos, music and software. P2P platforms and downloads of this type are responsible for a large proportion of the internet traffic. It is estimated that it makes up about 45 to 70 per cent of internet traffic depending on the region (Schulze/Mochalski, 2009).[3] This is not particularly surprising from an economic point of view since marginal costs of using these services are zero because of internet flat rates.

(2) Quality-sensitivity decribes the effect which internet overload has on the quality of a specific service, as from the viewpoint of the consumers and their willingness to pay. The reductions in quality due to congestion (delay, jitter, packet loss) differ extremely according to the service involved.

The qualities of some services are severely affected. These include interactive services (e.g. voice over IP, where delays over 150 milliseconds are not considered tolerable for the consumers, and online gaming, where delays of 50-100 milliseconds are already harmful) as well as many business applications and internet television. Other services will not be affected at all, only mildly, or only in extreme cases. These include elastic services where lost packets will be reordered from the source, such as e-mails, webbrowsing, and, especially, filesharing and other downloads.

(3) The economic value ist measured by the economic welfare per data packet which is the universal standard quantity unit in the internet. For the sake of simplicity it is assumed that producer surplus is zero. Then, economic value is consumer surplus per data packet. It varies significantly among services. Many business applications, a number of individual e-mails (often small number of packets), and some interactive services like voice over IP have high economic value. In general, file sharing platforms have low economic value. Consumers have a very limited willingness to pay for the download of music and movies. Since these downloads incur large data volume, the consumer surplus per data packet is low.

Let us now analyze the potential rivalry between two services, S_L and S_H, that use the internet traffic capacity as a common resource. S_H is assumed to be a highly quality-sensitive service with high economic value. By contrast,

---

[2] The term “file sharing” is used here according to normal customer usage or peer-to-peer (P2P), although the wording is actually not appropriate. This includes in particular (frequently large-volume) downloads and uploads of videos, music and software. It may be mentioned, that many of these contents are in fact illegal (copyright violations).

[3] In 2007 this figure has been even higher. The P2P-filesharing percentage in internet traffic was 83,5% for Eastern Europe, 63,9% for Southern Europe, 49% for Middle East, and 57,2% for Australia. In Germany, P2P-filesharing accounted for 69,25% of the internet workload, webbrowsing 10,05%, media streaminng (incl. YouTube etc.) 7,75%, VoIP 0,92%, E-Mail 0,37%. See for more details Schulze/Mochalski, 2007.

S_L is a low value service with low quality-sensitivity. These characteristics are summarized in table 4.1.

Table 4.1: Rivalry between two services

Service	Examples	Quality-Sensitivity	Economic value per data packet
S_L	File sharing platform, P2P downloads	None	Low
S_H	Business applications, Interactive services (Voice over IP)	High	High

We will now consider the demand function D_{L1} for S_L in the starting period t_1 as shown in F 4.1. We assume that D_{Li} (i=1,2,...) represents demand for a platform traffic which enables consumers to carry out file sharing and to make downloads free of charge. It is financed by advertising. Since this service is free, i.e. it has a price of zero, the saturation quantity X_1 determines the number of data packets that have to be processed by the internet infrastructure. In this case, consumer surplus (CS) for users is represented by the triangle $0X_1Z$ (CS=100*4/2=200).

The demand function D_W of advertisers must also be included. According to this demand, the advertising contact price P_W for quantity X_1 is one. The advertising contact price is defined as the advertising revenue per data packet. This generates advertising revenues of $E_1 = X_1 * P_W$ ($W_1Y_1X_10$) to the platform (E=100*1). Let us assume that this is just sufficient to cover the costs of the platform which are constant. It results in a consumer surplus of $W_1W_2Y_1$ for advertisers (CS_{W1}=100/2*(2-1)=50).

It is sufficient for the following analysis to represent the value of the service for the economy as a whole, $V_{L1,}$ by the sum of the producer and the consumer surplus on both markets. If we neglect for the comparison the constant costs of the platform ($W_1Y_1X_10$) for the sake of simplicity, V_{L1} is represented by the area $W_2Y_1X_1Z$ (4*100/2 + 100*1 + 100*1/2 = 350).

In F 4.2, D_{H1} represents the demand function for a high-value, quality-sensitive service S_H. The higher value of D_H (compared to D_L) is indicated by the numbers of the ordinate scale. Let us assume that there is sufficient internet capacity available at time t_1 so that there is no rivalry between the two services at that instant. This means that D_{H1} represents the overload-free

(congestion-free) demand function D_H* with a saturation quantity M_1 (100).[4] The value V_{H1} which is generated by the service to the economy as a whole then corresponds to the area OM_1Z (100*100/2= 5,000).

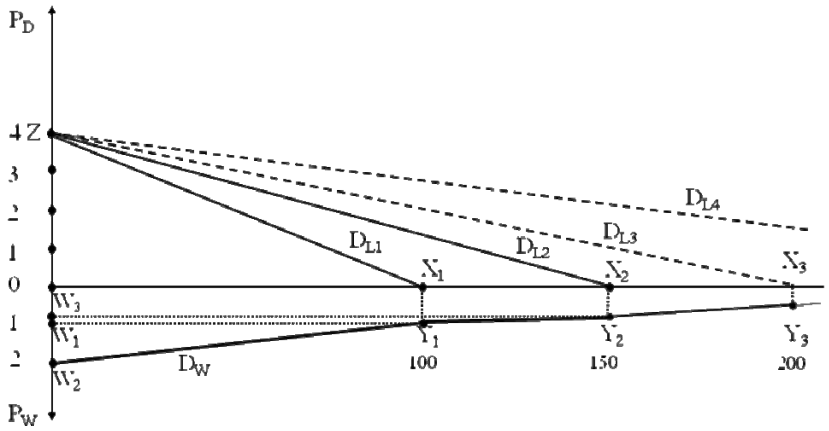

Figure 4.1: Non-quality-sensitive, low-value service S_L

It is assumed for the next period t_2 that demand for service S_L (F 4.2) increases to D_{L2}. This implies a growth of internet traffic from X_1 to X_2 (150) resulting in an increase in the users' consumer surplus to $0X_2Z$. (i.e. 150*4/2=300). Let us also assume that the advertising contact price drops to P_{W3}=0.8.[5] The advertising revenues are then $E_2=X_2*P_{W2}$ (150*0.8=120). The

[4] The saturation quantity M_1 of internet traffic is directly relevant in case the service is made available for free. If, instead, it were a pay-service with prices above zero, the relevant quantities were somewhat lower. However, the rivalry effect would still qualitatively be the same.

[5] This assumption is based on different potential links between the number of data packets and the attention for advertising messages. (1) If the increase in the number of data packets is based on additional users, it would be plausible to assume constant advertising contact prices. Advertising revenues would then increase with the number of users. (2) If the same users would download more than before, the advertising contact price will probably fall and the advertising revenues might still increase slightly. (3) If only the data rate would increase (high definition instead of conventional video), the advertising revenues would not grow, but the advertising contact price per data packet would fall considerably. All three effects may occur at the same time. The net effect on advertising contact price per data packet is unclear. Advertising revenues will increase with more data packets, if we assume that different effects occur.

advertising consumer surplus is 75. The economic value V_{12} corresponds to the area $W_2Y_1Y_2X_2Z$ (300+120+75=495).

Assuming that the increase of service S_L to X_2 induces overload, the quality of the quality-sensitive service S_H in F 4.2 is reduced, such that the new demand function D_{H2} applies. The economic value of S_H drops to $0M_2Z_2$ (80*70/2=2,800).

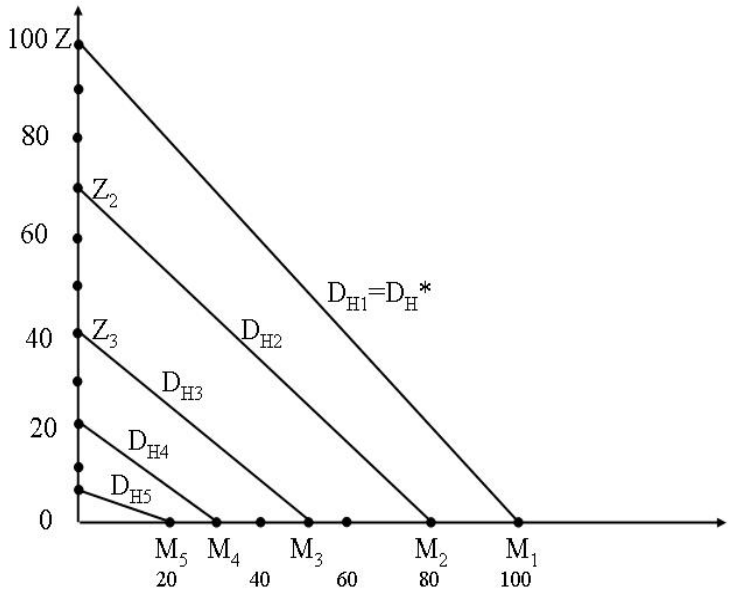

Figure 4.2: Quality-sensitive, high-value service S_H

Considering the value effects on both markets together, the growth of service S_L (495-350=145) corresponds with decrease of service S_H (5,000-2,800=2,200), resulting in a net loss (2,200-145=2,055) due to quality deterioration of S_H. If we assume that the demand function D_L of S_L continues to shift to the right (and data quantity increases), overload situations occur more frequently and more severely. The detrimental effect on the quality of service S_H is increasing such that demand functions D_H in F 4.2 shift further to the left. Potentially, S_H is no longer economically viable which would mean that the high-value service S_H has been crowded out by the low-value service S_L.

In such a rivalry situation between two selected services with the stated combination of quality-sensitivity and economic value, a “lower value service” can oust a high-value service, a process which is termed “crowding out”.

Although the “crowding out effect” has been formulated in a somewhat exaggerated manner it highlights a relevant problem of the internet. Considering the volumes of download and file sharing traffic which themselves are not quality-sensitive it can be anticipated that high-value services will be significantly harmed. Additionally, innovative services requiring high quality standards may not be developed at all even if they have high economic value.

Two economic factors specifically associated with the internet, user flatrates and a strict interpretation of network neutrality, can be identified to produce or foster the beforementioned problems. Substituting flatrates by pricing schedules that are based on traffic volume would prevent some of the problems, but would not be economically efficient, as has been mentioned in section 2 and will be further discussed in section 5. An interpretation of network neutrality that would allow to vary prices according to different qualities (but not discriminating specific services or data sending parties) seems to be necessary to reach efficient solution. This is priority pricing which will be discussed in the following.

5. PRIORITY PRICING AND QUALITY OF SERVICE

The congestion analysis in section 2 suggests that the optimal price is P_C, if the relevant demand function is D*. Since demand fluctuates and thus also the optimal congestion price, a peakload pricing structure with an efficient allocative effect, requires that the demand functions in future periods can be anticipated and that users will be able to effectively adjust their usage to prices. These conditions are unrealistic in the internet.

But even if these difficulties could be reasonably handled, there is another internet specific conceptual problem which is most severe. The congestion model assumes that the overload effects on quality, as from the viewpoint of the consumers and their willingness to pay, are basicly the same for all services. However, it has been mentioned in section 4 that the quality-sensitivities of individual internet applications and services are extremely different. Interactive services (like voice over IP), many business applications and internet television are highly quality-sensitive and will be significantly affected by internet congestion. Elastic services like filesharing and

other downloads, e-mailing and webbrowsing will often not be affected at all.

Overload situations at specific routers often only last for a few seconds before router capacity etc. becomes available again. If delay, jitter and packet loss cause noticeable reduction in quality to some services, but not to others which can handle these easier, it is merely a problem of treating the data packets differently according to the services involved. Technically, it is comparatively easy for network operators to prioritize specific services according to the information in the data packets headers.

Adequate prioritization means that (1) data packets of quality-sensitive services are forwarded immediately, and (2) data packets of non-quality-sensitive services have to wait or will be dropped if necessary. This is sometimes referred to as "needs-based discrimination". As such it does not infer anticompetitive practises (Ganley/Allgrove, 2006) and it may or may not comply with the network neutrality principle. Any service-specific prioritization is somewhat discriminating in the sense that different service providers and/or their customers do not have the same chance to get priority, even if their economic value would be high. It will almost inevitably lead to interest conflicts and debates which services should enjoy preferential treatment.

A solution that would avoid any of these problems is the concept of "priority pricing" (Telson, 1975; Chao/Wilson 1987, 1990; Kruse/Berger, 1998). This denotes a pricing scheme in which the right to receive priority service is available to any customer in exchange for higher prices. This may include a higher or lower number of different priority classes where each has a specific priority ranking. With priority pricing consumers express their willingness to pay for preferential treatment in case of an overload before it actually occurs. Customers are thereby given the ex ante choice of opting for different levels of delay, jitter, and packet loss (and thus different transport qualities). The higher price for preferential treatment then generally applies, regardless of whether overload actually occurs or not. Priority pricing is equivalent with the implementation of a "quality of service" (QoS) regime (Brenner/Dous/Zarnekow/Kruse, 2007).

A high quality of service is therefore synonymous with a high probability that the data packets will arrive with no or minimum delay, jitter and packet loss. Service providers (or senders of data packets in general) can individually select among two ore more quality classes which differ with regard to priorities and prices. Their willingness to pay for high priority (high quality of service) will depend mainly on two factors: (1) the quality-sensitivity and (2) the willingness to pay on the part of the users of the service.

(1) Only providers and/or users of quality-sensitive services will have any reason whatsoever to pay for priority since non-quality-sensitive services will not gain any advantage from it. Providers of non-quality-sensitive services (file sharing, e-mailing, web browsing,) will be adequately served with the least preferential "best-effort" class and will thus obtain the internet service cheaply.

(2) Providers of quality-sensitive services will only be willing to pay for the quality of the service (high priority of the data packets) if the users of these services (or indirectly the advertisers), for their part, are also ready to pay for the quality of these services. This means that, in general, only high-value services will choose a high priority class.

When overload occurs under a priority pricing regime, only those data packets will be subject to delay, jitter and/or packet loss where this occurrence infers the lowest disutility. This allows to say that priority pricing results in economically efficient rationing of scarce router and line capacity according to the value of the services. Priority pricing also guarantees that the above-mentioned crowding-out problem will not occur.

After a successful implementation of priority handling of quality-sensitive data packets, a flat rate for the best effort class of service will not be as disadvantageous from the economic viewpoint as before. Despite an end users' flat rate, with the existence of higher priority classes there will no longer be any crowding-out effect to the detriment of high-value services. Therefore, if there are other reasons (e.g. marketing for DSL services) for offering a flat rate for best effort service, this is acceptable if quality of service classes also exist.

Let's consider a specific situation of traffic overload. Priority pricing takes care that the quality-sensitive services will enjoy preferential treatment such that no quality problem will occur. But also most of the non-quality-sensitive services may not be harmed as long as capacity problems will last only for short time-slots such that delays are acceptable for the users and packet losses can be repaired by elastic applications.

The economic consequences of implementing a QoS regime can also be demonstrated in the capacity framework in F 5.1 (based on F 3.1). Priority pricing changes the earlier long-run utility function $LU_1(Y)$ into $LU_2(Y)$. Since priority pricing uses scarce capacity more efficient in overload situations, the same capacity (Y_1) now produces more economic value ($LU_2 > LU_3$). The same value (LU_2) requires less capacity ($Y_1 < Y_2$) und thus saves investment capital and operating costs.

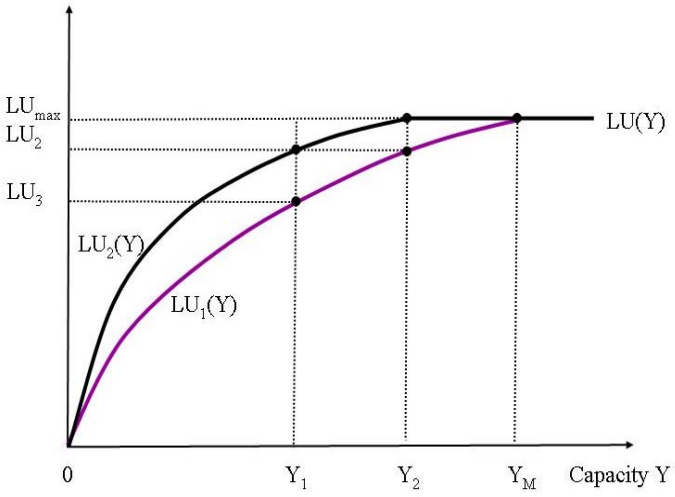

Figure 5.1: Effect of priority pricing on utility and capacity

6. SUMMARY AND CONCLUSIONS

The significant increase of internet traffic is to a large extent caused by more high-data-rate applications like file-sharing etc. Although network operators constantly increase router and line capacities, overload occurs from time to time, causing delays, jitter and packet-losses at the data packet level. At the service level this may significantly reduce the quality of certain applications. Among those are interactive services like VoIP, some business applications, online gaming etc. and other on-time-services like internet-television.

This will result in systematic inefficiencies which can mainly be attributed to two reasons. The first one is the widespread use of internet-flatrates and the second one is a much too strict interpretation of the network neutrality principle. The latter describes the fact that every single data packet will be handled strictly equal at every router, no matter what application it belongs to and what the technical and economic consequences of delayed or dropped packets will be.

A congestion model shows that flatrates which users' marginal outlays cause to be zero are inefficient as soon as positive marginal overload externalities exist. It comes up with a general pricing solution which will be

questioned later on. In this framework, optimal internet capacity is identified and the networks' overprovisioning policy is found to be inefficient.

The key issue for internet congestion problem is the fact that, although services are homogenous at the data packet level, they are very heterogenous at the service level. They are very different with respect to data rate, quality-sensitivity, and economic value. Under strict network neutrality rule it can be demonstrated that certain valuable, quality-sensitive services will be significantly harmed (and potentially be crowded out altogether) by non-quality-sensitive, high-data-rate services which may have low economic value.

Giving priority to certain services in overload situations looks like the adequate solution to the problem. However, this always bears the risk of discriminating some service providers and applications and will be heavily debated. A more appropriate solution is provided by priority pricing, whereby users express their willingness to pay for priority treatment in case of an overload. Customers have an ex ante choice between different qualities of service. The choice of a service provider to pay for high priority (high quality of service) will depend mainly on two factors, the quality-sensitivity and the end-users' willingness to pay for such services. Only providers of quality-sensitive services will have any reason whatsoever to pay for traffic prioritization. Providers of non-quality-sensitive services (file sharing, e-mailing, webbrowsing) will be adequately served by best effort traffic and will thus obtain it cheaply.

Priority pricing (quality of service) results in an economically efficient use of scarce router capacity according to the economic congestion effects of the specific service. It avoids the crowding-out problem. It allows to generate more economic value out of a given internet capacity.

7. REFERENCES

Brenner, Walter; M. Dous; R. Zarnekow und J. Kruse (2007), *Quality in the Internet. Technical and economic development prospects*, University of St. Gallen, March.

Brenner, W.; J. Kruse; R. Zarnekow and A. Sidler (2008), 'Qualität im Internet', *Elektrotechnik und Informationstechnik, Spezialausgabe Dynamik der Kommunikationsnetze*, 125 (7/8), 268–273.

Button, Kenneth (2004), 'The Rationale for Road Pricing: Standard Theory and Latest Advances', in G. Santos, (ed.), *Road Pricing: Theory and Evidence. Research in transportation economics*, Vol 9, S. 3–25.

Chao, H.-P. and R. B. Wilson (1987), 'Priority-Service: Pricing, Investment, and Market Organization', *American Economic Review*, 77 (5), 899–916.

Chao, H.-P. and R. B. Wilson (1990), 'Optimal Contract Period for Priority Service', *Operations Research*, 38 (), 598–606.

Cheng, H., S. Bandyopadhyay und H. Guo (2007), *The Debate on Net Neutrality: A Policy Perspective*, Mimeo, University of Florida.

Economides, N. (2007), 'Net Neutrality, Non-Discrimination and Digital Distribution of Content Through the Internet', *AEI-Brookings Joint Center for Regulatory Studies*.

Ganley, P. und B. Allgrove (2006), 'Net Neutrality: A user's guide', *Computer Law and Security Report* 22, 454-463.

Gomez-Ibanez, Jose A. und Kenneth A. Small (1994), 'Road Pricing for Congestion Management: A survey of International Practice', in: Tranportation Research Board (Hrsg.), *National Cooperative Practice 210*, Washington.

Hahn, R.W. und S. Wallsten (2006), 'The Economics of Net Neutrality', *Discussion Paper, AEI-Brookings Joint Center for Regulatory Studies*.

Knieps, Günter (2007), *Netzökonomie*, Wiesbaden (Gabler).

Kruse, J. (2008), 'Network Neutrality and Quality of Service', *Inter-economics, Review of European Economic Policy*, 43 (1), January/February, 25–30.

Kruse, Jörn und U.E. Berger (1995), ‚Stauprobleme und optimale Straßenkapazität', *Jahrbuch für Wirtschaftswissenschaften* Bd. 46, Heft 3, S. 295-305.

Kruse, Jörn und U.E. Berger (1998), ‚Priority Pricing und zeitkritische Rationierung', in: Tietzel, Manfred (Hrsg.), *Ökonomische Theorie der Rationierung*, München (Vahlen), S. 203-234.

Litan, R. and H. Singer (2007), 'Unintended Consequences of Net Neutrality', *Journal of Telecommunications and High Technology Law*, 5 (3).

Marcus, J. Scott, (2006), 'Interconnection in an NGN Environment', *ITU background paper*, April 15, commissioned for the ITU New Initiatives Programme workshop "What rules for IP-enabled Next Generation Networks?" held on 23-24 March 2006 at ITU Headquarters, Geneva. November 2006.

OECD (2006), *Internet Traffic Prioritisation: An Overview. Working Party of Telecommunication and Information Services Policies* (Tylor Reynolds), Dublin, November 7, 2006.

Santos, G. (ed.)(2004), *Road Pricing: Theory and Evidence*, Elsevier, Amsterdam.

Schulze, Hendrik und K. Mochalski (2007), 'The Impact of P2P File Sharing, Voice over IP, Skype, Joost, Instant Messaging, One-Click Hosting and Media Streamimg such as YouTube on the Internet', *Ipoque Internet Study* 2007,

Schulze, Hendrik und K. Mochalski (2009), Internet Study 2008/2009, Ipoque, http://www.ipoque.com/resources/internet-studies.

Sidak, G. (2006), 'A Consumer-Welfare Approach to Network Neutrality Regulation of the Internet', *Journal of Competition Law and Economics*, 2 (3), 349–474.

Telson, M. L. (1975), 'The Economics of alternative Levels of Reliability for Electric Power Generation Systems', *The Bell Journal of Economics*, 6, (2), 679–694.

Van Schewick, B. (2007), 'Towards an Economic Framework for Network Neutrality Regulation', *Journal on Telecommunications and High Technology Law*, 5, 329–391.

DISKUSSIONSPAPIERE DER FÄCHERGRUPPE VOLKSWIRTSCHAFTSLEHRE

DISCUSSION PAPERS IN ECONOMICS

Die komplette Liste der Diskussionspapiere ist auf der Internetseite veröffentlicht / for full list of papers see: http://fgvwl.hsu-hh.de/wp-vwl

2009

96 Kruse, Jörn. Priority and Internet Quality, August 2009.

95 Schneider, Andrea. Science and teaching: Two-dimensional signalling in the academic job market, August 2009.

94 Kruse, Jörn. Das Governance-Dilemma der demokratischen Wirtschaftspolitik, August 2009.

93 Hackmann, Johannes. Ungereimtheiten der traditionell in Deutschland vorherrschenden Rechtfertigungsansätze für das Ehegattensplitting, Mai 2009.

92 Schneider, Andrea; Klaus W. Zimmermann. Mehr zu den politischen Segnungen von Föderalismus, April 2009.

91 Beckmann, Klaus; Schneider, Andrea. The interaction of publications and appointments - New evidence on academic economists in Germany, März 2009.

90 Beckmann, Klaus; Schneider, Andrea. MeinProf.de und die Qualität der Lehre, Februar 2009.

89 Berlemann, Michael; Hielscher, Kai. Measuring Effective Monetary Policy Conservatism, February 2009.

88 Horgos, Daniel. The Elasticity of Substitution and the Sector Bias of International Outsourcing: Solving the Puzzle, February 2009.

87 Rundshagen, Bianca; Zimmermann, Klaus W.. Buchanan-Kooperation und Internationale Öffentliche Güter, Januar 2009.

2008

86 Thomas, Tobias. Questionable Luxury Taxes: Results from a Mating Game, September 2008.

85 Dluhosch, Barbara; Zimmermann, Klaus W.. Adolph Wagner und sein „Gesetz“: einige späte Anmerkungen, August 2008.

84 Zimmermann, Klaus W.; Horgos, Daniel. Interest groups and economic performance: some new evidence, August 2008.

83 Beckmann, Klaus; Gerrits, Carsten. Armutsbekämpfung durch Reduktion von Korruption: eine Rolle für Unternehmen?, Juli 2008.

82 Beckmann, Klaus; Engelmann, Dennis. Steuerwettbewerb und Finanzverfassung, Juli 2008.

81 Thomas, Tobias. Fragwürdige Luxussteuern: Statusstreben und demonstratives Konsumverhalten in der Geschichte ökonomischen Denkens, Mai 2008.

80 Kruse, Jörn. Hochschulen und langfristige Politik. Ein ordnungspolitischer Essay zu zwei Reformutopien, Mai 2008.

79 Kruse, Jörn. Mobile Termination Carrier Selection, April 2008.

78 Dewenter, Ralf; Haucap, Justus. Wettbewerb als Aufgabe und Problem auf Medienmärkten: Fallstudien aus Sicht der „Theorie zweiseitiger Märkte“, April 2008.

77 Kruse, Jörn. Parteien-Monopol und Dezentralisierung des demokratischen Staates, März 2008.

76 Beckmann, Klaus; Gattke, Susan. Status preferences and optimal corrective taxes: a note, February 2008.

75 Kruse, Jörn. Internet-Überlast, Netzneutralität und Service-Qualität, Januar 2008.

2007

74 Dewenter, Ralf. Netzneutralität, Dezember 2007

73 Beckmann, Klaus; Gerrits, Carsten. Making sense of corruption: Hobbesian jungle, bribery as an auction, and DUP activities, December 2007.

72 Kruse, Jörn. Crowding-Out bei Überlast im Internet, November 2007.

71 Beckmann, Klaus. Why do petrol prices fluctuate so much?, November 2007.