NET Institute*

www.NETinst.org

Working Paper #10-09

September 2010

Network Neutrality and Congestion Sensitive Content Providers: Implications for Service Innovation, Broadband Investment and Regulation

Jan Krämer, Lukas Wiewiorra
Karlsruhe Institute of Technology
Institute of Information Systems and Management

* The Networks, Electronic Commerce, and Telecommunications (“NET”) Institute, http://www.NETinst.org, is a non-profit institution devoted to research on network industries, electronic commerce, telecommunications, the Internet, “virtual networks” comprised of computers that share the same technical standard or operating system, and on network issues in general.

Network Neutrality and Congestion Sensitive Content Providers: Implications for Service Innovation, Broadband Investment and Regulation

Jan Krämer, Lukas Wiewiorra*

Karlsruhe Institute of Technology
Institute of Information Systems and Management
Karlsruhe, Germany

September 2010

Keywords:
Telecommunications, Net Neutrality, Quality of Service, Innovation, Investment, Regulation

JEL Classification: D42; L12; L43; L51; L52; L96

---

*Address correspondence to kraemer@kit.edu. Financial support from the Graduate School 895 "Information Management and Market Engineering" at the Karlsruhe Institute of Technology, funded by Deutsche Forschungsgemeinschaft (DFG), and from the NET Institute is gratefully acknowledged.

Abstract

We consider a two-sided market model with a monopolistic Internet Service Provider (ISP), network congestion sensitive content providers (CPs), and Internet customers in order to study the impact of Quality-of-Service (QoS) tiering on service innovation, broadband investments, and welfare in comparison to network neutrality. We find that QoS tiering is the more efficient regime in the short-run. However it does not promote entry by new, congestion sensitive CPs, because the ISP can expropriate much of the CPs' surplus. In the long-run, QoS tiering may lead to more or less broadband capacity and welfare, depending on the competition-elasticity of CPs' revenues.

1 Introduction

A decade after Lawrence Lessig (2001) first stated the principle of network neutrality, the debate on this regulatory issue has developed many faces and battlefields. Academics, practitioners, consumer rights groups and regulators alike continue to be caught in the debate that centers around the idea of Internet Service Providers (ISPs) exercising control over the data traffic that flows through their networks. In this context, the meaning of 'control' is often ambiguous and can mean anything from blocking certain types of undesired or unaffiliated traffic (Wu and Yoo, 2007), to termination fees (Lee and Wu, 2009), to offering differentiated services and taking measures of network management (Hahn and Wallsten, 2006). The most publicly visible part of the debate usually focuses on the relationship between ISPs and end customers. Here, advocates of network neutrality fear that increasing commercialization endangers the free and open spirit of the Internet that has made it a success and governed free speech and the availability of legal content (Lessig, 2001, 2002; Cerf, 2006). However, there seems to be little disagreement over the fact that ISPs should not act as Internet gatekeepers by preventing legal content from reaching consumers.

The most academically controversial part of the net neutrality debate is the question of the future relationship between ISPs and content providers (CPs). We seek to investigate in particular whether ISPs should be allowed to offer CPs differentiated service classes for the transmission of their data packets to end customers–known as Quality-of-Service (QoS) tiering (Lessig, 2001, p.46; Hahn and Wallsten, 2006). Under a network neutrality regime, the prioritization of paid-for traffic would be prohibited, even if QoS tiering was offered on a non-discriminatory basis.[1] This part of the debate is both more relevant and more controversial with respect to potential consequences for welfare and public policy (Sharma, 2010).

Proponents of network neutrality argue that only this regime can ensure a level playing field for competition among content providers and will thus lead to more content variety and service innovation (Lessig, 2001, p.168–175; Wu, 2003; Van Schewick, 2006; Sydell, 2006). At a given transmission capacity, the acceleration of priority traffic will lead unmistakably to a deceleration of the remaining best-effort traffic. Thus, those CPs who are not willing to pay for priority access are put at a disadvantage twice: first, from the speeding up of other providers' content, and second, from the slowing down of their own content. This disadvantage lies at the heart of the concern of network neutrality proponents. Moreover, advocates claim that in the long-run, broadband infrastructure investments are likely to be higher under network neutrality, because ISPs are forced to maintain sufficient quality of service in order to bring new content types on line and keep consumers satisfied (Lessig, 2001, p.47). They even express the fear that QoS tiering may in fact hinder the roll-out of additional transmission capacity, because ISPs seek to charge CPs for exactly this resource, which is only possible if it is scarce (Wu and Yoo, 2007).

Opponents of a network neutrality regime argue to the contrary that QoS tiering will stimulate more service innovation and broadband investment. A CP who offers an on-demand video streaming service, for example, is certainly more sensitive to network congestion than a simple e-mail service provider.[2] Consequently, they argue that the best-effort one-size-fits-all transmission regime of a neutral network is not appropriate anymore (Yoo, 2005). If customers' experience of use is unsatisfactory because a CP's service cannot be reliably offered, this CP's online advertisement revenues will decline, possibly

---

[1] Here, 'non-discrimination' means that content providers can self-select whether they want to buy priority treatment for their data packets or not. However, in the past network providers have often discriminated data packets based on their content type. Examples for such anti-competitive behavior are the blocking of voice-over-IP transmissions by mobile network operators (Hahn et al., 2007) and the degradation of peer-to-peer traffic (O'Connell, 2005).

[2] For expositional simplicity, in the following we will often use 'congestion' or 'speed' as a proxy for different transmission quality measures, such as bandwidth, latency or jitter.

up to the point where she is forced out of business (Crowcroft, 2007). Hence, QoS tiering may in fact be welfare-enhancing, because it explicitly enables entry and innovation by those CPs who crucially hinge on transmission quality requirements which the traditional neutral best-effort Internet may soon be unable to provide. By contrast, net neutrality could hinder innovation, because innovations can only occur if the corresponding CPs' business models are also sustainable under the best-effort domain. Furthermore, supporters of QoS tiering argue that investments in broadband infrastructure would be higher under this regime because CPs can be billed for the transmission quality they are using (Van Schewick, 2006; Yoo, 2005). Even if transmission capacity is not extended,[3] QoS tiering might still be welfare-enhancing because it avoids some of the the wasteful network capacity overprovisioning by alleviating congestion for the most demanding content providers when needed.

In light of the arguments for and against network neutrality regulation, some observers have noted that the debate seems stuck in the sense that "at this point, it is impossible to foresee which architecture will ultimately represent the best approach"(Wu and Yoo, 2007).[4] In an effort to advance the debate, we provide a formal framework which incorporates the arguments of either side. This allows us to compare QoS tiering with network neutrality in terms of their impacts on service innovation, broadband investment and overall welfare. More specifically, we model the Internet as a two-sided network (Armstrong, 2006; Rochet and Tirole, 2006) that connects congestion-sensitive content providers with consumers, and which is controlled by a monopolistic ISP.[5] Under network neutrality regulation all content providers experience the same transmission quality, whereas under the QoS tiering regime, every content provider can choose to buy priority access to consumers on the same non-discriminatory conditions. In addition to the externalities that are generated by either side, we explicitly consider the adverse effect that traffic prioritization has on the transmission quality of the remaining best-effort class as well as the positive effect that congestion is allocated away from the most congestion-sensitive content providers. In this framework, we investigate the effects of QoS tiering both in the short-run,

---

[3] Crowcroft (2007), for example, suggests that QoS tiering is a "zero sum game at any instant".

[4] For a similar argument see Owen and Rosston (2006)

[5] In the context of network neutrality, the two-sided market framework was first suggested by Sidak (2006a,b).

when network capacity is fixed, as well as in the long-run, when the ISP can strategically invest in broadband infrastructure.

Our main results are that in the short-run QoS tiering will lead to the same level of entry and innovation by content providers as network neutrality. However, because QoS tiering allocates congestion better to the congestion insensitive CPs, overall short-run welfare is higher under this regime. Nevertheless, it should also be clear that QoS tiering enables ISPs to expropriate some of the CPs' revenues and thus, in the short-run, all CPs are worse off under QoS tiering than under network neutrality. Indeed, this fact has driven much of the emotionality in the recent debate.[6] Although the shift of revenues from content providers to the ISP is welfare neutral per se, it will generally still need to be scrutinized by policy makers in order to evaluate the consequences.

Furthermore, our analysis reveals that the difference between the net neutrality and the QoS tiering regime with respect to investments and long-run welfare will crucially depend on the elasticity of CPs' advertisement revenues with respect to competition (i.e. an increase of active content providers in the market). If CPs' ad revenues are elastic, the ISP has stronger incentives to invest in infrastructure under network neutrality. In this case the private investments are generally above the efficient level. On the other hand, if CPs' advertisement revenues are inelastic, the ISP will invest more in broadband infrastructure under QoS tiering, but then private investments are generally below the efficient level. Thus, we find support for arguments on both sides of the debate: when QoS tiering is socially preferred to net neutrality, ISPs will invest too little in broadband infrastructure. When net neutrality is preferred, ISPs are overproviding network capacity.

The remainder of this article is structured as follows. In Section 2 we discuss our framework in the context of related work, before we formally introduce the model in Section 3. Next, we investigate the differences between the QoS tiering and network neutrality regimes in the short-run with respect to service innovation (Section 4), and in the long-run with respect to broadband investments (Section 5). In Section 6 we consider the scope

---

[6] The debate was particularly stimulated after a blunt statement by Ed Whitacre, the Chief Executive Officer of ATT, who said: "Now what [content providers] would like to do is use my pipes free, but I ain't going to let them do that because we have spent this capital and we have to have a return on it" (O'Connell, 2005). More recently, similar statements have been released by major European network operators (Lambert, 2010; Schneibel and Farivar, 2010).

for regulatory intervention and comment on price regulation and minimum quality standards. Finally, in Section 7 we briefly comment on the possibility of strategic quality degradation under QoS tiering before we conclude in Section 8 by summarizing our results.

2 Related Work

Compared to the total number of academic papers that have been published in the context of the net neutrality debate, the number of formal economic papers within this domain is rather small. Schuett (2010) provides a recent and comprehensive overview of this literature. The contributions that are most related to ours are Cheng et al. (2010) and Choi and Kim (2010). Like them, we investigate the ISP's incentive to invest in network infrastructure under QoS tiering in a model that embodies standard results from queuing theory to formalize the relationship between priority and best-effort traffic. However, their model set-up differs from ours. The authors investigate the effect of QoS tiering for two competing content providers that are located at the end of a standard Hotelling line. It is assumed that customers dislike congestion and visit one of the two content providers exclusively. However, the content providers can improve their competitive position by purchasing priority access from a monopolistic ISP. The hitch in Choi and Kim (2010) is that the ISP will sell priority access to only one of the two content providers. The authors make this unconventional assumption in order to exclude a possible prisoners' dilemma situation that is observed in Cheng et al. (2010), who also allow the ISP to sell priority transmission to both content providers. More precisely, Cheng et al. find that when the difference in profit margins between the two content providers is rather small, both will individually buy priority access–but if both exercise this option, neither gains an advantage and the price paid for priority access is forfeited.

Our model, on the contrary, considers a continuum of content providers that is heterogeneous in their sensitivity towards congestion. Content providers endogenously choose to join the network if their business model is sustainable under the current congestion level. This allows us also to study the effect of QoS tiering on entry and service innovation, which is not possible in Cheng et al. (2010) and Choi and Kim (2010). In this respect, our model is similar to that of Hermalin and Katz (2007), who, however, consider a vertical differentiation model with endogenous entry of content providers.

For Hermalin and Katz, QoS tiering means that the ISP offers a menu of 'qualities' from which content providers, who differ in their marginal willingness to pay for quality, can choose. By contrast, under network neutrality the ISP is allowed to offer only one quality. Unlike in our model, in Hermalin and Katz (2007) CPs' revenue elasticity is not considered. Furthermore, the paper does not explicitly study investment incentives, nor considers the adverse effect that the high priority class exerts on the remaining best-effort class under fixed network capacity.

Economides and Tåg (2008), finally, consider a similar two-sided market model. On one side of the market, there is a continuum of non-competing content providers and, on the other side of the market, there is a continuum of consumers. Each side experiences positive network externalities through the presence of the other side. However, the authors do not consider a QoS tiering regime and instead see a violation of network neutrality in the ISP's practice of charging content providers a termination fee for access to its customers.

While each of these models considers important facets of the net neutrality debate, none has been capable of addressing all of the previously mentioned aspects together. Thus, it is not surprising that previous results with respect to innovation, investment and welfare are mixed: Hermalin and Katz (2007) find that network neutrality leads to less innovation and has a tendency to be welfare reducing. Choi and Kim (2010) and Cheng et al. (2010), on the other hand, show for a large range of parameters, that the ISP's incentive to invest in infrastructure is higher under network neutrality, whereas QoS tiering is generally welfare-enhancing in the short-run. By taking a more holistic approach that involves a two-sided market framework, queuing theory, endogenous entry/innovation and infrastructure investments, we make an attempt to provide a more complete picture of the differences between network neutrality regulation and QoS tiering. In particular, we affirm that QoS tiering is unambiguously welfare-enhancing in the short-run, while the long-run welfare consequences depend on the content providers' advertisement revenue elasticity with respect to an increase of active content providers in the market. The ISP will invest more in broadband infrastructure under QoS tiering if content providers' revenues are inelastic, but investments are still below the efficient level. On the contrary, the ISP has stronger incentives to invest in infrastructure under network neutrality if content provider's revenues are elastic. However, in this case the private investments are likely to be above the efficient level.

3 The Model

We model the Internet as a two-sided market, with content providers and Internet customers (ICs) on either side, each of which value an increasing presence of the other side and dislike network congestion. In order to be able to isolate the arguments of the net neutrality debate we abstract from the full complexity of the networks forming the Internet and consider a single monopolistic ISP providing access to consumers and content providers.

Content Providers We consider a continuum of CPs. Whatever service the CPs offer, they provide it for free and receive revenues only indirectly through online advertisements.[7] In the model, a CP's advertisement revenue will depend on the average received traffic, the per-click advertisement revenue, and her individual click-through-rate, which is determined by the CP's innate sensitivity towards network congestion. Before these measures are formally introduced below, we make one fundamental assumption:

Assumption 1. *Each content provider receives the same average traffic from each customer, denoted by λ. This is independent of a content provider's business model and consequently its innate sensitivity to network congestion.*

For the remainder of this article, it will often be convenient to think of λ as the number of 'clicks' that a customer generates on each content provider's website. This assumption provides a neutral reference case, where customers have no preference for specific content and rather distribute their clicks evenly among the available content providers. On the content provider's end, usually only a fraction of these clicks can be turned into advertisement revenue. This measure is known as the click-through rate. We assume that each CP's click-through rate diminishes as network congestion increases. Moreover, each CP's business model has an innate sensitivity as to what extent network congestion affects the click-through rate. This individual congestion sensitivity is denoted by θ and is uniformly distributed on the unit interval. In summary, the individual click-through rate of CP x is

$$(1 - \theta_x w) \tag{1}$$

---

[7] In particular, this means that we rule out the possibility that content providers charge consumers directly for access to their content. However, this seems to be the more relevant case as empirical evidence suggest that customers are generally fairly reluctant to pay extra for specific content or services (Dou, 2004; Sydell, 2007).

where w denotes the CP's perceived average level of network congestion.[8] If r is the average revenue-per-click on advertisements depending on the mass of active content providers in the market, then each CPs' profit under net neutrality is[9]

$$U_{CP}^{N}(\theta_x) = \begin{cases} \left(1 - \theta_x w^N\right) \lambda \bar{\eta} r & \text{if active} \\ 0 & \text{if inactive,} \end{cases} \tag{2}$$

where $\bar{\eta}$ denotes the share of Internet customers in equilibrium. Under network neutrality all content providers perceive the same level of congestion, w^N. In the QoS tiering regime, however, content providers can opt for the priority transmission class with $w_1^D < w^N$ at a price of p^D per click. The content providers that remain in the best-effort class, on the other hand, experience a higher congestion level $w_2^D > w^N$.

$$U_{CP}^{D}(\theta_x) = \begin{cases} \left(1 - \theta_x w_1^D\right) \lambda \bar{\eta} r - \lambda \bar{\eta} p^D & \text{if active in priority class} \\ \left(1 - \theta_x w_2^D\right) \lambda \bar{\eta} r & \text{if active in best-effort class} \\ 0 & \text{if inactive.} \end{cases} \tag{3}$$

The CP that is indifferent between choosing the priority and the best-effort transmission class under a QoS tiering regime, is denoted by $\tilde{\theta}^D$. Furthermore, in both regimes, the content provider that is indifferent between becoming active and staying out of the market is characterized by a congestion sensitivity of $\bar{\theta}$. Recall that θ is normalized to the unit interval, such that $\bar{\theta}$ also reflects the mass of all active content providers. Thus, the share of content providers choosing the priority class under a QoS tiering regime is given by

$$\beta \equiv 1 - \frac{\tilde{\theta}^D}{\bar{\theta}^D}. \tag{4}$$

---

[8] Note that the click-through rate follows a Poisson thinning process. The thinning probability depends on the average waiting time (w) as a proxy for congestion in a transmission class and the sensitivity of the service (θ) itself. Therefore a CP with a high innate sensitivity has a lower probability of making money than a CP with a low innate sensitivity at any given congestion level.

[9] Throughout this paper, we distinguish between the network neutral regime and the QoS tiering (differentiated services) regime by superscript N and D, respectively.

In order to introduce competition among content providers in the model, we assume that the level CPs' gross advertisement revenues depends on the mass of active content providers, i.e. $r(\bar{\theta})$, and that $\frac{\partial r(\bar{\theta})}{\partial \bar{\theta}} \leq 0$.

Finally, note the two-sided market property that Internet customers generate positive externalities on the CPs' profits.

Customers Likewise, Internet customers value basic connectedness to the Internet as well as the presence of many content providers. In particular, we assume that connectedness adds a base utility of b whereas each additional content provider adds a marginal utility of v to a customer's utility.[10] On the other hand, customers dislike waiting for content due to network congestion. This is captured through consumers' average waiting time, $\hat{w}$. To summarize, a customer's utility is given by

$$U_{IC} = \begin{cases} b + v\bar{\theta} - i\hat{w} - z\eta - a & \text{if connected} \\ 0 & \text{if not connected,} \end{cases} \tag{5}$$

where i denotes a consumer's marginal opportunity costs in time (a proxy for impatience), and a the access fee charged by the ISP.

As outlined before, in our analysis we intend to focus on the effect of QoS tiering on the relationship between CPs and the ISP. Thus, for expositional clarity we set $z = 0$, which makes customers homogeneous and allows the ISP to extract all consumer surplus, but maintains the two-sided market property (cf. Rochet and Tirole, 2006). In equilibrium the ISP therefore sets the customer access price to[11]

$$a^* = b + v\bar{\theta} - i\hat{w}. \tag{6}$$

We assume that b is sufficiently large, such that all consumers connect to the ISP in equilibrium, i.e.

$$\bar{\eta}^* = 1. \tag{7}$$

---

[10] Recall that content providers are atomistic and customers have no preference for specific content. We therefore avoid making any judgment about the value of specific content or service innovations to consumers. Instead we assume that consumers derive a utility of v from every content provider entering the market.

[11] Now and in the following, an asterisk denotes equilibrium values.

Network Congestion Network congestion is measured through the consumers' average waiting time following a content request. We employ a $M/M/1$ queuing model to fix ideas on the relationship between average waiting time, network traffic and capacity.[12] Under a network neutral regime the $M/M/1$ model predicts that each consumer has an expected average waiting time of

$$w^N = \frac{1}{\mu - \Lambda}. \tag{8}$$

Here μ represents the average rate at which service requests are handled, which is interpreted as the overall *transmission capacity*; whereas Λ denotes the average rate at which customers' aggregate content requests arrive at the ISP's network, which is interpreted as *network traffic.* More precisely,

$$\Lambda = \bar{\eta}^* \bar{\theta}^* \lambda, \tag{9}$$

i.e. network traffic will depend on the share of connected consumers, $\bar{\eta}$, and content providers, $\bar{\theta}$, in equilibrium, and the average traffic of each CP by each customer, λ.[13]

Under a QoS regime, content providers are offered the choice between a priority and a best-effort transmission class. In the $M/M/1$ model this translates to introducing an additional queue which handles the request of the content providers in the priority class and which is processed ahead of the queue for the best-effort class. In this vein, the classical results of the $M/M/1$ queuing model represent the average waiting time in the priority class, w_1^D, and the best-effort class, w_2^D:

$$w_1^D = \frac{1}{\mu - \beta\Lambda} \tag{10}$$

---

[12] The $M/M/1$ queuing model assumes that (1) service requests arrive according to a Poisson process (i.e. arrivals happen continuously and independently of one another), (2) service time is exponentially distributed (i.e. request coming from a Poisson process are handled at a constant average rate) and (3) that service requests are processed by a single server. Furthermore it is assumed that the length of the queue as well as the number of users is potentially infinite. This model is standard and considered to be a good proxy for actual Internet congestion.

[13] In other words, $\Lambda = \int_{y=0}^{\bar{\eta}^*} \int_{x=0}^{\bar{\theta}^*} \lambda \; dx \; dy$. Moreover, because $\bar{\eta}^* = 1$, network traffic reduces to $\Lambda = \bar{\theta}^* \lambda$ in the following.

$$w_2^D = \frac{\mu}{\mu - \Lambda} w_1^D \tag{11}$$

It is easy to see that relation

$$w_1^D < w^N < w_2^D \tag{12}$$

is always fulfilled, assuming a fixed transmission capacity μ and $\beta < 1$.[14] This is an important feature of our model, because it shows formally that serving some content providers with priority will (in the short-run) unambiguously lead to a *degradation* of service quality for the remaining content providers in the best-effort class. These content providers are therefore put at a disadvantage twice: first, through the prioritization of foreign traffic, and second, through the degradation of their own traffic, compared to the congestion level under network neutrality.

Moreover, notice that the customers' average waiting time is independent of the introduction of service classes,[15] because each customer will visit every available content provider equally. Customers' requests to CPs in the priority class will be processed within a time of w_1^D, whereas requests to content providers in the best-effort class take w_2^D units of time. Consequently, the average level of congestion in the network is

$$\hat{w} = \beta w_1^D + (1 - \beta) w_2^D = w_1^D \left(\frac{\mu - \beta\Lambda}{\mu - \Lambda} \right) = w^N, \tag{13}$$

since $w_1^D = \frac{1}{\mu - \beta\Lambda}$.

Internet Service Provider The ISP controls the two-sided market through a number of variables which he sets strategically. First, he charges an access fee, a, from connected consumers. Under a network neutral regime, the consumer access fee is the only source of revenue for the ISP. Second, in the long-run the ISP also sets the level of network capacity, μ. As outlined before, customers and CPs dislike network congestion. The level of network congestion is captured by customers' average waiting time for content, $\hat{w}$, which is again controlled by the ISP through its choice of network capacity. Hence, under a network neutrality regime, the ISP's profit is[16]

---

[14] For $\beta = 1$, when all content providers are in the first priority class, the model trivially collapses to $w_1^D = w^N$

[15] And also independent of β .

[16] Subscript s and l denote short-run and long-run incentives, respectively.

$$\Pi^N = \bar{\eta}^* a^* - c(\mu), \tag{14}$$

where $c(\mu)$ denotes the costs of capacity expansion.[17] Under a QoS tiering regime, the ISP has an additional strategic variable, p^D, the price which he charges CPs to transmit data packets with priority. The ISP will choose p^D in order to maximize its additional revenues from selling priority access. More precisely, under QoS tiering the ISP's profit function is

$$\Pi^D = \bar{\eta}^* a^* + \beta \Lambda p^{D^*} - c(\mu). \tag{15}$$

We consider the ISP's previous investment decisions into transmission capacity as sunk in all regimes. Therefore, in the short-run μ can be considered an exogenous variable which is irrelevant for profit maximization.

4 Short-Run Effects on Innovation and Welfare

First, we compare the two network regimes in the short-run, i.e. when network capacity, μ, is exogenous and equal in both regimes.

4.1 Short-run Equilibrium and Innovation

Network Neutrality Regime Recall that CPs are arranged on the unit interval in order of ascending congestion sensitivity, θ. Those CPs with θ close to zero offer a service with waiting time insensitive advertisement revenues, whereas those with values of θ close to one are very sensitive to network congestion. Under network neutrality all content providers expect the same congestion level of w^N and enter the network only if they have non-negative utility at this level. Consequently, the last content provider to enter the network is located at: [18]

$$\bar{\theta}^{N^*} = \frac{\mu}{\lambda + 1} \tag{16}$$

---

[17] To ensure the existence of an interior solution to the ISP's investment decision, we assume a non-concave cost function, i.e. $\frac{\partial c}{\partial \mu} \geq 0$ and $\frac{\partial^2 c}{\partial \mu^2} \geq 0$.

[18] We restrict our analysis to the interesting case where (at least) the most congestion sensitive content provider, located at $\theta = 1$, remains inactive in equilibrium. This is ensured iff the average congestion level does not drop below $\hat{w} > 1$.

It is immediately obvious that an increase in network traffic per content provider, λ, has an adverse effect on network congestion and content variety.

$$\frac{\partial w^N}{\partial \lambda} > 0 \qquad \text{and} \qquad \frac{\partial \bar{\theta}^N}{\partial \lambda} < 0$$

We label this the *traffic effect*, which is central to the debate on network neutrality, because it exemplifies the network operators' concerns with respect to the expected increase in traffic.

Next, recall that under a network neutral regime, the ISP makes profits only by selling access to consumers. The optimal access charge and hence ISP short-run profit is[19]

$$\Pi_s^{N^*} = a^{N^*} = b + v\frac{\mu}{\lambda+1} - i\frac{\lambda+1}{\mu}. \tag{17}$$

Quality of Service Tiering Regime In the QoS tiering regime, the ISP can alleviate congestion for the most congestion-sensitive content providers through the provision of differentiated transmission classes. According to (12), in the short-run, when transmission capacity is fixed, the remaining providers will be handled in the best-effort class, but at higher congestion levels than under network neutrality. We may now distinguish three types of content providers:

1. CPs whose business model is relatively insensitive to network congestion. They will remain in the free-of-charge best-effort class.
2. CPs whose business model is sufficiently sensitive to network congestion. They will opt for priority access at a price of p^D.
3. CPs whose business model is extremely sensitive to network congestion. They will be foreclosed from market entry and remain inactive.

The content provider indifferent between the first two cases is denoted by $\tilde{\theta}^D$, whereas the content provider indifferent between the last two cases is denoted by $\bar{\theta}^D$. Obviously, it must hold that $0 < \tilde{\theta}^D < \bar{\theta}^D$. Contrary to Choi and Kim (2010), but similar to Hermalin and Katz (2007), in our

---

[19] Subscript 's' denotes short-run equilibria.

model the 'high cost' content providers are more likely to opt for the priority class. There are two reasons for this. First, we do not model the competitive aspect of obtaining a larger market share based on the prioritized connection. Second, in our model the incentive to buy priority is based on the individual business model's innate need for a higher connection quality. In a fulfilled expectations equilibrium, the marginal content providers are

$$\bar{\theta}^{D^*} = \frac{\mu}{\lambda + 1} \tag{18}$$

$$\tilde{\theta}^{D^*} = \frac{p}{\lambda(r(\bar{\theta}^{D^*}) - p)} \bar{\theta}^{D^*}. \tag{19}$$

It is easy to see that an increase in the price for priority transmission, p^D, will shift the indifferent CP downward, and therefore results in a larger share of content providers in the priority class. Notice, however, that the share of active content providers, $\bar{\theta}^D$, is independent of p^D. Furthermore, comparing equation (18) with (16) directly reveals that

$$\bar{\theta}_s^{D^*} = \bar{\theta}_s^{N^*} = \frac{1}{\hat{w}}. \tag{20}$$

This proves our first main result.

Proposition 1 (Innovation). *In the short-run, QoS tiering has no effect on innovation. The number of active content providers does not change, as compared to network neutrality, independent of the price for priority access. In both regimes the number of active content providers is inversely proportional to the average level of congestion in the network.*

This result has important implications for the network neutrality debate. Under the present assumptions the claims of both parties are flawed in the short-run: *QoS tiering will lead to neither more nor less innovation.* What is even more surprising is the fact that the precise nature of the price for priority access is irrelevant. This is because under the realistic $M/M/1$ queuing model, the benefit from congestion alleviation in the priority class and the price for priority are held in check–at any price level. As will be seen below, the last content provider to enter must pay exactly that amount for priority access which she has gained through extra advertisement revenues from congestion alleviation.

Furthermore, recall from equation (12) that QoS tiering does not affect consumers' average waiting time. Thus, consumers' network utility is not changed compared to the network neutrality regime, and consequently, the ISP will charge the same amount for access in the short-run, independent of the underlying assumption about network traffic:

$$a_s^{D^*} = a_s^{N^*} \tag{21}$$

However, under a QoS tiering regime the ISP can additionally extract rents from content providers through sales of priority access. In the short-run, he will do so by maximizing $\Lambda\beta p^D$, which is achieved through

$$p^{D^*} = \left(1 - \sqrt{\frac{1}{\lambda+1}}\right) r(\bar{\theta}^{D^*}) = \left(1 - \frac{w_1^{D^*}}{\hat{w}}\right) r(\bar{\theta}^{D^*}) \tag{22}$$

Intuitively, this shows that the ISP can extract a fraction of the content providers' gross advertisement revenue $r(\bar{\theta}^{D^*})$, depending on the congestion alleviation to the priority class compared to the average congestion level in the network. Furthermore, recall that the price has no influence on the number of active CPs under the $M/M/1$ specification and thus the equilibrium price is independent of the customers' valuation for variety, v. Under the QoS tiering regime, the ISP will make an extra profit of

$$\Delta\Pi^D \equiv \Pi^D - \Pi^N = \Lambda\beta p^{D^*} > 0 \tag{23}$$

compared to the network neutrality regime.

Proposition 2 (ISP Preferred Regime)**.** *The ISP always prefers the QoS tiering regime, because it can make extra profits by selling a priority transmission service to content providers. Internet customers, however, pay the same price for network access as under the network neutrality regime.*

4.2 Short-Run Welfare Implications

Now we investigate the short-run effect of QoS tiering on welfare. Total welfare, W, is the sum of Internet customers' surplus, W_{IC}, content providers' surplus, W_{CP}, and the ISP's profit, Π. Thus, the difference in social surplus between QoS tiering and network neutrality is given by

$$\Delta W = \left(W_{IC}^{D} - W_{IC}^{N}\right) + \left(W_{CP}^{D} - W_{CP}^{N}\right) + \Delta\Pi^{D} \tag{24}$$

Recall that Internet customers' surplus is zero under both regimes and that $\Delta\Pi^{D} > 0$ according to Proposition 2. What remains to be examined is the short-run effect of QoS tiering on content providers' surplus.

[ENTER FIGURE 1 ABOUT HERE]

To this extent, consider Figure 1 and notice that those content providers located at $\theta \in [0, \tilde{\theta}^{D})$ are evidently worse off under a QoS tiering regime, because for them network congestion has increased from w^{N} to w_{2}^{D}. Second, the content providers' welfare loss increases with congestion sensitivity on the interval $\theta \in [0, \tilde{\theta}^{D})$. The business model of the provider located at $\theta = 0$ is not affected at all through congestion, while the provider at $\theta = \tilde{\theta}^{D}$ is already suffering so much that she is indifferent between staying in the best-effort class and buying priority access. Third, by the converse argument, notice that the welfare loss decreases for those CPs in the priority class as $\theta \in [\tilde{\theta}^{D}, \bar{\theta}^{N,D})$ increases. To see this, recall from Proposition 1 that the last content provider to enter the market, $\bar{\theta}^{N,D}$, is identical under both regimes and receives a surplus of zero. For her, the benefit through reduced congestion (compared to the network neutrality regime) is just offset by the price that she pays for priority access. Consequently, for all content providers with less congestion sensitivity, i.e. $\theta \in [\tilde{\theta}^{D}, \bar{\theta}^{N,D})$, the price that is paid for priority is higher than the benefit of being in the priority class. Nevertheless, by definition of $\tilde{\theta}^{D}$, for these providers the welfare loss is still less severe in the first priority class than in the best-effort class. In this line of argumentation, it is also obvious that content provider $\tilde{\theta}^{D}$ incurs the greatest welfare loss. In summary, we can conclude that in the short-run all active content providers are (weakly) worse off under a QoS tiering regime.

However, the price that CPs pay for priority access is merely a welfare shift to the ISP by a total of $\Delta\Pi^{D}$ (diagonally hatched area in Figure 1). The sign of the overall welfare effect, ΔW, will therefore only depend on the difference between the gross surplus gain through less congestion of those content providers in the priority class and the gross surplus loss through increased congestion of those providers remaining in the best-effort class.

$$
\begin{aligned}
\Delta W &= \Delta \Pi^D + \left(W_{CP}^D - W_{CP}^N\right) \\
&= \left(\bar{\theta} - \tilde{\theta}\right) \lambda p^D - \int_{\theta=\tilde{\theta}}^{\bar{\theta}} \lambda p^D d\theta + \\
&\qquad \lambda r(\bar{\theta}) \left(\int_0^{\bar{\theta}} \left(1 - w^N \theta\right)\ d\theta - \int_0^{\tilde{\theta}} \left(1 - w_2^D \theta\right)\ d\theta - \int_{\tilde{\theta}}^{\bar{\theta}} \left(1 - w_1^D \theta\right)\ d\theta \right) \\
&= \lambda r(\bar{\theta}) \left(\underbrace{\left(w^N - w_1^D\right) \int_{\tilde{\theta}}^{\bar{\theta}} \theta\ d\theta}_{\substack{\text{congestion alleviation} \\ \text{to priority class}}} \quad - \quad \underbrace{\left(w_2^D - w^N\right) \int_0^{\tilde{\theta}} \theta\ d\theta}_{\substack{\text{congestion aggravation} \\ \text{to best-effort class}}} \right) \qquad (25)
\end{aligned}
$$

Equation (25) reveals that the overall effect of QoS tiering on welfare depends on the relative size of the *congestion alleviation effect* to providers in the priority class versus the *congestion aggravation effect* to providers in the best-effort class. These effects relate directly to the main argument of proponents and opponents of net neutrality, respectively. Figure 2 exemplifies the relative size of these effects. Net of the price paid for priority, which is welfare neutral, content providers in the first priority class have higher surplus than under network neutrality (vertically hatched area). In comparison, for those content providers in the remaining best-effort class, some surplus is destroyed (horizontally hatched area).

[ENTER FIGURE 2 ABOUT HERE]

Proposition 3 (Short-run Welfare)**.** *In the short-run, QoS tiering unambiguously increases welfare with respect to the network neutrality regime, because congestion is alleviated for the most congestion sensitive content providers in lieu of the less congestion sensitive content providers. However, all content providers are worse off under a QoS tiering regime because the increased surplus is expropriated by the ISP.*

QoS tiering will therefore always be welfare improving in the short-run, because it 'allocates' congestion more efficiently. Those content providers who are relatively inelastic to congestion with respect to their advertisement revenues are allocated more congestion than the providers with relatively congestion elastic advertisement revenues.

5 Long-Run Effects on Broadband Investments and Welfare

In this section we extend the analysis of the model to long-run investments in network transmission capacity. Much of the neutrality debate is rooted in the ISPs' concerns about infrastructure investments. On the one hand, ISPs would like to accommodate new innovative content because this is valued by customers. However, on the other hand ISPs disapprove of content providers who free-ride on their infrastructure investments. QoS tiering seems to be a plausible way out of this dilemma, but it is unclear whether this regime will lead to more or less incentives for infrastructure investments than will network neutrality regulation in the long-run. In our model, transmission capacity is represented by the average service rate, μ, at which customer requests can be handled. An increase of μ allows the ISP to handle more service requests to content providers simultaneously.

5.1 Investment Incentives

Formally, the ISP's investment decision is a discrete decision stage which precedes the previous analysis. The ISP chooses the network capacity level, μ, first, and subsequently sets the customer access charge, a, and the priority price, p^D, if applicable. In the subgame perfect long-run equilibrium, the ISP will set the optimal capacity level at the point where the marginal revenues of capacity expansion, $MR \equiv \frac{\partial \Pi_l}{\partial \mu}$, equal marginal costs $MC \equiv \frac{\partial c(\mu)}{\partial \mu}$. Consequently, the ISP's optimal capacity level will be higher if marginal revenues from capacity expansion are higher.[20] In both network regimes the ISP makes revenues from customer access. More precisely, we can distinguish the following two marginal effects of capacity expansion on ISP revenue:

- The *variety incentive* $\left(v\frac{\partial \bar{\theta}}{\partial \mu}\right)$ denotes the ISP's marginal revenue effect from the entry of new, innovative content providers.
- The *congestion incentive* $\left(-i\frac{\partial \hat{w}}{\partial \mu}\right)$ describes the ISP's marginal revenue effect from a change of the overall congestion level.

---

[20] Thereby we assume that the ISP's marginal revenues with respect to μ are decreasing, while marginal costs are increasing. The former is shown in the appendix, whereas the latter is warranted by the assumption of a convex cost function.

Furthermore, notice that these investment incentives are always positive and identical under both network regimes. Hence, potential differences in investments between the two regimes may only be a result of the additional incentive that an ISP has under QoS tiering:

- The *priority revenue incentive* $\left(\frac{\partial \Delta \Pi^D}{\partial \mu}\right)$ denotes the ISP's marginal revenue effect from selling priority access.

Consequently, the sign of the priority revenue incentive is definitive for the comparison between investment incentives under QoS tiering and network neutrality.

Proposition 4 (Investment Incentives)**.** *The ISP's optimal capacity level is higher under QoS tiering if content providers' gross advertisement revenues are inelastic with respect to an increase of active content providers in the market (*$\varepsilon^r > -1$*). On the contrary, if the content providers' gross advertisement revenues are elastic (*$\varepsilon^r < -1$*), the ISP has higher investment incentives under network neutrality. At unit elasticity* $\varepsilon^r = -1$ *the ISP's investment levels coincide under both regimes.*

5.2 Innovation at the Edge and Long-Run Welfare

The ISP's investments into network infrastructure have direct ramifications for welfare. At higher capacity levels customers enjoy lower network congestion (congestion incentive) and the entry of new content providers (variety incentive). Figure 3 illustrates the effect of capacity expansion for content providers under QoS tiering. The reduction of network congestion increases content providers' click-through rate and thus the slope of their surplus curve in both transmission classes. The CPs in the best-effort class and also some CPs in the first priority class may still be worse off than under network neutrality. However, as a consequence of the overall decreased congestion level, both marginal content providers $\tilde{\theta}^D$ and $\bar{\theta}^D$ are shifted to the right. This means that new, congestion sensitive content providers are able to enter the network. This is often termed 'innovation at the edge', because the services of these content providers were not previously available. Obviously the surplus of the new CPs (crosswise hatched area), but also the surplus of some of the previously most congestion sensitive CPs (vertically hatched area), are thus increased compared with a network neutrality regime.

[ENTER FIGURE 3 ABOUT HERE]

Accordingly, higher capacity levels will ceteris paribus lead to higher gross utility for Internet customers and content providers.

QoS tiering will therefore be the efficient regime if CPs' ad revenues are inelastic. In this case, QoS tiering leads to more welfare in the long-run and, by Proposition 3, also in the short-run. However, a non-negligible share of this surplus is immediately expropriated by the ISP. On the other hand, if CPs' ad revenues are elastic, the network neutrality regime will provide higher investment incentives. In this case, the choice of the socially preferred network regime depends on the trade-off between long-run and short-run welfare effects. While the QoS tiering regime increases welfare in the short-run through better allocation of network congestion, network neutrality will reduce the average congestion level in the long-run and thereby lead to more innovation. The precise nature of the trade-off between the two regimes depends on the costs of capacity expansion. It is likely, however, that in this case network neutrality will eventually lead to higher overall welfare. Proposition 5 summarizes the results.

Proposition 5 (Long-Run Welfare)**.** *The regime that provides more incentives for infrastructure investments is more efficient in the long-run. If investment incentives are higher under network neutrality, the long-run welfare gain must be weighed against the short-run welfare loss compared to QoS tiering.*
If QoS tiering provides more investment incentives, it is unambiguously the more efficient regime. However, in this case the ISP is able to expropriate much of the content providers' surplus through the price for priority transmission. Nevertheless, the most congestion sensitive content providers will be better off than under network neutrality.

6 Regulatory Implications

We conclude our analysis with an investigation of the suitability of two prominent policy instruments in this domain: price regulation and minimum quality standards. Price regulation seems to be a promising policy tool with respect to controlling the ISP's expropriation of content providers' surplus, and is thus targeted at short-term regulation. On the other hand, minimum quality standards are targeted at infrastructure regulation and have recently been proposed to mitigate the congestion aggravation effect to the content providers in the best-effort class under QoS tiering.

6.1 Price Regulation

Although QoS tiering is the more efficient regime in the short-run, policy makers may be suspicious of the fact that much of the social surplus is accommodated by the ISP. It seems logical at first to regulate the revenue stream between content providers and the ISP through price controls. The regulated price should thus be set to maximize the social surplus under QoS tiering, W^D. In particular, the social planner's and ISP's maximization problems differ by the term

$$W^D - \Pi^D = \frac{\hat{w} - w_1^D(p)}{\hat{w} \quad w_1^D(p)}. \tag{26}$$

In other words, the social planner seeks to set the regulated price such that the optimal share of content providers selects the priority transmission class. In this vein content providers' gross surplus is maximized. However, the ISP pursues the same goal, because he can subsequently extract a fraction of this surplus.

Proposition 6 (Price Regulation)**.** *A welfare maximizing social planner would select the same price for priority transmission as the revenue maximizing ISP. Therefore, regulating the price below the ISP's price level will inevitably lead to welfare reductions.*

To provide further intuition for this result, assume that a regulatory agency would lower the price for prioritized traffic under the level which the ISP has chosen according to (22). Consequently, more content providers would subscribe to the priority transmission service,[21] and in turn increase the average waiting time in the priority class. However, because the priority traffic is handled before the best-effort traffic, congestion is also increased for content providers in the remaining best-effort class. Nevertheless, the average congestion level remains unchanged because the content providers that have switched to the priority class experience a lower congestion level than before. But overall, content providers' gross surplus is reduced. Of course,

---

[21] Remember that a price reduction does not result in the entry or exit of content providers.

price regulation would indeed also shift some of the surplus back from the ISP to the content providers. Unfortunately, this effort proportionally destroys welfare, as long as the ISP's and CPs' surplus are not considered different. In the most extreme case, regulation could impose a zero-pricing-rule which would in fact reestablish the network neutrality regime.

6.2 Infrastructure Regulation

Price controls are not a suitable policy instrument to correct for potential welfare distortions, because in the short-run social and private incentives are in line. This is generally not true for the ISP's long-run incentives to invest in infrastructure. Opponents of net neutrality regulation have often objected that this regime forces ISPs to invest above the efficient level, which is known as overprovisioning. On the contrary, opponents of QoS tiering argue that this regime induces ISPs to keep transmission capacity scarce, and thus broadband investments are generally below the efficient level (underprovisioning). We can show that the difference between the efficient and private level of infrastructure investments is in fact independent of the network regime, but depends again on the elasticity of content providers' revenues.

Proposition 7 (Infrastructure Regulation)**.** *The social planner has a higher incentive to invest in network capacity than the ISP, if content providers' gross advertisement revenues are inelastic with respect to an increase of active content providers in the market ($\varepsilon^r > -1$). On the contrary, if content providers' ad revenues are elastic ($\varepsilon^r < -1$), the ISP provides network capacity above the efficient level. This result holds for both network regimes, QoS tiering and network neutrality.*

The wit of Proposition 7 is that the arguments of either side are to some extent right. When network neutrality provides higher investment incentives ($\varepsilon^r < -1$), and is thus more efficient in the long-run, the ISP will invest too much in network infrastructure (overprovisioning). On the other hand, if QoS tiering is the socially preferred regime ($\varepsilon^r > -1$), private investments will be too low (underprovisioning). Especially in the latter case, policy makers are faced with a dilemma. Even though QoS tiering is the socially preferred regime, the ISP is able to expropriate much of the social surplus and private investments are still below the efficient level. Clearly, additional efforts to subsidize the ISP's broadband investments are not politically feasible and thus alternative policy instruments must be considered.

Minimum Quality Standards It has recently been argued that a minimum quality standard (MQS) could be an appropriate policy instrument in this context (Brennan, 2010). After all, MQSs have found to be generally welfare-enhancing in competitive settings (Ronnen, 1991). For example, here the MQS could be set such that the ISP is required to offer content providers under QoS tiering a congestion level in the best-effort class that is as least as low as the equilibrium best-effort congestion level under network neutrality. Consequently, under the QoS tiering regime no content provider would be set at a disadvantage anymore. Moreover, in order to meet the MQS, the ISP is required to increase the network's capacity, potentially to the extent that the gap between the level of private and efficient investments is closed. More precisely, by requiring the MQS $w^N(\mu^{N^*}) \equiv w_2^D(\mu^{MQS})$ the regulator implicitly defines the new capacity level $\mu^{MQS} > \mu^{N^*}$.

First, consider the case where $\varepsilon^r < -1$. By Proposition 4 we know that the ISP's former privately optimal capacity level yields $\mu^{D^*} < \mu^{N^*}$. Moreover, by Proposition 7 it is known that the equilibrium capacity is above the efficient level, $\mu^{D^*} > \mu^{D^{**}}$. Taken together, it follows that

$$\mu^{MQS} > \mu^{N^*} > \mu^{D^*} > \mu^{D^{**}}.$$

Consequently, in this case the imposed MQS would lead to a further increase of the ISP's already excessive network capacity, well above the efficient level $\mu^{D^{**}}$.

Now assume $\varepsilon^r > -1$. Here the order of relevant capacity levels is

$$\mu^{D^{**}} > \mu^{D^*} > \mu^{N^*}.$$

Remember that $\mu^{MQS} > \mu^{N^*}$, and thus we can differentiate between three different cases. First, if $\mu^{D^*} \geq \mu^{MQS}$, the MQS is not a binding condition for the ISP's capacity choice an hence is simply ineffective. Second, if $\mu^{D^{**}} \geq \mu^{MQS} > \mu^{D^*}$ the MQS is effective in raising the ISP's network capacity level, potentially up to the efficient level. Third, if $\mu^{MQS} > \mu^{D^{**}}$ the MQS policy may again lead to an excessive investment in network infrastructure.

In summary, MQS are only effective in one out of three cases when content providers' ad revenues are inelastic, and never when ad revenues are elastic.

Proposition 8 (Minimum Quality Standard Regulation)**.** *An MQS policy, which requires the ISP to guarantee a best-effort congestion level under QoS* *tiering which is equal to the equilibrium congestion level under network neutrality, results in excessive infrastructure investments when content providers' revenues are elastic. When content providers' revenues are inelastic, QoS tiering will already lead to more infrastructure investments than network neutrality. An additional MQS policy may still increase welfare, but may also lead to excessive investments or be ineffective.*

7 Strategic Quality Degradation

In the preceding analysis we have neglected the possibility that the monopolistic ISP may also engage in non-price discrimination, for example by degrading the quality of the best-effort class under QoS tiering. The concern for strategic quality degradation under a QoS tiering regime has been expressed by network neutrality proponents, but also previous empirical and theoretical research has identified several circumstances under which such practice is indeed profitable (Economides, 1998; Foros et al., 2002; Crawford and Shum, 2007). Absent the possibility to degrade the quality of the best-effort class, the ISP's only control over the share of CPs that buy priority transmission in equilibrium (β) is through the price p^D. In this case, as has been shown in Section 6, the ISP will choose the efficient price which maximizes CPs' gross surplus. Quality degradation, however, provides the ISP with an additional means through which it can manipulate the relative attractiveness of the priority class over the best-effort class and thus the mass of CPs that buy priority transmission in equilibrium. It is inevitable that such practice will destroy some CPs' surplus and therefore questions the previously positive welfare results of QoS tiering. However, ex-ante it is not clear whether there exist scenarios under which quality degradation may actually be profitable to the ISP in the first place.

To this end, consider the extreme scenario where the ISP degrades the best-effort class under a QoS tiering regime maximally ($w_2^d \to \infty$), such that in equilibrium no content provider wants to remain in the best-effort class ($\tilde{\theta}^{d^*} \to 0, \beta^{d^*} \to 1$).[22] Furthermore, let $r(\overline{\theta}) = r$ be constant in this example.[23] We will show that there exist circumstances under which even

---

[22] Superscript 'd' denotes the QoS tiering regime with maximum quality degradation.

[23] This does not affect the generality of the existence of settings in which the ISP prefers to degrade the best-effort class under QoS tiering. In fact, as will be readily seen later, assuming r to be constant, is the most conservative assumption one can make in this

this extreme form of quality degradation is profitable. More precisely, by rendering the best-effort class useless, the ISP effectively forces all content providers into the priority transmission class. It is easy to see that this is equivalent to a scenario in which the ISP demands a termination fee from each content provider for transmitting content to its connected consumers.[24] At a fixed transmission capacity, this has detrimental effects on innovation and welfare.

To see this, recall that without quality degradation, the last content provider to enter the network was located at $\overline{\theta}^{D,N} = \frac{1}{\hat{w}}$, independent of the network regime and independent of the price for priority transmission. In contrast, the last content provider to enter under quality degradation is located at

$$\overline{\theta}^{d} = \frac{1 - \frac{p^d}{r}}{w_1^d}. \tag{27}$$

Since all content providers are forced into the priority transmission class, congestion is the same for all content providers and at a similar level than under network neutrality. Thus, maximum quality degradation not only destroys the source of the positive welfare effects of the QoS tiering regime, but also forces the most congestion sensitive content providers out of the network: CPs experience a similar congestion level as under network neutrality, but have to pay a price $p^d > 0$ as if they were under QoS tiering. While under the $M/M/1$ queuing assumption, the congestion alleviation and the price for priority were held in check, such that the level of innovation was in fact independent of the price level, this is no longer true under quality degradation. The higher the price for priority, the less content providers will enter the network. To be precise, it must be mentioned that the smaller mass of active content providers will also slightly reduce the average congestion level compared to network neutrality or QoS tiering, i.e. $w_1^d < \hat{w}$. However, this type of congestion alleviation effect cannot outweigh the detrimental effect to innovation and welfare.

---

context.

[24] Consequently the analysis in this section bridges the gap between the formal strand of the literature that considers net neutrality as a zero-price rule (i.e. no termination fees) and the literature that associates net neutrality with a non-discrimination rule (see. cf. Schuett, 2010).

Proposition 9 (Innovation and Welfare under Quality Degradation)**.** *When the ISP degrades the quality of the best-effort class under QoS tiering such that all content providers choose to buy priority transmission, then, compared to network neutrality, less content providers enter the network in equilibrium and overall welfare is lower.*

Consequently, quality degradation is undesirable from a policy perspective and tarnishes the short-run welfare results of QoS tiering. The question remains, however, whether quality degradation is in fact a profitable option to the ISP under QoS tiering and thus constitutes an actual source of concern to policy makers.

The effect of quality degradation on the ISP's revenue depends on the trade-off of two opposing effects. By Proposition 9 quality degradation results in less content variety and consequently the ISP can charge consumers less for access. On the other hand, quality degradation forces *all* content providers to pay for their traffic and thus revenues from priority sales are potentially larger than before. Obviously, this trade-off is driven by the relative size of the marginal valuations of consumers and content providers, respectively. This can be exemplified by the equilibrium price formula:

$$p^{d^*} = \frac{r(1+\lambda) - \sqrt{r\left(v + r(1+\lambda)\right)}}{\lambda}. \tag{28}$$

First notice that p^d depends on v, as one would expect from a two-sided market model, because the ISP takes into account the reduction in variety that would result from an increase of the price.[25] Prices are positive as long as $v < r\lambda(1+\lambda)$, i.e. as long as the consumers' marginal utility for variety is not too large with respect to the CPs' marginal valuation for gross traffic (generated by consumers). On the contrary, if $v > r\lambda(1+\lambda)$, the ISP would theoretically like to subsidize the content providers and thus promote innovation in order to extract consumers' high utility from variety.

Proposition 10 (Profitability of Quality Degradation)**.** *For all $v < \underline{v}$, where $\underline{v} < r\lambda(1+\lambda), \forall\lambda > 0$, the ISP makes larger profits under a QoS tiering regime in which the transmission quality of the best-effort class is degraded, such that all content providers choose to buy priority transmission in comparison to a QoS tiering regime without quality degradation.*

---

[25] Recall that under the $M/M/1$ specification the price was irrelevant for innovation and thus v did not enter the pricing formula.

Proposition 10 establishes first that the ISP never subsidizes content providers by imposing negative prices under maximum quality degradation, but rather prefers to refrain from quality degradation and revert to the untempered QoS tiering regime instead. This also implies that the ISP will not privately establish a network neutrality regime, which could be the result of QoS tiering with maximum quality degradation and a price of $p^d = 0$. Secondly, and more importantly, the proposition highlights that strategic quality degradation is in fact a profitable strategy for the ISP as long as consumers' marginal valuation for variety is sufficiently small.[26]

Given the detrimental welfare consequences of quality degradation under a QoS tiering regime, policy makers should be aware of this strategic option. In particular, if policy makers suspect the ISP to engage in quality degradation, some of the previously reviewed policy instruments may now regain attention. As mentioned before, price regulation (i.e. $p^d = 0$) can at least ensure the current status quo of the network neutrality regime. However, such regulation also excludes the potentially positive welfare effects of an untempered QoS tiering regime. In this context, minimum quality standards and transparency obligations seem to provide a more appropriate policy tool. If applied effectively, such obligations can preclude the ISP's negative strategic incentives under QoS tiering, while maintaining the generally positive welfare effects of this regime; after all, the ISP is still left better off than under network neutrality.

8 Conclusions

Network neutrality has become a prime topic for many regulatory authorities, but the effect of such regulation is still unclear. Scholarly papers often find contradictory results with respect to the consequences of network neutrality on service innovation, broadband investments and welfare. We contribute to the ongoing debate on network neutrality by providing a formal framework that incorporates the relevant arguments of net neutrality proponents and opponents in a two-sided market framework with Internet customers, content providers and an Internet service provider. Our analysis focuses on the

---

[26] If instead we would have assumed again that $\frac{\partial r(\bar{\theta})}{\partial \theta} < 0$, then the profitability of quality degradation would even be increased. This is because the aggregate loss in CPs' gross advertisement revenues that is caused by the reduction in content variety according to Proposition 9, would be less pronounced.

relationship between content providers and the ISP, and compares network neutrality to a Quality-of-Service tiering regime, in which content providers may pay for the prioritized transmission of their data packets. We explicitly consider the negative externality that prioritization has on the remaining best-effort class, but acknowledge that content providers' services differ in their sensitivity toward network congestion, and may offer their services only if they are sustainable under the given congestion level.

We find that QoS tiering increases welfare in the short-run because the installed level of network capacity is used more efficiently. Network congestion is alleviated for the most congestion sensitive content providers and this effect offsets the congestion aggravation for the content providers in the remaining best-effort class. However, QoS tiering does not immediately promote the entry of new content provides with innovative services that are even more congestion sensitive. In fact, in the short-run, all content providers are worse off under a QoS tiering regime, because the ISP is able to expropriate some of the content providers' surplus through priority pricing. Consequently, the ISP always prefers the QoS tiering regime. It is subject to the authority of policy makers to evaluate the shift of surplus from content providers to ISPs, which is welfare neutral per se, but lies at the heart of the net neutrality debate. On the one hand, those content providers with intermediate congestion sensitivity (who are indifferent between buying priority and remaining in the best-effort class) will especially suffer from a switch to QoS tiering. On the other hand, ISPs argue that they will use the additional revenues to invest more in broadband infrastructure. We show that this is only true if competition among content providers is weak, i.e. if content providers' gross advertisement revenues are inelastic with respect to an increase of active content providers in the market. In this case, the ISP will indeed invest more in broadband infrastructure than under network neutrality. On the contrary, if competition among content providers is strong (gross advertisement revenues are elastic), network neutrality regulation will eventually lead to more private infrastructure investments.

Accordingly, QoS tiering is unambiguously the more efficient regime when content providers' revenues are inelastic: the ISP invests more into broadband infrastructure, lowers the overall congestion level in the long-run and thereby encourages entry of new content providers with highly congestion sensitive services. The most congestion sensitive content providers will therefore be better off under QoS tiering, and hence it is not surprising that Google and Verizon have privately agreed on a tiered system (Wyatt, 2010).[27] However, when content providers' revenues are elastic, investment incentives are higher under network neutrality. In this case the long-term efficiency gains of net neutrality must be weighed against its short-term inefficiencies.

Furthermore, our analysis reveals that the level of private investments is generally not efficient. When network neutrality is the socially preferred regime, investments are too high (overprovisioning), whereas investments are too low (underprovisioning) when QoS tiering is the more efficient regime. In this sense, our analysis unifies both sides of the debate, but presents policy makers with a dilemma; and we show that there is no 'easy way' out of this dilemma: although price regulation can shift some of the congestion alleviation gains back to content providers, it is unsuitable as a policy instrument, because welfare is proportionally destroyed in the process. Therefore, if regulatory correction is desired, it should address the regulation of broadband infrastructure more directly. Certainly, in the light of the ISP's additional revenues under QoS tiering, it seems politically impossible to further subsidize the ISP's infrastructure investments. Minimum quality standards have recently been proposed as a more appropriate policy instrument. However, we show that a policy that requires the ISP to guarantee a congestion level in the best-effort class under QoS tiering which is at least as good as the best-effort congestion level under network neutrality is likely to induce excessive broadband investments. Yet future research is needed to find more pertinent ways to regulate broadband infrastructure. One promising possibility that we have not addressed here is to mitigate quality degradation and to stimulate broadband investments through the infrastructure-based competition among ISPs. This, however, is subject to the perennial effort of regulatory authorities.

The caveat is that the previous results hold only as long as the ISP has no incentive to strategically degrade the quality of the best-effort class. Strategic quality degradation is detrimental to innovation and welfare, but leaves the ISP with higher profits if consumers' marginal utility for content variety is small. It can possibly be counteracted, however, by Internet transparency obligations, or by setting a minimum quality standard for the best-effort class. In the latter case, our results advise that the minimum standard

---

[27] Interestingly, Google CEO Eric Schmidt argues that such an agreement would be in line with net neutrality, because it does not discriminate against specific content providers (Fehrenbacher, 2010).

should not be set too high (i.e. not necessarily at the congestion level of the best-effort class under network neutrality), but for instance at a congestion level that would have emerged for the best-effort class in an untempered QoS tiering regime. With respect to transparency obligations, it is noticeable that FCC chairman Julius Genachowski has recently proposed the amendment of such a new principle for Internet policy–together with an accompanying non-discrimination principle (Genachowski, 2009). While we find strong support for the transparency principle, the non-discrimination principle, on the other hand, can potentially be harmful if it prevents ISPs from offering QoS tiering on a non-discriminatory basis.

References

Armstrong, M. (2006). Competition in two-sided markets. *The RAND Journal of Economics 37*(3), 668–691.

Brennan, T. (2010, June). Net neutrality or minimum quality standards: Network effects vs. market power justifications. mimeo, Social Science Research Network.

Cerf, V. G. (2006, February 7). Hearing on Network Neutrality. Testimony before the United States Senate, Committee on Commerce, Science and Transportation.

Cheng, H. K., S. Bandyopadhyay, and H. Guo (2010). The Debate on Net Neutrality: A Policy Perspective. *Information Systems Research*, isre.1090.0257.

Choi, J. P. and B.-C. Kim (2010). Net neutrality and investment incentives. *RAND Journal of Economics 41*(3), 446–471.

Crawford, G. and M. Shum (2007). Monopoly quality degradation and regulation in cable television. *The Journal of Law and Economics 50*, 181–219.

Crowcroft, J. (2007). Net neutrality: The technical side of the debate - a white paper. *International Journal of Communication 1*, 567–579.

Dou, W. (2004). Will internet users pay for online content? *Journal of Advertising Research 44*(4), 349–359.

Economides, N. (1998). The incentive for non-price discrimination by an input monopolist. *International Journal of Industrial Organization 16*(3), 271–284.

Economides, N. and J. Tåg (2008). Net neutrality on the internet: A two-sided market analysis. mimeo, University of New York, School of Law.

Fehrenbacher, K. (2010, August 5). Caught on video: Google ceo dishes on google wave, verizon & social strategy. Gigaom.com.

Foros, Ø., H. Kind, and L. Sørgard (2002). Access pricing, quality degradation, and foreclosure in the Internet. *Journal of Regulatory Economics 22*(1), 59–83.

Genachowski, J. (2009, September 21). Preserving a free and open internet: A platform for innovation, opportunity, and prosperity. Speech at The Brookings Institution, Washington DC.

Hahn, R., R. Litan, and H. Singer (2007). The economics of wireless net neutrality. *Journal of Competition Law and Economics 3*, 399–451.

Hahn, R. and S. Wallsten (2006). The economics of net neutrality. *The Berkeley Economic Press - Economists' Voice 3(6)*, 1–7.

Hermalin, B. and M. Katz (2007). The economics of product-line restrictions with an application to the network neutrality debate. *Information Economics and Policy 19*, 215–248.

Lambert, P. (2010, February 24). Vodafone and telefonica are overplaying their hand with google. Telecoms.com. last accessed on 12/08/2010.

Lee, R. and T. Wu (2009). Subsidizing creativity through network design: Zero-pricing and net neutrality. *The Journal of Economic Perspectives 23*(3), 61–76.

Lessig, L. (2001). *The future of ideas.* Random House New York.

Lessig, L. (2002, October 1). The Government's Role in Promoting the Future of Telecommunications Industry and Broadband Deployment. Testimony before the United States Senate, Committee on Commerce, Science and Transportation.

O'Connell, P. (2005, November 7). At sbc, it's all about "'scale and scope"'. Businessweek.

Owen, B. and G. Rosston (2006). Local broadband access: Primum non nocere or primum processi? a property rights approach. In T. Lenard and R. May (Eds.), *Net Neutrality or Net Neutering: Should Broadband Internet Services be Regulated*, pp. 163–194. Springer US.

Rochet, J. and J. Tirole (2006). Two-sided markets: A progress report. *The RAND Journal of Economics 37*(3), 645–667.

Ronnen, U. (1991). Minimum quality standards, fixed costs, and competition. *The RAND Journal of Economics 22*(4), 490–504.

Schneibel, G. and C. Farivar (2010, April 7). Deutsche telekom moves against apple, google and net neutrality. Deutsche Welle.

Schuett, F. (2010). Network neutrality: A survey of the economic literature. *Review of Network Economics 9*(2), Article 1.

Sharma, P. (2010, August 9). Winners, losers from the new net neutrality. The Source - Wall Street Journal.

Sidak, G. J. (2006a, February 7). Hearing on Network Neutrality. Testimony before the United States Senate, Committee on Commerce, Science and Transportation.

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

Sydell, L. (2006, April 25). Internet debate - preserving user parity. "All Things Considered" - National Public Radio, USA.

Sydell, L. (2007, September 19). Firms abandon online subscription plans. "All Things Considered" - National Public Radio, USA.

Van Schewick, B. (2006). Towards an economic framework for network neutrality regulation. *Journal on Telecommunications & High Technology Law 5*, 329.

Wu, T. (2003). Network neutrality, broadband discrimination. *Journal on Telecommunications & High Technology Law 2*, 141.

Wu, T. and C. Yoo (2007). Keeping the Internet Neutral?: Tim Wu and Christopher Yoo Debate. *Federal Communications Law Journal 59*(3), 575–592.

Wyatt, E. (2010, August 4). Google and verizon near deal on web pay tiers. New York Times.

Yoo, C. (2005). Beyond network neutrality. *Harvard Journal of Law & Technology 19*, 1–77.

Appendix

Proof of Proposition 2

Proof.

$$\Delta\Pi^D = \Lambda\beta p^{D^*} = \mu r(\bar{\theta}^{D^*})\left(1 + \frac{1}{1+\lambda} - \frac{2}{\sqrt{1+\lambda}}\right),$$

which is always greater than zero for $\mu, r(\bar{\theta}^{D^*}), \lambda > 0$. □

Proof of Proposition 3

Proof. From (25) it is clear that

$$\Delta W = \frac{\lambda r(\bar{\theta})}{2}\left(\left(w^N - w_1^D\right)\left(\bar{\theta}^2 - \tilde{\theta}^2\right) - \left(w_2^D - w^N\right)\tilde{\theta}^2\right).$$

Thus,

$$\Delta W > 0 \Leftrightarrow \frac{w^N - w_1^D}{w_2^D - wN} > \frac{\tilde{\theta}^2}{\bar{\theta}^2 - \tilde{\theta}^2} \Leftrightarrow \frac{1-\beta}{\beta} > \frac{(1-\beta)^2}{1-(1-\beta)^2} \Leftrightarrow 0 < \beta < 1$$

□

Proof of Proposition 4

Proof. Incentives to invest into network capacity are higher under QoS tiering iff marginal revenues from priority sales are greater than zero, provided that the ISP revenues are concave, and the costs of capacity expansion convex in μ. The latter is warranted by assumption. To ensure that the ISP's revenues are concave the property $\frac{\partial^2\Pi^D}{\partial\mu^2} \leq 0$ has to be fulfilled. It is easy to verify, that this is always true, if ad-revenues are constant or concave and monotonic decreasing. If ad-revenues are convex and monotonic decreasing, consider the following derivative.

$$\frac{\partial^2\Pi^D}{\partial\mu^2} = -\frac{i\,(1+\lambda)}{\mu^3} + \underbrace{\left[\frac{\partial^2 r(\bar{\theta})}{\partial\bar{\theta}^2}\frac{\bar{\theta}}{2} + \frac{\partial r(\bar{\theta})}{\partial\bar{\theta}}\right]}_{A}\frac{\partial\bar{\theta}}{\mu}\underbrace{\left(1 + \frac{1}{(1+\lambda)} - \frac{2}{\sqrt{1+\lambda}}\right)}_{B}$$

Since $B \geq 0$ always holds, the ISP revenues are concave if

$$\frac{\partial^2 \Pi^D}{\partial \mu^2} \leq 0 \begin{cases} A \leq 0 & \text{always} \\ A > 0 & \text{if} \quad \frac{i(1+\lambda)^2}{\mu^3 B} \geq A \end{cases}$$

Therefore we have to assume, that the condition in the second case holds. Now consider $\Delta\Pi^D$ from (23). Differentiating with respect to μ yields

$$\frac{\partial \Delta \Pi^D}{\partial \mu} = \frac{\sqrt{\lambda+1}\left((\lambda+1) - \sqrt{\lambda+1}\right)}{(\lambda+1)\sqrt{\lambda+1}} \left[\frac{\partial r(\bar{\theta}^D)}{\partial \bar{\theta}^D} \frac{\partial \bar{\theta}^D}{\partial \mu} \mu + r(\bar{\theta}^D)\right]$$

The sign of the derivative is determined by the part in square brackets. Consequently,

$$\frac{\partial \Delta \Pi^D}{\partial \mu} \lesseqgtr 0 \Leftrightarrow \varepsilon^r = \frac{\partial r(\bar{\theta}^D)}{\partial \bar{\theta}} \frac{\bar{\theta}}{r(\bar{\theta}^D)} \lesseqgtr -1$$

$\square$

Proof of Proposition 5

Proof. To see that the overall congestion level, $\hat{w}$ decreases with capacity expansion, we show that $\frac{\partial \hat{w}}{\partial \mu} = \frac{\partial \frac{1}{\mu - \Lambda}}{\partial \mu} < 0$: Notice that $\Lambda = \bar{\theta}^* \lambda = \frac{\lambda \mu}{\lambda+1}$, so that $\frac{\partial \Lambda}{\partial \mu} = \frac{\lambda}{\lambda+1} < 1$. Therefore, it holds that $\frac{\partial(\mu-\Lambda)}{\partial \mu} > 0$ and consequently, $\frac{\partial \frac{1}{\mu-\Lambda}}{\partial \mu} < 0$. By equation (3) and (5) it is immediately obvious that the gross utility of customers and content providers increases as the congestion level decreases. The homogeneity of customers allows the ISP to fully expropriate the additional customer utility. Capacity expansion also increases the amount of active content providers, since $\frac{\partial \bar{\theta}}{\partial \mu} > 0$ by equation (18). Before the capacity expansion occurred, these content providers had a surplus of zero and are therefore unambiguously better off.

Additionally some content providers already active in the market are better off. To determine the critical content provider one has to solve the following equation:

$$\left(1 - \theta_x w^N\right) \lambda r(\bar{\theta}^D) = \left(1 - \theta_x w_1^D\right) \lambda r(\bar{\theta}^D) - \lambda p \tag{29}$$

$$\breve{\theta} = \frac{\left(\hat{w} - w_1^D\right)}{\left(w_N - w_1^D\right) \hat{w}} \tag{30}$$

Therefore all content providers in the interval $[\breve{\theta}, \bar{\theta}]$ are better of, compared to the network neutral regime.

$\square$

Proof of Proposition 6

Proof. The social extra term

$$W^D - \Pi^D = \frac{\mu p \left(\lambda \left(r(\bar{\theta}^D) - p\right)\right)}{(\lambda + 1)\left(r(\bar{\theta}^D) - p\right)} \tag{31}$$

is maximized by

$$p^{D^{**}} = \left(1 - \sqrt{\frac{1}{\lambda + 1}}\right) r(\bar{\theta}^{D^{**}}), \tag{32}$$

which is identical to the price p^{D^*} from equation (22) that maximizes the ISP's profit Π^D. $\square$

Proof of Proposition 7

Proof. We consider each regime separately and show that the conditions with respect to efficient investments coincide. First, we derive the conditions for which $\frac{\partial(W^N - \Pi^N)}{\partial\mu}$ is larger/equal/smaller zero:

$$\begin{aligned} W^N - \Pi^N &= \frac{\lambda}{2(\lambda+1)}\mu r(\bar{\theta}^D) \\ \frac{\partial\left(W^N - \Pi^N\right)}{\partial\mu} &\lesseqqgtr 0 \quad \Leftrightarrow \quad \frac{\partial r(\bar{\theta}^D)}{\partial\bar{\theta}^D}\frac{\bar{\theta}^D}{r(\bar{\theta}^D)} \lesseqqgtr -1 \quad \Leftrightarrow \quad \varepsilon^r \lesseqqgtr -1 \end{aligned}$$

The difference of private and efficient investment incentives under the QoS tiering regime is:

$$\begin{aligned} W^D - \Pi^D &= \frac{\sqrt{\lambda+1} - 1}{\lambda + 1}\mu r(\bar{\theta}^D) \\ \frac{\partial\left(W^D - \Pi^D\right)}{\partial\mu} &= \frac{\sqrt{\lambda+1} - 1}{\lambda + 1}\left(\frac{\partial r(\bar{\theta}^D)}{\partial\bar{\theta}^D}\mu + r(\bar{\theta}^D)\right) \lesseqqgtr 0 \quad \Leftrightarrow \\ \frac{\partial r(\bar{\theta}^D)}{\partial\bar{\theta}^D}\frac{\bar{\theta}^D}{r(\bar{\theta}^D)} &\lesseqqgtr -1 \quad \Leftrightarrow \quad \varepsilon^r \lesseqqgtr -1 \end{aligned}$$

$\square$

Proof of Proposition 8

Proof. To show that the ISP under a minimum quality standard enforcement of $w_2^D = w^N$ has a higher incentive to invest in capacity than under network neutrality we have to show that $\mu^{MQS} > \mu^N$.

$$\begin{aligned} w_2^D &= w^N \\ \Leftrightarrow \quad \frac{\mu^{MQS}}{\mu^{MQS} - \lambda\bar{\theta}^D}\frac{1}{\mu^{MQS} - \lambda\beta\bar{\theta}^D} &= \frac{1}{\mu^N - \lambda\bar{\theta}^N} \\ \Leftrightarrow \quad \mu^{MQS} &= \frac{1+\lambda}{1+\lambda\left(1-\beta\right)}\mu^N \end{aligned}$$

Since $\beta < 1$ it is easy to see, that $\mu^{MQS} > \mu^N$ always holds true. ☐

Proof of Proposition 9

Proof. In the QoS tiering regime with maximum quality degradation the last CP to enter the market is located at $\overline{\theta}^d = \frac{\mu\left(1-\frac{p^d}{r}\right)}{1+\lambda\left(1-\frac{p^d}{r}\right)}$. In contrast, the last CP to enter under network neutrality, or equivalently under the untempered QoS tiering regime, is located at $\overline{\theta}^{N,D} = \frac{\mu}{1+\lambda}$. Obviously, for $p^d = 0$, which corresponds to a network neutrality regime, the indifferent content providers coincide. However, $\forall p^d > 0$ it easy to see that $\overline{\theta}^d < \overline{\theta}^{N,D}$ for $\lambda > 0 > \frac{r}{p^d - r}$. This proves the first part of the proposition.

The ISP's profit under maximum quality degradation is $\Pi^d = a(p^d) + \lambda\,\overline{\theta}^d(p^d)\,p^d$, which is maximized by a price of $p^{d^*} = \frac{r(1+\lambda)-\sqrt{r(v+r(1+\lambda))}}{\lambda}$. At this price level, $W^d < W^N$ iff $v < r\lambda\left(1+\lambda\right)$, which is the same condition as for a positive equilibrium price. Thus, as long as $p^d > 0$ (which is shown in Proposition 10), welfare is lower under QoS tiering with maximum quality differentiation compared to network neutrality (which again has lower welfare than the untempered QoS tiering regime in the short-run). ☐

Proof of Proposition 10

Proof. At the optimal price level for Π^d and Π^D, solving $\Pi^d > \Pi^D$ for v yields

$$v < r\left(\lambda\left(3+2\lambda-2\sqrt{\lambda+1}\right)-2\sqrt{\lambda\,(\lambda+1)^2\left(\lambda+2-2\sqrt{\lambda+1}\right)}\right) \equiv \underline{v} \tag{33}$$

Furthermore, $\forall\lambda > 0$ it holds that $\underline{v} < r\lambda(1+\lambda)$ at which $p^d = 0$. Thus, the ISP never engages in maximum quality degradation at $p^d \leq 0$, but prefers the untempered QoS tiering regime instead. $\square$

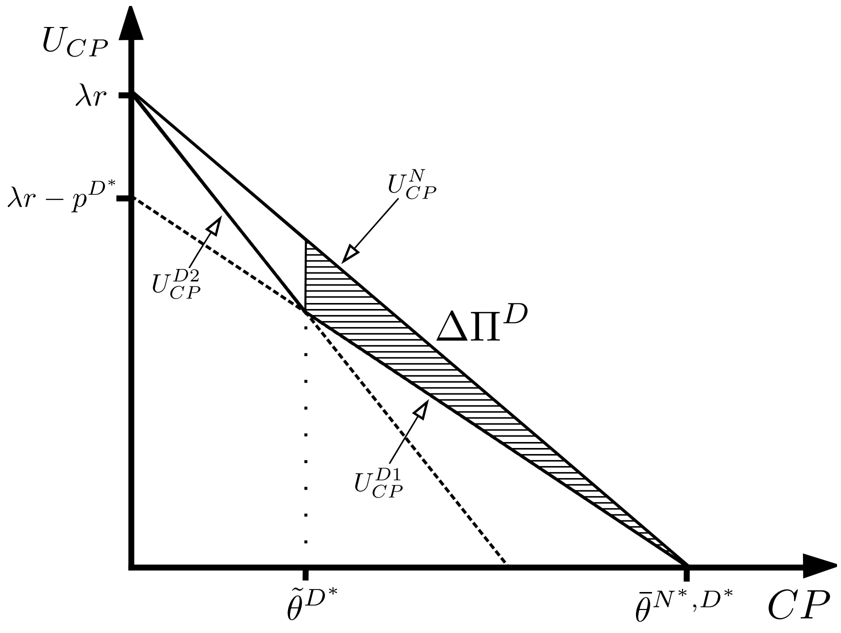

Figure 1: The Short-Run Effect of QoS Tiering on Content Providers' Surplus

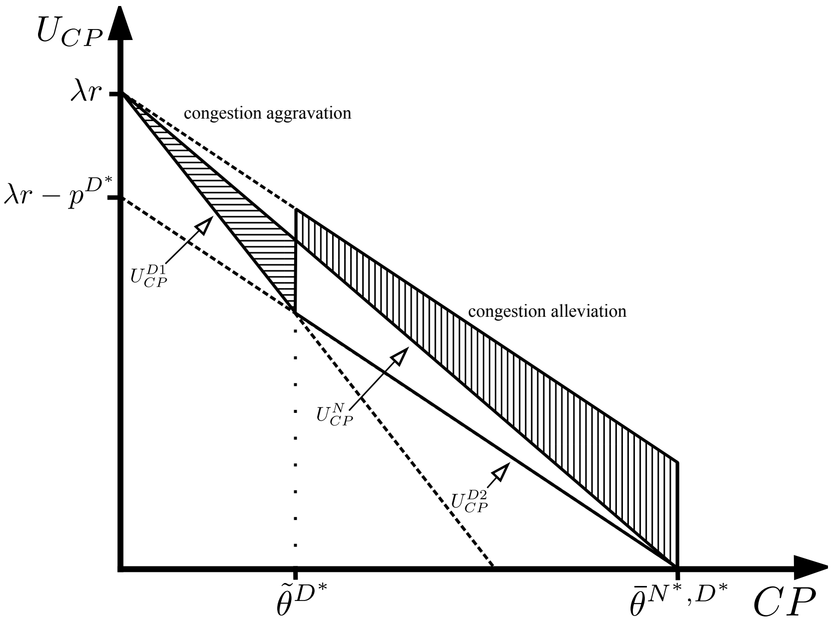

Figure 2: Congestion Alleviation vs. Congestion Aggravation Effect of QoS Tiering

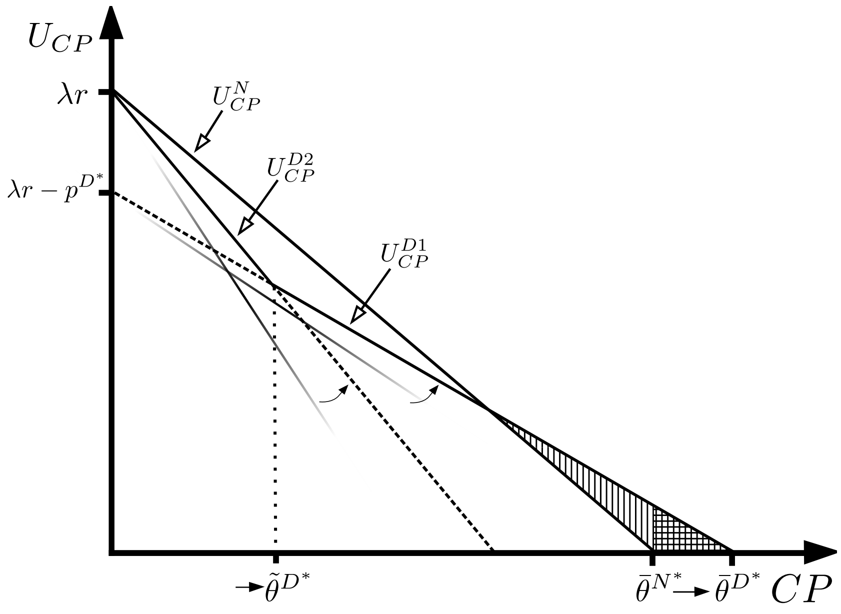

Figure 3: The Long-run Effect of QoS Tiering on Innovation and Welfare for $\varepsilon^r > -1$.