Net Neutrality, Prioritization and the Impact of Content Delivery Networks

Pio Baake (DIW Berlin)
Slobodan Sudaric (HU Berlin)

Discussion Paper No. 102

June 25, 2018

Collaborative Research Center Transregio 190 | www.rationality-and-competition.de
Ludwig-Maximilians-Universität München | Humboldt-Universität zu Berlin
Spokesperson: Prof. Dr. Klaus M. Schmidt, University of Munich, 80539 Munich, Germany
+49 (89) 2180 3405 | info@rationality-and-competition.de

Net Neutrality, Prioritization and the Impact of Content Delivery Networks

Pio Baake
DIW Berlin[*]

Slobodan Sudaric
Humboldt-Universität zu Berlin[†]

June 7, 2018

Abstract

We analyze competition between Internet Service Providers (ISPs) where consumers demand heterogeneous content within two Quality-of-Service (QoS) regimes, Net Neutrality and Paid Prioritization, and show that paid prioritization increases the static efficiency compared to a neutral network. We also consider paid prioritization intermediated by Content Delivery Networks (CDNs). While the use of CDNs is welfare neutral, it results in higher consumer prices for internet access. Regarding incentives to invest in network capacity we show that discriminatory regimes lead to higher incentives than the neutral regime as long as capacity is scarce, while investment is highest in the presence of CDNs.

JEL–classification numbers: L13, L51, L96

Keywords: content delivery network, investment, net neutrality, prioritization

[*]DIW Berlin, Mohrenstr. 58, 10117 Berlin, Germany, Email: pbaake@diw.de.

[†]HU Berlin, Spandauer Str. 1, 10178 Berlin, Germany, Email: sudarics@hu-berlin.de.

We wish to thank Marc Bourreau, Philipp Dimakopoulos, Tianchi Li, Robert Somogyi, Roland Strausz as well as two anonymous referees for helpful comments and suggestions. We also thank participants of the 2016 IODE workshop in Louvain-la-Neuve, 2016 EARIE conference in Lisbon and 2017 ITS European conference in Passau. Financial support through the German Research Foundation (RTG 1659 and CRC TRR 190) is gratefully acknowledged. A previous version of this paper circulated under the title "Interconnection and Prioritization".

1 Introduction

This paper contributes on the ongoing debate on 'net neutrality' – a concept that broadly requires that all internet traffic should be treated equally (Wu, 2003). One central aspect within the debate revolves around differentiation with respect to Quality-of-Service (QoS), i.e. whether or not all content classes should face identical service quality within the network. While opponents of net neutrality argue that QoS differentiation is part of reasonable network management and should therefore be allowed if not encouraged, net neutrality proponents argue that this benefits mainly network providers as it opens up new revenue models, and picks a few winners amongst the landscape of content providers (CPs). Indeed this ambivalence can be found e.g. in EU guidelines (EP and Council of the EU, 2015; BEREC, 2016) where a neutral treatment of internet traffic appears as a central pillar of the new regulation, while internet service providers (ISPs) may still offer differentiated QoS under certain conditions.[1] While there are various ways of QoS alterations within the management of a network, we would like to focus on the practice of 'paid prioritization' where CPs pay ISPs directly for prioritization of their content. We also consider the impact of Content Delivery Networks (CDNs) such as Akamai or Limelight. Instead of contracting with network operators directly, content providers can contract with an intermediary, the CDN, which then delivers the traffic to the ISPs.[2]

The purpose of this paper is therefore to analyze how paid prioritization affects, firstly, the static efficiency for a given network infrastructure, and secondly, the dynamic efficiency regarding incentives for investment in network capacity. In a neutral regime ISPs are only allowed to offer one quality level, i.e. all participants experience potential network congestion to the same extent. In a paid prioritization regime ISPs can charge CPs for bypassing

[1] As long as there is no discrimination within content classes, differentiated QoS measures can be applied to different content classes if they are considered to be 'reasonable'. While traffic management measures can not be put in place based on purely commercial considerations, the guidelines remain silent on pricing of differentiated QoS in the case they are technologically reasonable. For further details we refer to BEREC (2016).

[2] CDNs often have direct interconnection points with last-mile networks which can lead to higher traffic quality when delivering content to consumers. However, this quality improvement is not seen as a violation of the principle of net neutrality, as all traffic within the last-mile network is continued to be treated equally, even though from a consumer point of view a quality differentiation takes place. For example, the Netflix-Comcast dispute was not about offering priority lines for Netflix's services, but rather about Comcast's decision to demand interconnection charges from CDNs with a large amount of outgoing data traffic (caused by Netflix). See for example 'Comcast vs. Netflix: Is this really about Net neutrality?' (Retrieved May 17,2018 from https://www.cnet.com/news/comcast-vs-netflix-is-this-really-about-net-neutrality).

the network congestion by having access to a 'priority lane'. In a CDN environment ISPs offer access to their priority lanes to CDNs instead, which then resell the access to CPs. This setup reflects the idea of capacity bottlenecks in the regional or last-mile segment where congestion occurs because of high consumer demand (e.g. in legacy copper or coaxial networks).

We present a two-sided market model where two symmetric ISPs compete for consumers and CPs. Consumers are assumed to single-home, i.e. they purchase internet access only once, while CPs are free to multi-home with respect to their QoS choice. Content is differentiated with respect to connection quality sensitivity and quality levels are derived from a M/M/1 queuing system, where the non-priority quality ('best-effort') always remains free of charge, while the priority quality becomes a possible revenue source.

Using this framework, we show that the two regimes of QoS differentiation are welfare superior to the neutral regime. As content is differentiated, a tiered quality regime allocates priority to highly sensitive content classes while it leaves content classes with low quality sensitivity in the waiting queue, resulting in a more efficient use of existing network capacity. In particular we show that from a welfare perspective it is irrelevant whether this is achieved by direct paid prioritization or through the use of a CDN. Differences emerge once we take into account strategic effects of the QoS regimes on competition for consumers. Here we argue that QoS differentiation makes the consumer market more elastic leading to lower consumer prices in regimes of QoS differentiation compared to the neutral regime. In particular under paid prioritization consumer prices are lowest as here a price increase on the user market has an additional negative effect on the CP market, while this is not the case in a CDN environment. Lastly, we analyze unilateral incentives to increase network capacity from a symmetric equilibrium perspective and show that as long as network capacity is scarce, both discriminatory regimes lead to higher investment in network capacity than the neutral regime, while investment is highest in the CDN case irrespective of the initial capacity level.

Related literature

From a modeling perspective we build on the literature on competition in two-sided markets in general and applications in the telecommunications industry in particular. The general setup follows the competitive bottleneck idea in Armstrong (2006) in the sense that we consider single-homing consumers and allow for multi-homing on the CP side. Applications of a two-sided approach to telephone networks (Armstrong, 1998; Laffont et al., 1998a,b) and to the internet industry (Laffont et al., 2003) can also already be found in earlier work. The key difference is that we explicitly model network congestion and resulting questions of QoS differentiation, while the early stream of literature largely disregards questions of network quality.

This aspect is analyzed in detail in the younger but growing literature on net neutrality.[3] Hermalin and Katz (2007) compare a neutral network where ISPs are restricted to offer a single quality level as opposed to a discriminatory regime where ISPs can offer multiple quality levels to CPs. They conclude with ambiguous welfare effects: offering a single quality level drives some content types out of the market and provides an inefficient low quality level for other content types. However, CPs 'in the middle' are likely to benefit from it. Economides and Hermalin (2012) expand on this result by explicitly modeling bandwidth limits where different qualities could introduce welfare gains in light of congested networks. Following a similar QoS approach, Economides and Hermalin (2015) further show that net neutrality leads to lower investment incentives. Guo and Easley (2016) consider QoS differentiation with respect to effective bandwidth and demonstrate that net neutrality is beneficial for content innovation. Another stream of literature tackles the congestion problem using a queuing approach. Choi and Kim (2010) and Cheng et al. (2011) present a model where a monopoly ISP offers a prioritization service to two CPs. This framework is extended by Krämer and Wiewiorra (2012) to a model with a continuum of heterogeneous CPs. While Choi and Kim (2010) and Cheng et al. (2011) derive mixed results regarding welfare and investment incentives, Krämer and Wiewiorra (2012) show that a discriminatory regime is more efficient and provides higher investment incentives in the long run. While we follow the same direction in terms of CP heterogeneity and the use of queuing, our model differs substantially as we consider platform competition.

This aspect is captured to some extent by Economides and Tåg (2012) and Njoroge et al. (2013) where platform competition is considered but the congestion issue is ignored.

[3] Greenstein et al. (2016) provide an excellent overview over the inherent trade-offs of the net neutrality debate as well as the associated literature.

Choi et al. (2015) present a closely related model in terms of content differentiation and analyze how the business model of CPs affects the optimal price-quality choice of platforms. The key difference is that while we keep the business model fixed in our model, qualities are endogenous in the sense that they are affected by congestion. Secondly, in the case of competition the authors consider cooperative quality choice, while we consider competition in the quality dimension through the platforms' pricing strategies. Kourandi et al. (2015) also consider the case of competing ISPs but focus on the aspect of internet fragmentation when ISPs obtain exclusivity over content. The work most closely related to our model is the paper by Bourreau et al. (2015) where competing ISPs offer queuing based prioritization to differentiated CPs. The main difference from a modeling perspective is how surplus is generated in the economy, as the authors consider an elastic number of CPs and interpret the exclusion of CPs as decrease of content variety. In our model consumers' utility depends on the connection quality of consumed content and not on variety per se. One could therefore see our modeling setup as a combination of the models presented in Choi et al. (2015) and Bourreau et al. (2015). Further, we additionally introduce CDNs as intermediaries which are not considered in any of the previously mentioned papers.

In general the topic of CDNs has largely been disregarded in the net neutrality debate. Hosanagar et al. (2008) study the optimal pricing policy of CDNs but do not perform any welfare comparisons. This is done to some extent in Hau et al. (2011), where different QoS regimes are analyzed in the market for internet interconnection. The overall model differs substantially from ours and in particular the authors do not consider competition between ISPs for consumers, which is a main driver for our results. Interestingly, however, the authors also find that a CDN shifts rents away from consumers to ISPs, a result which qualitatively reoccurs in our analysis, although the underlying mechanics differ. In particular, our results show that CDNs soften competition for consumers compared to a regime where CPs directly contract with ISPs.

Our analysis supports the results obtained by Krämer and Wiewiorra (2012) and Bourreau et al. (2015): a discriminatory regime is superior in terms of static efficiency and tends to provide higher investment in network capacity. At the same time our work complements the existing literature in terms of the role of CDNs. While total efficiency is identical to paid prioritization, consumers face higher prices when CDNs are used. Regarding the ongoing debate on net neutrality our results therefore suggest that if QoS differentiation is to be allowed (see e.g. recent advances in the US), direct prioritization agreements between CPs and ISPs should be preferred over the indirect contracting via CDNs from a (static) consumer perspective, as they lead to lower consumer prices, while investment in network infrastructure is highest in the presence of CDNs.

The paper is structured as follows. Section 2 introduces the model. Section 3 presents equilibrium outcomes for the different QoS regimes. Section 4 compares the different outcomes regarding efficiency and investment incentives. Section 5 concludes.

2 Model

We study different QoS regimes in a two-sided market setting where ISPs deliver content from CPs on one market side to consumers on the other market side. CPs strike QoS deals either with ISPs directly (section 3.1 and 3.2) or with a CDN in section 3.3.

Internet service providers: There are two identical ISPs $i = 1, 2$ located at the ends of a Hotelling line (location $\lambda_i = 0$ for $i = 1$ and $\lambda_i = 1$ for $i = 2$). ISPs sell internet access to consumers at price p_i and make QoS offers (f_i, q_i) to CPs, such that in exchange for a fee f_i consumers in network i can be reached at quality q_i. In the case of net neutrality the only offer ISPs can make is of the form $(0, q_i^n)$ where q_i^n denotes the best-effort quality in network i, which is free of charge. This reflects the idea that there is ubiquitous interconnectivity in the economy such that CPs can reach consumers of network i irrespective of whether there is an existing agreement with the network. Under paid prioritization ISPs can offer in addition to the free best-effort quality a prioritization service (f_i, q_i^p) with $f_i \geq 0$ where q_i^p denotes the priority quality level in network i.

The quality levels q_i^n and q_i^p are derived from a M/M/1 queuing model with an arrival rate of content requests equal to one such that waiting times are given by

$$w_i^p = \frac{1}{k_i - N_i Y_i} \text{ with prioritization,} \tag{1}$$

$$w_i^n = \frac{k_i}{k_i - N_i} w_i^p \text{ without prioritization,} \tag{2}$$

where $N_i \in [0, 1]$ denotes the mass of consumers connected to ISP i, $Y_i \in [0, 1]$ denotes the mass of CPs who purchased prioritization in network i, and k_i is the network capacity of ISP i. Quality levels in network i are then defined as

$$q_i^p = 1 - w_i^p \text{ with prioritization,} \tag{3}$$

$$q_i^n = 1 - w_i^n \text{ without prioritization.} \tag{4}$$

Further, we make the following assumption regarding network capacities such that quality levels remain non-negative.

Assumption 1 *Network capacities are sufficiently large* $k_i \in (2, \infty)$.

This assumption ensures that waiting times do not explode for low capacity levels such that we have $w_i^p, w_i^n \in (0, 1)$ and therefore $q_i^p, q_i^n \in (0, 1)$. Also this assumption implies that each network could shoulder the whole traffic by itself such that there are not any purely allocative reasons behind our setup. Also note that $q_i^p > q_i^n$ and $\lim_{k\to\infty}(q_i^p - q_i^n) = 0$, i.e. if capacities are large waiting times in all queues converge to zero and quality differences disappear.

Content providers: There is a continuum of differentiated CPs with total mass normalized to one. CPs are differentiated with respect to their quality sensitivity $\theta \in \Theta \equiv [0, 1]$ which we assume to be uniformly distributed. Low values of θ correspond to content-types with low sensitivity with respect to transmission quality (e.g. e-mails) whereas high values represent quality-sensitive services (e.g. live streaming).

We assume the CPs' business model is entirely passive (e.g. ad-based) and that the delivery of content of type θ at quality level q to one consumer generates advertisement revenues $r(\theta, q) = \theta q$ such that θ measures the importance of quality for the revenue generation. A CP of type θ decides whether to purchase prioritization ($h_i^\theta = p$) in network i or not ($h_i^\theta = n$) such that profits obtained from network i are given by

$$\pi_i(\theta, h_i^\theta) = \begin{cases} r(\theta, q_i^n)N_i & \text{if } h_i^\theta = n \\ r(\theta, q_i^p)N_i - f_i & \text{if } h_i^\theta = p \end{cases} \tag{5}$$

resulting in total profits of a CP of type θ with QoS plan $h^\theta = \{h_1^\theta, h_2^\theta\}$ of

$$\pi(\theta, h^\theta) = \sum_{i \in \{1,2\}} \pi_i(\theta, h_i^\theta). \tag{6}$$

We can then define $\mathcal{P}_i = \{\theta \in \Theta \mid h_i^\theta = p\}$ as the set of content types prioritizing in network i and $Y_i = \int_{\theta \in \mathcal{P}_i} d\theta$ as the total mass of prioritized content in network i.

Consumers: There is a continuum of differentiated consumers with total mass normalized to one. Consumers have a uniformly distributed location $x \in [0,1]$ and obtain utility $v(\theta, q) = \theta q$ from consuming one unit of content from a CP of type θ delivered with quality q.[4] The total utility V_i from content consumption in network i is then given by

$$V_i = \int_{\theta \notin \mathcal{P}_i} v(\theta, q_i^n) d\theta + \int_{\theta \in \mathcal{P}_i} v(\theta, q_i^p) d\theta \tag{7}$$

and can be thought of as a summary statistic for the network quality of ISP i. Overall utility $u_i(x)$ from connecting to network i is then given by

$$u_i(x) = \underline{u} + V_i - p_i - |\lambda_i - x| \tag{8}$$

and depends on the aggregate utility from content consumption V_i, the internet access price p_i and the location of the consumer.[5] Lastly, $\underline{u}$ captures utility which is derived from connecting to the internet but not covered by our CP model and is assumed to be sufficiently high such that market coverage is ensured.

Timing In a first step, ISPs set consumer prices p_i and (if allowed) prioritization fees f_i. Secondly, consumers decide which network to join and CPs decide in which network to purchase prioritization (if applicable) simultaneously. The solution concept is sub-game perfection.

[4]Note, in our model $v(\theta, q) = r(\theta, q)$ which is a simplifying assumption. We could also allow for a setting where consumers receive a fraction s of the surplus θq_i and CPs the remaining fraction $(1-s)$. Our results would not change qualitatively.

[5]We omit the arguments of V_i where it does not lead to confusion.

3 Equilibrium analysis

In this section we present equilibrium outcomes for the three different QoS regimes which we will refer to by the superscripts given in brackets: Net neutrality (n), Paid prioritization (p) and Content Delivery Networks (c). Details of the formal analysis are delegated to Appendix A and proofs can be found in Appendix B.

3.1 Net neutrality

In this section we consider the benchmark scenario of net neutrality. In this scenario ISPs can not sell prioritization and their only source of revenue is selling internet access to consumers, i.e. we have $\mathcal{P}_i = \emptyset$ and therefore $Y_i = 0$ in both networks. As the best-effort quality level is free of charge, CPs will reach consumers of network i at quality level q_i^n such that we have $h^\theta = (n, n)\ \forall\ \theta$. Total profits from content delivery obtained by a CP of type θ are then given by

$$\pi(\theta, h^\theta) = \theta\,(q_1^n N_1 + q_2^n N_2) \tag{9}$$

with $q_i^n = 1 - 1/(k_i - N_i)$ for $i = \{1, 2\}$. Turning to consumers the aggregate utility from content consumption without prioritization V_i is then given by

$$V_i = q_i^n \int_0^1 \theta d\theta. \tag{10}$$

Since there is only one quality level in the neutral regime all content types arrive at the uniform quality level q_i. The consumer market shares of both ISPs are then given by the indifferent consumer $\hat{x} : u_1(\hat{x}) = u_2(\hat{x})$ on the Hotelling line.

$$N_1 = \hat{x} \equiv \frac{1}{2} + \frac{1}{2}\left[(V_1 - p_1) - (V_2 - p_2)\right] \text{ and } N_2 = 1 - \hat{x}. \tag{11}$$

Note, that (11) defines N_i only implicitly as the quality levels q_i^n also depend on the consumer market shares. We therefore make use of the implicit function theorem to obtain market share reactions $\partial N_i/\partial p_i$. Details can be found in Appendix A. The first order conditions to the ISPs' maximization problems

$$\max_{p_i} \Pi_i = p_i N_i \tag{12}$$

can then be written using the market share reactions obtained in (34) such that we get

$$p_i = \frac{N_i}{-\partial N_i/\partial p_i}. \tag{13}$$

Note, that even though this is a very simple maximization problem it is not the standard Hotelling problem. The endogeneity of q_i^n leads to less elastic market shares N_i. Utilizing symmetry in network capacities $k_i = k_j = k$ we obtain the unique symmetric solution $p^n := \arg\max_{p_i} \Pi_i|_{p_j=p^n}$ where

$$p^n = \underbrace{1 + \frac{2}{(2k-1)^2}}_{=-N_i/(\partial N_i/\partial p_i)}. \tag{14}$$

Equilibrium market shares are then given by $N_i = N_j = 1/2$. Regarding comparative statics we see that $\partial p^n/\partial k < 0$, i.e. consumer prices are lower for higher (symmetric) capacity levels. The reason is that higher capacity levels make consumer demand more elastic $\partial^2 N_i/\partial p_i \partial k < 0$. Consider the case where k is very large. Then congestion is basically irrelevant and quality levels in both networks effectively do not depend on the ISPs' market shares, such that ISPs only compete in prices. If capacity is scarce the congestion problem dampens consumers willingness to switch networks as by joining the rival network the rival's quality decreases. Hence, demand is less elastic and consumer prices increase. The property $\partial p^n/\partial k < 0$ will reoccur throughout the analysis and we will refer to it as 'capacity effect'.

3.2 Paid prioritization

In this section we consider the case where ISPs directly offer CPs paid prioritization agreements. The proposed offer consists of content delivery to all consumers in network i at priority quality q_i^p in exchange for a fee f_i, while content delivery at the best-effort quality level q_i^n remains free of charge.[6]

CPs make the decision whether to purchase prioritization for each network separately. The decision depends on how the profit of reaching consumers connected to ISP i at best-effort quality q_i^n compares to the profit under a prioritization agreement with access to the priority quality q_i^p. By comparing the profit levels given in (5) we can pin down an indifferent CP of type $\hat{\theta}_i$ such that $\pi_i(\hat{\theta}_i, n) = \pi_i(\hat{\theta}_i, p)$ with

$$\hat{\theta}_i = \frac{f_i}{(q_i^p - q_i^n)N_i}. \tag{15}$$

CPs will therefore engage in a prioritization contract if they offer sufficiently quality-sensitive content $\theta \geq \hat{\theta}_i$, and stick to the best-effort quality if their content type is insensitive $\theta < \hat{\theta}_i$. The set of prioritizing CPs in network i is then given by $\mathcal{P}_i = [\hat{\theta}_i, 1]$ such that the mass of prioritized traffic in network i is given by $Y_i = 1 - \hat{\theta}_i$. Turning to consumers the aggregate utility from content consumption V_i under prioritization is given by

$$V_i = q_i^n \int_0^{\hat{\theta}_i} \theta d\theta + q_i^p \int_{\hat{\theta}_i}^1 \theta d\theta \tag{16}$$

and consists of prioritized ($\theta \geq \hat{\theta}_i$) and non-prioritized ($\theta < \hat{\theta}_i$) content. The consumers' decision which network to join is given as in (11) by pinning down an indifferent consumer. The profit maximization problem of an ISP can then be written as

$$\max_{p_i, f_i} \Pi_i = p_i N_i + f_i Y_i. \tag{17}$$

Due to the endogeneity of the quality levels, we again apply the implicit function theorem to obtain consumer market share reactions $\partial N_i/\partial p_j, \partial N_i/\partial f_j$ and CP share reactions $\partial Y_i/\partial p_j, \partial Y_i/\partial f_j$ for $i, j = \{1, 2\}$. Further we introduce the following intermediary result which provides us assurance of an interior solution to the maximization problem.

Lemma 1 *Each ISP has an incentive to offer prioritization.*

Proof. See Appendix. ■

[6] As we consider unit demand the distinction between a linear per-consumer fee and a lump-sum fee to reach all consumers in network i is irrelevant. We stick to the latter specification for reasons of conciseness.

First, prioritization introduces additional revenue streams on the CP side of the market. Secondly, compared to no prioritization the network's overall quality V_i increases as some highly sensitive content types now arrive at high quality, while the quality of the remaining content types barely changes. This pushes more consumers into the network offering prioritization which increases the ISP's profit even further. As this argument holds for each ISP irrespective of whether the other ISP offers prioritization or not, offering prioritization is a strictly dominant strategy. Given Lemma 1 we can now focus on the interior solution given by the first order conditions to the maximization problem in (17) such that

$$p_i = \frac{N_i}{-\partial N_i/\partial p_i} - f_i \frac{\partial Y_i/\partial p_i}{\partial N_i/\partial p_i} \text{ and } f_i = \frac{Y_i}{-\partial Y_i/\partial f_i} - p_i \frac{\partial N_i/\partial f_i}{\partial Y_i/\partial f_i}. \quad (18)$$

Comparing (18) to (13) we see that optimal consumer price setting now takes into account the effect on the CP market, where an increase in prices reduces the number of consumers and hence reduces the revenue from the prioritization business as it decreases the share of prioritizing CPs. Going back to the definition of the indifferent content class in (15) we see that there are two effects affecting the share of prioritizing CPs. First, there is a direct effect when increasing consumer prices, as the share of consumers N_i decreases. Secondly, the indifferent content class depends on the difference in quality levels $q_i^p - q_i^n$. As a reduction in the number of consumers reduces the total traffic in the network and hence the congestion problem, the difference in quality levels decreases when the number of consumers goes down, pushing the indifferent content class upwards and hence reducing the share of prioritizing CPs.[7] In summary, introducing prioritization therefore restricts the ability of ISPs to raise consumer prices.

The optimal prioritization fee similarly balances the revenue generation across both market sides. While an increase in fees reduces the share of prioritizing CPs, the effect on the consumer market is not necessarily monotone. Coming from a situation of no prioritization, a higher share of prioritized content increases utility from content consumption as quality sensitive content arrives at high quality. However, if the share of prioritized content is too large, the congestion externality imposed on the priority queue might outweigh the

[7] It is easy to verify that $\partial(q_i^p - q_i^n)/\partial N_i > 0$.

benefits of prioritizing additional content classes which would decrease overall network quality.

Continuing with the analysis we again find a symmetric equilibrium such that $(p^p, f^p) = \arg\max_{p_i,f_i} \Pi_i|_{p_j=p^p,f_j=f^p}$ resulting in $Y_i = Y_j = Y^p \equiv 1-\hat{\theta}^p = 2k-\psi$ and $N_i = N_j = 1/2$ as well as $\partial N_i/\partial f_i = 0$ for $i \neq j$ with equilibrium values

$$p^p = 5 - \underbrace{\frac{(8k-6)\psi}{(2k-1)^2}}_{=\frac{N_i}{-\partial N_i/\partial p_i}} - f^p \underbrace{\frac{4k}{(2k-1)}}_{=\frac{\partial Y_i/\partial p_i}{\partial N_i/\partial p_i}} \text{ and } f^p = Y^p \underbrace{\frac{1}{2k(2k-1)}}_{=\frac{1}{-\partial Y_i/\partial f_i}} \tag{19}$$

where $\psi := \sqrt{2k(2k-1)}$. Note that $\partial Y^p/\partial k < 0$ (or equivalently $\partial\hat{\theta}^p/\partial k > 0$), i.e. the higher the capacity level in the market the lower share of prioritized content classes. If capacity levels rise, networks become less congested and the quality gain from prioritization decreases.[8] Therefore, only CPs with extremely sensitive content types opt for prioritization. Regarding consumer prices we obtain $\partial p^p/\partial k < 0$ which is in line with the capacity effect described in section 3.1. The effect of the quality level on prioritization fees is given by $\partial f^p/\partial k < 0$ which reflects the decreasing advantage of prioritization if overall capacity is large. Further, this effect prevails even in presence of an increased elasticity on the consumer market such that standard platform logic would predict a price increase on the CP market side.

Note that from equations (14) and (19) we can infer the equilibrium market share elasticity $\partial N_i/\partial p_i$ in both regimes. Comparing the two cases we see that the consumers' reaction to price changes is stronger in a prioritization regime.[9] The reason for this is that introducing a priority queue already eases the congestion problem in the networks. Therefore by switching to the rival network the overall network quality decreases less, hence market shares are more elastic in a prioritization regime. We will refer to this effect simply as 'elasticity effect'. Note this effect is very similar to the capacity effect described in section 3.1. However, while the capacity effect states that market shares become more elastic when the capacity level k increases, the elasticity effect states that for a given level of k market shares are more elastic in a prioritization regime.

[8] To see this remember that $\lim_{k\to\infty}(q_i^p - q_i^n) = 0$.

[9] The comparison boils down to $\psi > 2(2k-1)^2/(4k-3)$ which is satisfied under Assumption 1.

Lastly, we get $\partial N_i/\partial f_i = 0$ in equilibrium. To gain intuition for this result consider the case where coming from a neutral regime (f_i prohibitively high) the prioritization fee is reduced such that $Y_i > 0$. This increases the revenue on the CP side and at the same time increases the network's quality which attracts more consumers. This 'double benefit' is exploited fully in equilibrium, resulting in $\partial N_i/\partial f_i = 0$.

3.3 Content Delivery Network

In this section we consider an alteration to the prioritization setup presented in section 3.2. In particular we introduce a Content Delivery Network (CDN) as an additional player which serves as an intermediary between CPs and ISPs. The idea is that the CDN enters an agreement with ISPs such that traffic coming from the CDN is prioritized, while traffic not coming from the CDN remains unprioritized.[10]

For this we introduce an additional 'offer stage' at the beginning of the game. In the offer stage the CDN publicly announces lump-sum transfers $F_i \in \mathbb{R}$, which the ISPs can either accept or reject.[11] If ISP i accepts offer F_i, the CDN is free to set the prioritization fee f_i for reaching costumers in network i just like in section 3.2 while ISP i only sets consumer prices p_i. If ISP i rejects offer F_i, prioritization in network i is offered by ISP i instead.[12] In any case prices p_i, p_j and prioritization fees f_i, f_j are set simultaneously as before. This setting resembles the industry practice, where ISPs and CDNs make long-term infrastructure level decisions, while offers made to consumers and CPs are made once those decisions are made.[13]

To avoid multiplicity of equilibria we apply the payoff dominance refinement (Harsanyi and Selten, 1988) to the coordination game in the offer stage, such that in case there

[10]For simplicity reasons we abstract from any additional quality improvements due to the use of CDNs.

[11]One can alternatively consider a two-part tariff $T_i = (t_i, F_i) \in \mathbb{R}^2$ where t_i is an additional linear fee. It is clear that t_i introduces a double marginalization inefficiency which would reduce the total obtainable profit of the CDN. We therefore restrict our analysis to the case of $t_i = 0$ which reduces the proposal to the lump-sum fee F_i.

[12]We implicitly assume that ISPs commit to not offer prioritization themselves in case they accept the offer such that the offer F_i can be seen as an exclusive dealing arrangement. Without commitment the standard Bertrand argument would apply, as in particular the CDN would undercut any positive fee set by the ISP.

[13]The fact that offers F_i are public is a simplifying assumption which allows us to focus on the induced change in competition dynamics. If we consider private offers instead, existence of the presented equilibrium remains unchanged.

are multiple equilibria when deciding whether to accept offers F_i, we select the Pareto-dominant equilibrium in terms of ISP profits.

Suppose both offers have been accepted. The maximization problem of the CDN is then given by

$$\max_{f_i,f_j} \Pi_c = f_i Y_i + f_j Y_j - F_i - F_j,\ i \neq j \tag{20}$$

where Y_i and Y_j are obtained as in section 3.2 and F_i, F_j denote (sunk) lump-sum transfers to both ISP i and j. ISPs in this case only compete on the consumer market:

$$\max_{p_i} \Pi_i = p_i N_i + F_i,\ i = 1, 2 \tag{21}$$

Market share reactions $\partial N_i/\partial p_j, \partial N_i/\partial f_j$ as well as $\partial Y_i/\partial p_j, \partial Y_i \partial f_j$ for $i, j = \{1, 2\}$ are again obtained as in (36) giving rise to first order conditions to maximization problems (20) and (21) of:

$$p_i = \frac{N_i}{-\partial N_i/\partial p_i} \text{ and } f_i = \frac{Y_i}{-\partial Y_i/\partial f_i} - f_j \frac{\partial Y_j/\partial f_i}{\partial Y_i/\partial f_i},\ i \neq j. \tag{22}$$

We can immediately see that the maximization problem of the ISPs now closely resembles the maximization problem under the neutral regime. In particular ISPs now do not internalize the negative effect of a price increase on the share of prioritized content as they did in section 3.2. However, remember that there is a tiered queue on the content market, such that the market share reaction differs compared to the neutral regime due to the elasticity effect. The CDN on the other side now internalizes the effect of a fee setting in market i on the share of prioritized content in network j while in section 3.2 the fee setting internalized the effect on the consumer market share in network i.

Continuing with the analysis we again obtain a symmetric equilibrium $(f^c, f^c) = \arg\max_{f_i,f_j} \Pi_c|_{p_i=p_j=p^c}$ and $p^c = \arg\max_{p_i} \Pi_i|_{f_i=f_j=f^c, p_j=p^c}$ for $i \neq j$ resulting in $N_i = N_j = 1/2$ and $Y_i = Y_j = Y^c \equiv 1 - \hat{\theta}^c = 2k - \psi$ and $\partial Y_j/\partial f_i = 0$. Equilibrium values are given by

$$p^c = \underbrace{5 - \frac{(8k-6)\psi}{(2k-1)^2}}_{=\frac{N_i}{-\partial N_i/\partial p_i}} \text{ and } f^c = Y^c \underbrace{\frac{1}{2k(2k-1)}}_{=\frac{1}{-\partial Y_i/\partial f_i}} \tag{23}$$

and $\psi = \sqrt{2k(2k-1)}$ as in section 3.2. It now remains to show that this sub-game is actually reached, i.e. the CDN makes offers which are accepted by the ISPs.

Lemma 2 *The optimal offer is symmetric $F_i = F_j = F^c$ and is accepted by both ISPs in equilibrium.*

Proof. See Appendix. ∎

Given Lemma 2 we know that the CDN prefers contracting with both ISPs compared to contracting with only one ISP. Further, the proposed offers are accepted by the ISPs such that the presented equilibrium outcome is indeed sub-game perfect which allows us to compare derived equilibrium values to the previous QoS regimes. We immediately see that $f^c = f^p$ and $Y^c = Y^p$ while $p^c \neq p^p$ which gives rise to the following result.

Proposition 1 *The use of CDNs is welfare-equivalent to paid prioritization.*

Proof. See Appendix. ∎

Proposition 1 implies that from a total welfare perspective it is irrelevant whether prioritization is achieved by direct paid prioritization offers made by ISPs, or whether prioritization is offered through the use of a CDN. In particular the CDN will pick prioritization fees which are equivalent to the paid prioritization scenario, resulting in an identical share of prioritized content classes.

Going back to the definition of the critical content class in (15), we can see that the only effect f_i has on Y_j is via the consumer market share N_j. Now consider the case of f_i, f_j being large such that $Y_i = Y_j = 0$ and start decreasing f_i such that we obtain $Y_i > 0$. This increases the network quality V_i in network i and hence pulls consumers from network i into network j, increasing the revenue obtained from network i. Now consider a decrease in f_j such that $Y_j > 0$. Consumers are pulled away from network i into network j, decreasing the revenue obtained from network i and increasing the revenue obtained from network j. Given these 'push-and-pull' effects, it is optimal for the CDN to set its prioritization fees such that the marginal effect on the consumer market vanishes, resulting in $\partial N_i/\partial f_i = 0$ and thus $\partial Y_j/\partial f_i = 0$, which in turn leads to identical equilibrium fees as in section 3.2.

We can also immediately see that $p^c > p^p$ as the optimal consumer prices now do not take into account the adverse revenue effect on the CP side $\partial Y_i/\partial p_i$ as is the case under paid prioritization. Unsurprisingly, we therefore observe higher consumer prices in the CDN case. As we consider a covered consumer market the total welfare is unaffected by this price increase, resulting in Proposition 1.

4 Comparison

This section compares the different QoS regimes from section 3. In the first part we look at profits and consumer surplus separately to gain a better understanding of the underlying dynamics before combining our results in a single welfare measure. The second part compares incentives to invest in network capacities.

4.1 Welfare

We start this section by defining simplified surplus metrics for symmetric equilibrium outcomes. First, remember that in our symmetric outcomes $N_i = N_j = 1/2$ while the share of prioritized content is pinned down by an indifferent content class $\hat{\theta}$ such that the share of prioritized content takes the form $Y_i = Y_j = Y = 1 - \hat{\theta}$. It turns out to be helpful to denote equilibrium quality levels as functions of $\hat{\theta}$ such that we have $q_i^n = q_j^n = q^n(\hat{\theta}), q_i^p = q_j^p = q^p(\hat{\theta})$. Note, that in a regime of net neutrality we have $Y = 0$ or equivalently $\hat{\theta} = \hat{\theta}^n := 1$. Starting with the definition of consumer utility (8) we can then denote consumer surplus S_C as a function of symmetric consumer prices p and a cutoff level $\hat{\theta}$:

$$S_C(p, \hat{\theta}) = 2\int_0^{1/2} u_i(x)dx = \underline{u} + V(\hat{\theta}) - p - \frac{1}{4} \tag{24}$$

where

$$V(\hat{\theta}) = q^n(\hat{\theta})\int_0^{\hat{\theta}} \theta d\theta + q^p(\hat{\theta})\int_{\hat{\theta}}^1 \theta d\theta. \tag{25}$$

Similarly, we can define total CP industry profits S_{CP} as a function of a cutoff content class $\hat{\theta}$ and a symmetric prioritization fee f in the case of prioritization.[14]

$$S_{CP}(f,\hat{\theta}) = \begin{cases} V(\hat{\theta}) - 2f(1-\hat{\theta}) & \text{for } \hat{\theta} < 1 \\ V(\hat{\theta}) & \text{for } \hat{\theta} = 1 \end{cases} \tag{26}$$

Finally, we can define total ISP (incl. CDN in section 3.3) industry profits S_{ISP} as a function of prices p, f and critical content class $\hat{\theta}$.

$$S_{ISP}(p,f,\hat{\theta}) = \begin{cases} p + 2f(1-\hat{\theta}) & \text{for } \hat{\theta} < 1 \\ p & \text{for } \hat{\theta} = 1 \end{cases} \tag{27}$$

Combining all three measures into a total surplus measure TS we obtain

$$TS(\hat{\theta}) = S_C + S_{CP} + S_{ISP} = \underline{u} + 2V(\hat{\theta}) - \frac{1}{4}. \tag{28}$$

We immediately see that the network quality $V(\hat{\theta})$ plays a central role and we therefore introduce the following intermediate result which will become useful when we compare the different QoS regimes.

Lemma 3 *The network quality $V(\hat{\theta})$ is higher in a prioritization regime:*

$$V(1) < V(\hat{\theta}),\ \ \hat{\theta} \in (0,1)$$

Proof. See Appendix. ■

To gain intuition for Lemma 3 it is helpful to consider the extreme case $\hat{\theta} = 1$, i.e. no prioritization and all traffic taking place in the best-effort queue. Marginally decreasing $\hat{\theta}$ then implies that highly quality-sensitive content arrives at priority quality, while the quality for all the remaining traffic remains effectively unchanged, i.e. overall network quality increases. A similar argument can be made for the other extreme case $\hat{\theta} = 0$, where all content is 'prioritized', i.e. again the entire traffic takes place in a single quality queue,

[14] In a prioritization regime CP industry profits are given by $S_{CP}(f,\hat{\theta}) = \int_0^{\hat{\theta}} \pi(\theta,(n,n))\,d\theta + \int_{\hat{\theta}}^1 \pi(\theta,(p,p))\,d\theta = \int_0^{\hat{\theta}} 2r\left(q^n(\hat{\theta}),\theta\right)d\theta + \int_{\hat{\theta}}^1 2\left(r\left(q^p(\hat{\theta}),\theta\right) - f\right)d\theta = V(\hat{\theta}) - 2f(1-\hat{\theta})$ while in the neutral regime we have $S_{CP}(\emptyset,1) = \int_0^1 \pi(\theta,(n,n))\,d\theta = \int_0^1 2r\left(q^n(1),\theta\right)d\theta = V(1)$.

while for intermediate levels $\hat{\theta} \in (0,1)$ content distributes across both queues and some sensitive content classes arrive at priority quality. Hence $V(\hat{\theta})$ is high for intermediate levels of $\hat{\theta}$. Starting with consumers we then obtain the following result by comparing consumer surplus under the different QoS regimes.

Proposition 2 *Consumers benefit from prioritization as consumer prices decline and network quality increases. In particular we have:*

i.) $V(\hat{\theta}^p) = V(\hat{\theta}^c) > V(\hat{\theta}^n)$

ii.) $p^n > p^c > p^p$

iii.) $S_C(p^p, \hat{\theta}^p) > S_C(p^c, \hat{\theta}^c) > S_C(p^n, \hat{\theta}^n)$

Proof. See Appendix. ∎

Prioritization has two main benefits for consumers. First, it allocates existing capacity more efficiently such that highly quality sensitive content arrives at priority quality, while content classes for which transmission quality plays a minor role are put in a waiting queue. This increases the total utility from content consumption. Secondly, prioritization makes it harder for ISPs to raise consumer prices as the consumer market becomes more elastic, and since losing consumers to the rival network has an additional negative effect on the revenue obtained on the CP market side. The last effect is not present in the case of CDNs as here ISPs do not internalize the negative effect on the CP side. However, in both cases consumers benefit from prioritization. Turning to the content industry, the following proposition summarizes the main finding.

Proposition 3 *The content industry does not benefit from prioritization.*

$$S_{CP}(f^n, \hat{\theta}^n) > S_{CP}(f^p, \hat{\theta}^p) = S_{CP}(f^c, \hat{\theta}^c)$$

Proof. See Appendix. ∎

There are two main reasons why the content industry does not profit from prioritization. For content classes which are not prioritized $\theta < \hat{\theta}$, the free best-effort quality decreases as we have $\partial q^n(\hat{\theta})/\partial\hat{\theta} > 0$, resulting in lower profits for CPs with low quality sensitivity. CPs who purchased prioritization now have their content delivered at higher quality, however, the content delivery is no longer free of charge. The content class which is indifferent between prioritization and best-effort quality $\hat{\theta}$ is worse off under prioritization, as the best-effort quality decreases compared to a neutral regime. Only those CPs with very high quality sensitivity potentially benefit from prioritization. However, in total the content industry is worse off under prioritization. When it comes to the comparison between a paid prioritization regime and a CDN based model this result predicts that CPs are indifferent between the two as the outcome is equivalent.

Proposition 4 *ISPs do not benefit from prioritization.*

$$S_{ISP}(p^n, f^n, \hat{\theta}^n) > S_{ISP}(p^c, f^c, \hat{\theta}^c) > S_{ISP}(p^p, f^p, \hat{\theta}^p) \tag{29}$$

Proof. See Appendix. ■

Even though prioritization opens up new revenue streams on the CP side, the induced competition dynamics on the consumer market leads to lower industry profits. Consumer prices decrease as the consumer market becomes more elastic and losing consumers now has additional negative effects on the CP side of the business. This reduction in revenue outweighs any additional revenue which can be obtained from selling prioritization to CPs, resulting in lower ISP industry profits under prioritization. In the case of CDNs the ability to raise consumer prices is less restricted compared to the paid prioritization case resulting in higher industry profits in the presence of CDNs compared to paid prioritization. However, ISPs would be better off if they would not introduce prioritization offers even if they would be allowed to do so.

Corollary 1 *ISPs face a prisoner's dilemma when deciding whether to offer prioritization.*

Proof. Follows from Lemma 1 and Proposition 4. ■

As ISPs have an unilateral incentive to introduce prioritization offers (see Lemma 1), they end up in a situation where competition for consumers is strengthened to such an extent, that the negative effect on the consumer market outweighs the additional revenues made on the CP market side. This result supports the finding in Bourreau et al. (2015). Delegating the prioritization business to a CDN can then be seen as a remedy to soften competition on the consumer market.

Proposition 5 *Welfare is higher under prioritization*

$$TS(\hat{\theta}^p) = TS(\hat{\theta}^c) > TS(\hat{\theta}^n)$$

Proof. Follows from Lemma 3. ■

As prices are transfers from consumers to ISPs, and fees from CPs to ISPs / CDN, the welfare comparison boils down to the aggregate network quality. Under a prioritization regime the existing network capacity is allocated more efficiently, resulting in a higher total surplus.

4.2 Investment incentives

In this section we want to shed light on how the different QoS regimes affect investment in network infrastructure. For this we compare investment incentives from a symmetric equilibrium perspective. The idea is that capacity investments are typically long-term decisions such that the industry is in equilibrium before the next investment decisions are made. We assume that investment costs for capacity expansion are identical in all regimes and therefore restrict our analysis to the comparison of marginal profits gross of investment costs. The changes of p_i, f_i, N_i and Y_i with respect to k_i are again obtained by applying the implicit function theorem, while we evaluate all expressions at respective equilibrium values which allows us to make use of the envelope theorem for simplification. Detailed derivations can be found in Appendix A.

Starting with the neutral regime we obtain

$$\frac{d\Pi_i}{dk_i} = p_i \overbrace{\left(\underbrace{\frac{\partial N_i}{\partial k_i}}_{+} + \underbrace{\frac{\partial N_i}{\partial p_j}}_{+} \underbrace{\frac{\partial p_j}{\partial k_i}}_{-} \right)}^{\text{Consumer effect}} > 0. \tag{30}$$

The marginal profit of capacity investment mainly depends on the direct effect $\partial N_i/\partial k_i > 0$ of investment in network quality and thereby attracting consumers, and the strategic effect $\partial p_j/\partial k_i < 0$ of network j in order to recapture lost market share by decreasing prices. The former effect outweighs the latter, such that the overall effect is positive, and we will refer to the overall effect simply as 'consumer effect'.

Turning to the paid prioritization case we obtain investment incentives of

$$\frac{d\Pi_i}{dk_i} = p_i \overbrace{\left(\underbrace{\frac{\partial N_i}{\partial k_i}}_{+} + \underbrace{\frac{\partial N_i}{\partial p_j}}_{+} \underbrace{\frac{\partial p_j}{\partial k_i}}_{-} \right)}^{\text{Consumer effect}} + f_i \overbrace{\left(\underbrace{\frac{\partial Y_i}{\partial k_i}}_{-} + \underbrace{\frac{\partial Y_i}{\partial p_j}}_{+} \underbrace{\frac{\partial p_j}{\partial k_i}}_{-} \right)}^{\text{CP effect}} > 0. \tag{31}$$

The dynamics behind the consumer effect are identical as in the neutral regime but differ in magnitude (see detailed discussion below). The main difference is that we now have an additional effect on the CP market side which we will refer to as 'CP effect'. We again distinguish two different sub-effects: A direct effect $\partial Y_i/\partial k_i < 0$ and a strategic effect $\partial p_j/\partial k_i < 0$. As a capacity increase eases the congestion problem, less CPs opt for prioritization, resulting in a negative direct effect. Similar to the consumer effect network j reacts by lowering consumer prices, reducing the market share of network i and thereby making prioritization in network i even less attractive, resulting in a second negative (strategic) effect. As the business model of prioritization relies on a congestion problem, investment in capacity expansion directly reduces the obtainable profit from the CP side of the market. In total the positive consumer effect, however, outweighs the negative CP effect, resulting in positive investment incentives.

For the CDN case we need to take into account the effect of the investment decision on the business relationship with the CDN. It turns out to be helpful to denote the lump-sum transfer F^c as a fraction $\alpha \in [0,1]$ of the CDN profit where $\alpha := F^c/\Pi_c$. The investment incentive of an ISP can then be written as $d\tilde{\Pi}_i/dk_i$ where $\tilde{\Pi}_i := \Pi_i + \alpha\Pi_c$ such that $d\tilde{\Pi}_i/dk_i$ is given by

$$\frac{d\tilde{\Pi}_i}{dk_i} = p_i \overbrace{\left(\underbrace{\frac{\partial N_i}{\partial k_i}}_{+} + \underbrace{\frac{\partial N_i}{\partial p_j}}_{+} \underbrace{\frac{\partial p_j}{\partial k_i}}_{-} \right)}^{\text{Consumer effect}} + \alpha \overbrace{\left(f_i \underbrace{\frac{\partial Y_i}{\partial k_i}}_{-} + f_j \underbrace{\frac{\partial Y_j}{\partial k_i}}_{-} \right)}^{\text{CP effect}} + \overbrace{\left(\underbrace{\frac{d\alpha}{dk_i}}_{>0} \Pi_c^c \right)}^{\text{CDN effect}} > 0. \tag{32}$$

We now observe three separate effects: a positive consumer effect, a negative CP effect and a positive 'CDN effect'. The dynamics in the consumer effect are as before, however, the CP effect now consists only of two direct effects. On the one hand a capacity increase in network i reduces the congestion problem and hence the benefit from prioritization $\partial Y_i/\partial k_i < 0$. On the other hand it also attracts consumers from network j to join network i and makes thereby prioritization in network j less attractive $\partial Y_j/\partial k_i < 0$, resulting again in a negative CP effect in total. The CDN effect reflects the fact that as capacity increases the ISP obtains a larger share of the CDN profits. This has mainly two reasons. First, increasing capacity k_i reduces CDN profits $d\Pi_c/dk_i < 0$ (see CP effect). Secondly, increasing capacity increases the outside option of an ISP, resulting in a higher share of obtainable CDN profits and a positive CDN effect.[15]

In figure 1 we now illustrate the magnitude of the different effects for various levels of initial symmetric network capacity k. Proposition 6 summarizes the main findings of the illustrated results.

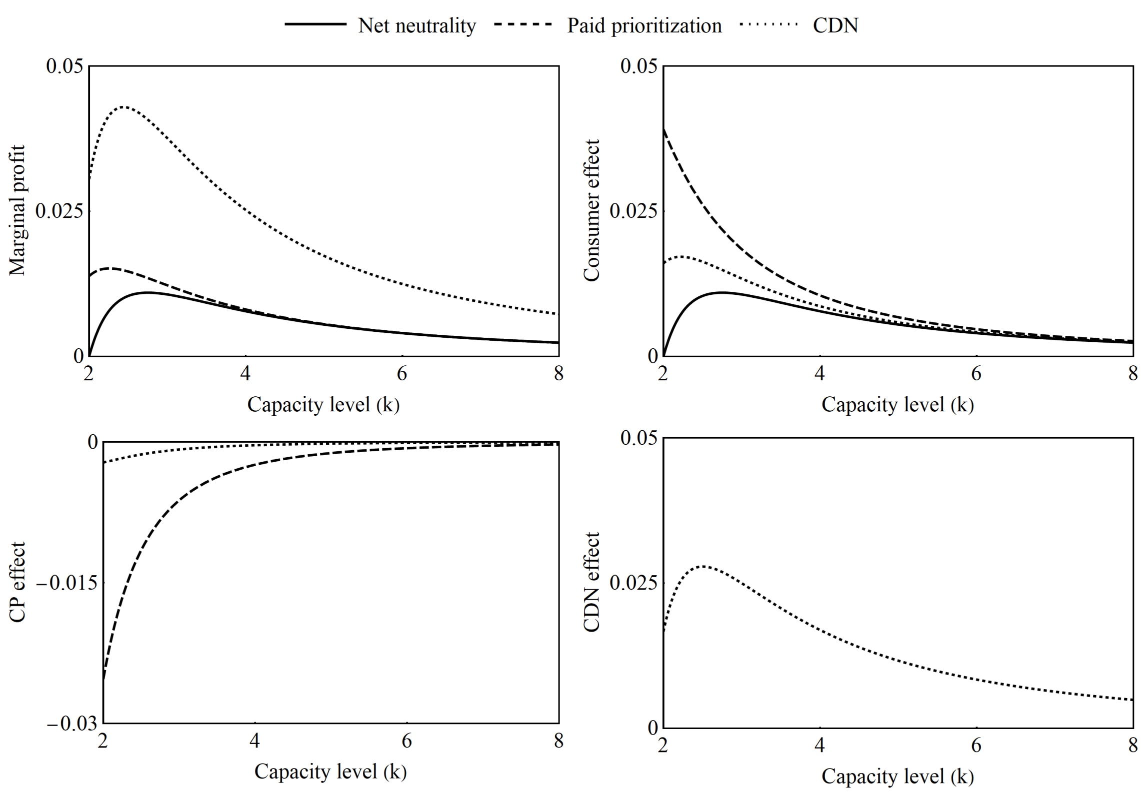

Figure 1: Comparison of investment incentives

[15]We refer to the proof of Proposition 6 for details.

Proposition 6 *Incentives to invest in network capacity are highest in the CDN case. Paid prioritization leads to higher investment incentives than a neutral regime if capacity is scarce* $k \leq \overline{k}$.

Proof. See Appendix. ■

First, note that the consumer effect is positive in all regimes while the CP effect is negative in the two prioritization regimes. In the CDN case there is an additional positive CDN effect. Those effects, however, differ in magnitude.

Starting with the consumer effect we would first like to mention that the direct effect $\partial N_i/\partial k_i$ is strongest in the neutral regime. Since a tiered quality scheme already reduces the congestion problem, the marginal effect of capacity expansion is higher when the congestion problem is severe, as in the single-queue (neutral) regime. This means, however, that the strategic effect introduced by $\partial p_j/\partial k_i$ must be the driving force behind the ranking in magnitude displayed in the top right graph of figure 1. As explained in section 3.2, prioritization leads to a more elastic consumer market. Hence, the market share reaction with respect to rival's prices $\partial N_i/\partial p_j$ is more strongly pronounced in the prioritization regimes. For the same reason, however, the strategic response by ISP j to an increase in capacity k_i is less pronounced in the prioritization regimes. As the consumer market share is more elastic, prices p_j are decreased to a lower extent than in the neutral regime. The strategic effect combined with a more elastic consumer market results in a stronger consumer effect in the prioritization regimes.

The CP effect is negative in both prioritization cases, however, their composition differs. While under paid prioritization only the direct effect on the proprietary network $\partial Y_i/\partial k_i < 0$ is taken into account, in the CDN solution direct effects of both networks are taken into account. The main driver for the difference in magnitude is, however, the weighting factor α in the CDN case, such that in a CDN environment ISPs do not fully internalize the negative effect on the CP market side when deciding on investment in network capacity. In addition to the less pronounced CP effect, ISPs obtain an additional positive CDN effect resulting in highest total marginal profits from capacity investment.

The comparison between the neutral and the paid prioritization regime depends on initial capacity levels.[16] If capacity is scarce $k \leq \overline{k}$ the stronger consumer effect in the discriminatory regime dominates the neutral regime even in light of the negative CP effect. If capacity is abundant $k > \overline{k}$ on the other hand, consumer effects are virtually identical as high overall capacity makes prioritization irrelevant such that the negative CP effect prevails yielding higher investment incentives in the neutral regime. In light of existing QoS differentiation measures and global efforts to incentivize broadband investment, we consider the case of scarcity to be more relevant.

5 Conclusion

We analyzed equilibrium outcomes under different QoS practices and showed that discriminatory regimes are superior in terms of static efficiency as they allocate existing capacity more efficiently while at the same time competition for consumers is strengthened, resulting in lower consumer prices and higher network quality in both discriminatory regimes compared to a neutral regime. The extent to which consumers benefit, however, depends on the way how prioritization is achieved. While prices are lowest under paid prioritization, consumer prices increase with the use of CDNs as ISPs lack the additional incentive to attract consumers to make prioritization more valuable.

Regarding investment incentives we showed that both discriminatory regimes lead to higher investment in network capacity than the neutral regime as long as capacity is scarce, while investment is highest in a CDN environment irrespective of the initial capacity level. Under paid prioritization marginal profits obtained from the consumer market side are higher than in the neutral regime, while marginal profits obtained from the CP side are negative as capacity expansion makes prioritization less valuable. In a CDN scenario this detrimental effect on the CP side is not fully internalized while at the same time ISPs are able to capture a larger fraction of CDN profits when network capacity is expanded, resulting in high investment incentives.

We would like to mainly draw two policy conclusions where the first is driven by our efficiency result. As long as content is heterogeneous and network capacity is scarce, a tiered-quality scheme increases efficiency. This result is not driven by the assumption that

[16]The critical capacity level is given by $\overline{k} \approx 6.45$. Details can be found in the proof of Proposition 6.

total demand on the consumer market is inelastic, as a discriminatory regime simultaneously reduces consumer prices. Also, since the best-effort quality remains free of charge no CPs are excluded from the market. For the second conclusion one should note that the general debate on net neutrality tends to focus on ISP practices, while the use of CDNs is barely mentioned. Our results suggest that while the outcome with CDNs is welfare equivalent to the classical paid prioritization, consumer surplus is lower in the presence of CDNs due to higher prices. Focusing on static efficiency and having consumer welfare in mind, a regime of paid prioritization is therefore to be preferred. If the primary policy goal is, however, investment in network infrastructure then our results suggest that a CDN environment is to be preferred over paid prioritization.

We would also like to point out limitations of our analysis and where future research could be headed. First, we implicitly assume that from a technical perspective contracting with a CDN is equivalent to direct prioritization between ISPs and CPs. Here, a more nuanced analysis could refine the comparison with respect to efficiency. Also, we modeled the contractual relationship between ISPs and CDNs in rather general way. Here, industry specific payment structures (access pricing, etc.) could provide further insights. Lastly, one could alter the industry structure in the upstream market and introduce competition between CDNs.

Appendix

A Omitted analysis

Market share reactions

For the consumer market share we define an ancillary equation

$$\Delta_N = N_1 - \hat{x} \tag{33}$$

where $\hat{x}$ denotes the indifferent consumer on the Hotelling line as in (11). To obtain market share reactions $\partial N_i/\partial p_i$ we then totally differentiate $\Delta_N = 0$ with respect to consumer prices p_i while $\partial N_j/\partial p_i = -\partial N_i/\partial p_i$ for $j \neq i$ due to full market coverage. Market share reactions are therefore given by

$$\frac{\partial N_i}{\partial p_i} = -\frac{\partial \Delta_N/\partial p_i}{\partial \Delta_N/\partial N_i}. \tag{34}$$

In the case of a prioritization regime we additionally define ancillary equations for the share of prioritized content:

$$\Delta_{Y_i} = Y_i - \left(1 - \hat{\theta}_i\right),\ i = \{1, 2\}. \tag{35}$$

Reactions with respect to consumer prices $\partial Y_i/\partial p_j, \partial N_i/\partial p_j$ and prioritization fees $\partial Y_i/\partial f_j, \partial N_i/\partial f_j$ for $i, j = \{1, 2\}$ are then obtained by totally differentiating equations $\Delta_N = 0$, $\Delta_{Y_1} = 0$ and $\Delta_{Y_2} = 0$. Market share reactions $\partial Y_i/\partial p_j, \partial N_i/\partial p_j$ are then determined by the solution to

$$\frac{d\Delta_Z}{dp_i} = \frac{\partial \Delta_Z}{\partial p_i} + \frac{\partial \Delta_Z}{\partial N_i}\frac{\partial N_i}{\partial p_i} + \frac{\partial \Delta_Z}{\partial Y_i}\frac{\partial Y_i}{\partial p_i} + \frac{\partial \Delta_Z}{\partial Y_j}\frac{\partial Y_j}{\partial p_i} = 0 \tag{36}$$

where $\Delta_Z = \{\Delta_N, \Delta_{Y_i}, \Delta_{Y_j}\}$ and $i, j = \{1, 2\}$. Market share reactions with respect to prioritization fees $\partial Y_i/\partial f_j, \partial N_i/\partial f_j$ can be obtained by an equivalent procedure.

Investment incentives

First we outline how we obtain reactions $\partial N_j/\partial k_i, \partial Y_j/\partial k_i, \partial p_j/\partial k_i$ and $\partial f_j/\partial k_i$ for $j = \{1, 2\}$. We will demonstrate the procedure in the paid prioritization case as it can easily be adjusted to yield the reactions in the other regimes. We make extensive use of the implicit function theorem by totally differentiating the first order conditions of ISPs i and j as well as the ancillary equations

$\Delta_N = 0$, $\Delta_{Y_i} = 0$ and $\Delta_{Y_j} = 0$ with respect to k_i. The result is the following system of equations where for $Z = \{\Delta_N, \Delta_{Y_i}, \Delta_{Y_j}, \partial\Pi_i/\partial p_i, \partial\Pi_i/\partial f_i, \partial\Pi_j/\partial p_j, \partial\Pi_j/\partial f_j\}$ we have

$$\begin{aligned}\frac{dZ}{dk_i} &= \frac{\partial Z}{\partial k_i} + \frac{\partial Z}{\partial p_i}\frac{\partial p_i}{\partial k_i} + \frac{\partial Z}{\partial p_j}\frac{\partial p_j}{\partial k_i} + \frac{\partial Z}{\partial f_i}\frac{\partial f_i}{\partial k_i} + \frac{\partial Z}{\partial f_j}\frac{\partial f_j}{\partial k_i} \\ &\quad + \frac{\partial Z}{\partial N_1}\frac{dN_1}{dk_i} + \frac{\partial Z}{\partial Y_i}\frac{dY_i}{dk_i} + \frac{\partial Z}{\partial Y_j}\frac{dY_j}{dk_i} = 0, \end{aligned} \tag{37}$$

where for $W = \{N_1, Y_i, Y_j\}$ we have

$$\frac{dW}{dk_i} = \frac{\partial W}{\partial k_i} + \frac{\partial W}{\partial p_i}\frac{\partial p_i}{\partial k_i} + \frac{\partial W}{\partial p_j}\frac{\partial p_j}{\partial k_i} + \frac{\partial W}{\partial f_i}\frac{\partial f_i}{\partial k_i} + \frac{\partial W}{\partial f_j}\frac{\partial f_j}{\partial k_i}. \tag{38}$$

We now turn to the definition of investment incentives. Starting with the neutral regime, marginal profits $d\Pi_i/dk_i$ with $\Pi_i = p_i N_i$ can be written as

$$\frac{d\Pi_i}{dk_i} = \frac{\partial \Pi_i}{\partial p_i}\frac{\partial p_i}{\partial k_i} + \frac{\partial \Pi_i}{\partial N_i}\left(\frac{\partial N_i}{\partial k_i} + \frac{\partial N_i}{\partial p_i}\frac{\partial p_i}{\partial k_i} + \frac{\partial N_i}{\partial p_j}\frac{\partial p_j}{\partial k_i}\right).$$

Imposing symmetry and evaluating at equilibrium values $p_i = p_j = p^n$ allows us to further make use of the envelope theorem, yielding the final expression.

$$\frac{d\Pi_i}{dk_i} = \underbrace{\left(N_i + p_i\frac{\partial N_i}{\partial p_i}\right)}_{=0}\frac{\partial p_i}{\partial k_i} + p_i\left(\frac{\partial N_i}{\partial k_i} + \frac{\partial N_i}{\partial p_j}\frac{\partial p_j}{\partial k_i}\right). \tag{39}$$

Similarly, in the case of paid prioritization with $\Pi_i = p_i N_i + f_i Y_i$ we obtain

$$\begin{aligned}\frac{d\Pi_i}{dk_i} &= \frac{\partial \Pi_i}{\partial p_i}\frac{\partial p_i}{\partial k_i} + \frac{\partial \Pi_i}{\partial N_i}\left(\frac{\partial N_i}{\partial k_i} + \frac{\partial N_i}{\partial p_i}\frac{\partial p_i}{\partial k_i} + \frac{\partial N_i}{\partial p_j}\frac{\partial p_j}{\partial k_i} + \frac{\partial N_i}{\partial f_i}\frac{\partial f_i}{\partial k_i} + \frac{\partial N_i}{\partial f_j}\frac{\partial f_j}{\partial k_i}\right) \\ &\quad + \frac{\partial \Pi_i}{\partial f_i}\frac{\partial f_i}{\partial k_i} + \frac{\partial \Pi_i}{\partial Y_i}\left(\frac{\partial Y_i}{\partial k_i} + \frac{\partial Y_i}{\partial p_i}\frac{\partial p_i}{\partial k_i} + \frac{\partial Y_i}{\partial p_j}\frac{\partial p_j}{\partial k_i} + \frac{\partial Y_i}{\partial f_i}\frac{\partial f_i}{\partial k_i} + \frac{\partial Y_i}{\partial f_j}\frac{\partial f_j}{\partial k_i}\right). \end{aligned} \tag{40}$$

Imposing symmetry $k_i = k_j = k$, evaluating at equilibrium values $p_i = p_j = p^p$, $f_i = f_j = f^p$, and applying the envelope theorem yields

$$\begin{aligned}\frac{d\Pi_i}{dk_i} &= \underbrace{\left(N_i + p_i\frac{\partial N_i}{\partial p_i} + f_i\frac{\partial Y_i}{\partial p_i}\right)}_{=0}\frac{\partial p_i}{\partial k_i} + p_i\left(\frac{\partial N_i}{\partial k_i} + \frac{\partial N_i}{\partial p_j}\frac{\partial p_j}{\partial k_i} + \underbrace{\frac{\partial N_i}{\partial f_j}}_{=0}\frac{\partial f_j}{\partial k_i}\right) \\ &\quad + \underbrace{\left(Y_i + p_i\frac{\partial N_i}{\partial f_i} + f_i\frac{\partial Y_i}{\partial f_i}\right)}_{=0}\frac{\partial f_i}{\partial k_i} + f_i\left(\frac{\partial Y_i}{\partial k_i} + \frac{\partial Y_i}{\partial p_j}\frac{\partial p_j}{\partial k_i} + \underbrace{\frac{\partial Y_i}{\partial f_j}}_{=0}\frac{\partial f_j}{\partial k_i}\right). \end{aligned} \tag{41}$$

Using $\tilde{\Pi}_i := \Pi_i + \alpha\Pi_c$ with $\Pi_i = p_i N_i$, $\Pi_c = f_i Y_i + f_j Y_j$ and $\alpha = F^c/\Pi_c$ in the CDN case we obtain

$$
\begin{aligned}
\frac{d\tilde{\Pi}_i}{dk_i} &= \frac{d\Pi_i}{dk_i} + \alpha\frac{d\Pi_c}{dk_i} + \frac{d\alpha}{dk_i}\Pi_c \\
&= \frac{\partial \Pi_i}{\partial k_i}\frac{\partial p_i}{\partial k_i} + \frac{\partial \Pi_i}{\partial N_i}\left(\frac{\partial N_i}{\partial k_i} + \frac{\partial N_i}{\partial p_i}\frac{\partial p_i}{\partial k_i} + \frac{\partial N_i}{\partial p_j}\frac{\partial p_j}{\partial k_i} + \frac{\partial N_i}{\partial f_i}\frac{\partial f_i}{\partial k_i} + \frac{\partial N_i}{\partial f_j}\frac{\partial f_j}{\partial k_i}\right) \\
&+ \alpha\left[\frac{\partial \Pi_c}{\partial f_i}\frac{\partial f_i}{\partial k_i} + \frac{\partial \Pi_c}{\partial Y_i}\left(\frac{\partial Y_i}{\partial k_i} + \frac{\partial Y_i}{\partial p_i}\frac{\partial p_i}{\partial k_i} + \frac{\partial Y_i}{\partial p_j}\frac{\partial p_j}{\partial k_i} + \frac{\partial Y_i}{\partial f_i}\frac{\partial f_i}{\partial k_i} + \frac{\partial Y_i}{\partial f_j}\frac{\partial f_j}{\partial k_i}\right)\right. \\
&+ \left.\frac{\partial \Pi_c}{\partial f_j}\frac{\partial f_j}{\partial k_i} + \frac{\partial \Pi_c}{\partial Y_j}\left(\frac{\partial Y_j}{\partial k_i} + \frac{\partial Y_j}{\partial p_i}\frac{\partial p_i}{\partial k_i} + \frac{\partial Y_j}{\partial p_j}\frac{\partial p_j}{\partial k_i} + \frac{\partial Y_j}{\partial f_i}\frac{\partial f_i}{\partial k_i} + \frac{\partial Y_j}{\partial f_j}\frac{\partial f_j}{\partial k_i}\right)\right] \\
&+ \frac{d\alpha}{dk_i}\Pi_c. \qquad (42)
\end{aligned}
$$

Evaluating at equilibrium values $p_i = p_j = p^c$, $f_i = f_j = f^c$ for the symmetric case $k_i = k_j = k$ and applying the envelope theorem yields

$$
\begin{aligned}
\frac{d\tilde{\Pi}_i}{dk_i} &= \underbrace{\left(N_i + p_i\frac{\partial N_i}{\partial p_i}\right)}_{=0}\frac{\partial p_i}{\partial k_i} + p_i\left(\frac{\partial N_i}{\partial k_i} + \frac{\partial N_i}{\partial p_j}\frac{\partial p_j}{\partial k_i} + \underbrace{\frac{\partial N_i}{\partial f_i}}_{=0}\frac{\partial f_i}{\partial k_i} + \underbrace{\frac{\partial N_i}{\partial f_j}}_{=0}\frac{\partial f_j}{\partial k_i}\right) \\
&+ \alpha\left[\underbrace{\left(Y_i + f_i\frac{\partial Y_i}{\partial f_i} + f_j\frac{\partial Y_j}{\partial f_i}\right)}_{=0}\frac{\partial f_i}{\partial k_i} + f_i\left(\frac{\partial Y_i}{\partial k_i} + \frac{\partial Y_i}{\partial p_i}\frac{\partial p_i}{\partial k_i} + \frac{\partial Y_i}{\partial p_j}\frac{\partial p_j}{\partial k_i}\right)\right. \\
&+ \left.\underbrace{\left(Y_j + f_i\frac{\partial Y_i}{\partial f_j} + f_j\frac{\partial Y_j}{\partial f_j}\right)}_{=0}\frac{\partial f_j}{\partial k_i} + f_j\left(\frac{\partial Y_j}{\partial k_i} + \frac{\partial Y_j}{\partial p_i}\frac{\partial p_i}{\partial k_i} + \frac{\partial Y_j}{\partial p_j}\frac{\partial p_j}{\partial k_i}\right)\right] \\
&+ \frac{d\alpha}{dk_i}\Pi_c. \qquad (43)
\end{aligned}
$$

Investment incentives in the CDN case can then be written as:

$$
\begin{aligned}
\frac{d\tilde{\Pi}_i}{dk_i} &= p_i\left(\frac{\partial N_i}{\partial k_i} + \frac{\partial N_i}{\partial p_j}\frac{\partial p_j}{\partial k_i}\right) \\
&+ \alpha\left[f_i\left(\frac{\partial Y_i}{\partial k_i} + \frac{\partial Y_i}{\partial p_i}\frac{\partial p_i}{\partial k_i} + \frac{\partial Y_i}{\partial p_j}\frac{\partial p_j}{\partial k_i}\right) + f_j\left(\frac{\partial Y_j}{\partial k_i} + \frac{\partial Y_j}{\partial p_i}\frac{\partial p_i}{\partial k_i} + \frac{\partial Y_j}{\partial p_j}\frac{\partial p_j}{\partial k_i}\right)\right] \\
&+ \frac{d\alpha}{dk_i}\Pi_c. \qquad (44)
\end{aligned}
$$

Finally, using $\partial Y_i/\partial p_i = -\partial Y_j/\partial p_i$ for $i \neq j$ we obtain the final simplified form. Furthermore, using $\alpha = F^c/\Pi_c = (\Pi_i^A - \Pi_i^c)/\Pi_c^c$ we obtain

$$
\frac{d\alpha}{dk_i} = \frac{\partial \alpha}{\partial \Pi_i^A}\frac{d\Pi_i^A}{dk_i} + \frac{\partial \alpha}{\partial \Pi_i^c}\frac{d\Pi_i^c}{dk_i} + \frac{\partial \alpha}{\partial \Pi_c^c}\frac{d\Pi_c^c}{dk_i} \qquad (45)
$$

where $d\Pi_i^c/dk_i$ and $d\Pi_c^c/dk_i$ are derived above, while $d\Pi_i^A/dk_i$ is obtained precisely as in (37) and (38) applied to the asymmetric game structure outlined in the proof of Lemma 2.

B Omitted proofs

Proof of Lemma 1

Proof. To prove the result define $\overline{f}_i$ such that

$$\hat{\theta}_i = \frac{\overline{f}_i N_i}{N_i(q_i^p - q_i^n)} = 1.$$

Then, solving $\partial\Pi_i/\partial p_i = 0$ for p_i and substituting we get (independently of $\widehat{\theta}_j \in (0,1)$ or $\widehat{\theta}_j = 1$)

$$\left.\frac{\partial \Pi_i}{\partial f_i}\right|_{f_i=\overline{f}_i} = -\frac{3k}{2(k-\widehat{x})} < 0$$

where $\widehat{x}$ denotes the indifferent consumer obtained as in (11). Hence, decreasing $\overline{f}_i$ and thus offering prioritization would lead to higher profits Π_i. ■

Proof of Lemma 2

Proof. We consider three different sub-game outcomes: a symmetric outcome where both ISPs contract with the CDN, an asymmetric outcome where only one ISP contracts with the CDN and the case where no ISP contracts with the CDN. The last case leads to zero profits for the CDN such that it suffices to show that profits in the other two cases are weakly positive to exclude this case.

To ease notation let $\Pi_c^c = 2f^cY^c$ and $\Pi_i^c = p^c/2$ denote the equilibrium profits of the sub-game outlined in section 3.3, and $\Pi_i^p = p^p/2 + f^pY^p$ be the equilibrium ISP profit from section 3.2. Further consider the following asymmetric game where ISP j delegates prioritization to the CDN, while ISP i does not contract with the CDN. As the lump-sum transfers do not affect the price choice in the subsequent simultaneous move game, we can solve for the asymmetric solution $(p_i^A, p_j^A, f_i^A, f_j^A)$ by following the same steps as outlined in section 3.3 such that

$$(p_i^A, f_i^A) := \arg\max_{p_i, f_i} \Pi_i^A = p_iN_i + f_iY_i|_{p_j=p_j^A, f_j=f_j^A}, \tag{46}$$

$$p_j^A := \arg\max_{p_j} \Pi_j^A = p_jN_j|_{p_i=p_i^A, f_i=f_i^A, f_j=f_j^A}, \tag{47}$$

$$f_j^A := \arg\max_{f_j} \Pi_c^A = f_jY_j|_{p_i=p_i^A, p_j=p_j^A, f_i=f_i^A}. \tag{48}$$

Equilibrium values are then given by

$$f_i^A = \frac{kN_i^A\left(-k+N_i^A+\sqrt{k(k-N_i^A)}\right)}{(k(k-N_i^A))^{3/2}}, \tag{49}$$

$$f_j^A = \frac{(N_i^A-1)^2}{(k+N_i^A-1)\left(k+\sqrt{k(k+N_i^A-1)}\right)}, \tag{50}$$

$$p_j^A = \frac{1-N_i^A}{N_i^A}\left(p_i^A+f_i^A\frac{k}{(k-N_i^A)N_i^A}\right), \tag{51}$$

$$\begin{aligned} p_i^A =& \frac{1}{N_i^A}\left(1-\frac{k}{\sqrt{k(k-N_i^A)}}\right)+\frac{-2k+2N_i^A+3\sqrt{k(k-N_i^A)}}{2(k-N_i^A)^2} \\ &+\frac{1}{2}N_i^A\left(\frac{(-2k-3N_i^A+3)k^2}{(N_i^A-1)^2(k(k+N_i^A-1))^{3/2}}-\frac{2}{(k-N_i^A)^2}+\frac{2}{(N_i^A-1)^2}+4\right). \end{aligned} \tag{52}$$

resulting in

$$Y_i^A = \frac{k-\sqrt{k(k-N_i^A)}}{N_i^A} \quad \text{and} \quad Y_j^A = 1-\frac{k+N_i^A-1}{\sqrt{k(k+N_i^A-1)}+k+N_i^A-1} \tag{53}$$

while N_i^A is implicitly defined by $\Delta_{N^A}=0$ with

$$\begin{aligned} \Delta_{N^A} =& \frac{k-2\sqrt{k(k-N_i^A)}}{(k-N_i^A)^2} \\ &+\frac{1}{1-N_i^A}\left[4+\frac{N_i^A(k+2N_i^A-2)k^2}{(1-N_i^A)(k(k+N_i^A-1))^{3/2}}-3N_i^A+\frac{\sqrt{k(k+N_i^A-1)}}{2(k+N_i^A-1)^2}-\frac{1}{1-N_i^A}\right] \\ &+\frac{1}{(N_i^A)^2}\left[1-6(N_i^A)^3+\frac{1}{2}\left(\frac{k^2(2k-2N_i^A+1)}{(k(k-N_i^A))^{3/2}}-2\right)N_i^A-\frac{k}{\sqrt{k(k-N_i^A)}}\right]. \end{aligned} \tag{54}$$

We denote the payoffs evaluated at the asymmetric equilibrium solution simply as Π_i^A, Π_j^A and Π_c^A.

Starting with the symmetric outcome outlined in section 3.3, an ISP will not deviate from the acceptance of an offer F^c if $\Pi_i^c + F^c \geq \Pi_i^A$ such that the optimal offer from the point of view of an CDN is given by $F^c = \Pi_i^A - \Pi_i^c$. Also, the CDN's participation constraint must be satisfied such that $\Pi^c - 2F^c \geq 0$. Substituting F^c and rearranging we obtain the sufficient condition

$$\Pi_c^c + 2\Pi_i^c - 2\Pi_i^A \geq 0 \tag{55}$$

for the existence of an offer F^c which is accepted by both ISPs.

Turning to the asymmetric case we need acceptance of an offer $\hat{F}$ such that $\Pi_j^A + \hat{F} \geq \Pi_i^p$ where $\hat{F} = \Pi_i^p - \Pi_j^A$ is the lowest offer accepted by one ISP, while the offer to the other ISP can simply be set to $-\infty$ to make sure that only one ISP contracts with the CDN. Similarly, we need to make sure that the participation constraint of the CDN is satisfied such that $\Pi_c^A - \hat{F} \geq 0$ or after substituting

$$\Pi_c^A + \Pi_j^A - \Pi_i^p \geq 0. \tag{56}$$

In order for the symmetric case to be sub-game perfect we additionally require that the achievable payoff from contracting with both ISPs is higher than in the asymmetric case such that $\Pi^c - 2F^c \geq \Pi_c^A - \hat{F}$ or after substituting and rearranging

$$\Pi_c^c + 2\Pi_i^c + \Pi_i^p - \Pi_c^A - 2\Pi_i^A - \Pi_j^A \geq 0. \tag{57}$$

Conditions (55) - (57) are sufficient to prove existence of symmetric fees F^c which are accepted by both ISPs and satisfy the participation constraint of the CDN, and furthermore these conditions assure that the symmetric outcome is preferred to the asymmetric case, assuring that the equilibrium of the sub-game characterized in section 3.3 is indeed sub-game perfect. As all three conditions depend on the asymmetric solution, we address the implicit definition of N_i^A in (54) by evaluating equilibrium payoffs Π_c^A, Π^A and Π_j^A for a given level k at the root to $\Delta_{N^A} = 0$. We can then perform numerical analysis for arbitrary values of k to verify that all three conditions are satisfied. An illustration of the numerical analysis can be seen in the following figure where 'sym' refers to the LHS of (55), 'asym' to the LHS of (56) and 'comp' to the LHS of (57).

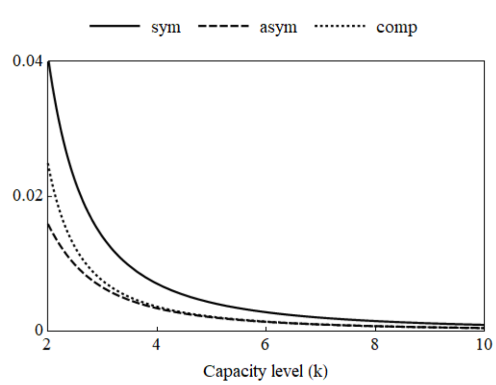

Figure 2: Illustration of conditions (55)-(57)

Following the same procedure one can verify that for given offers $F_i = F_j = F^c$ there also exists the reject/reject equilibrium in the offer stage which leads to payoffs Π_i^p for both ISPs. This equilibrium is payoff dominated by the accept/accept equilibrium and therefore is subject to our refinement. ■

Proof of Lemma 3

Proof. This result follows from the definition of $V(\hat{\theta})$ in (25). After basic simplifications we obtain

$V(\hat{\theta}) = 3/2 - \hat{\theta}/(2k-1) - 2k/(2k-(1-\hat{\theta}))$ and $V(1) = (2k-3)/(4k-2)$. Comparing $V(\hat{\theta}) > V(1)$ then reduces to $(1-\hat{\theta}) > 0$, which is satisfied for $\hat{\theta} \in (0,1)$. ■

Proof of Proposition 1

Proof. As the consumer market is inelastic, the equivalence follows directly from $\hat{\theta}^c = \hat{\theta}^p$. ■

Proof of Proposition 2

Proof. The result in i.) follows from Lemma 3. For ii.) we rearrange $p^n > p^c$ to $\psi > (8k^2 - 8k + 1)/(4k-3)$. Squaring both sides and basic simplification steps lead to $(4k-3)^2(4k^2 - 2k - 1) > 0$ which is true by Assumption 1. For $p^c > p^p$ we immediately see that $p^c - p^p = f^p 4k/(2k-1) > 0$. Point iii.) follows directly from i.) and ii.). ■

Proof of Proposition 3

Proof. The inequality $S_{CP}(f^p, \hat{\theta}^p) = S_{CP}(f^c, \hat{\theta}^c) < S_{CP}(f^n, \hat{\theta}^n)$ reduces to $2\hat{\theta}^c = 2\hat{\theta}^p < 1$. As $\partial\hat{\theta}^p/\partial k = \partial\hat{\theta}^c/\partial k > 0$ it follows from $\lim_{k\to\infty} \hat{\theta}^p = 1/2$ that $2\hat{\theta}^p = 2\hat{\theta}^c < 1$. ■

Proof of Proposition 4

Proof. First note that $S_{ISP}(p^c, f^c, \hat{\theta}^c) > S_{ISP}(p^p, f^p, \hat{\theta}^p)$ follows from the proof of Proposition 2 as it boils down to the difference in consumer prices. The second inequality $S_{ISP}(p^n, f^n, \hat{\theta}^n) > S_{ISP}(p^c, f^c, \hat{\theta}^c)$ can be rearranged to $2k(1 + 12k^2 + 5\psi + \psi^2) < \psi^2 + 4k^2(5 + 4\psi)$. Substituting $\psi^2 = 2k(2k-1)$ and rearranging yields $(16k^2 - 14k + 2)/(8k-5) < \psi$. Squaring both sides and rearranging yields $(5-8k)^2(3k-2) > 0$, which is true. ■

Proof of Proposition 5

Proof. Follows from Lemma 3 and Proposition 2. ■

Proof of Proposition 6

Proof.

Following the procedure outlined in Appendix A we obtain closed form solutions for investment incentives in all three QoS regimes. In the case of net neutrality and paid prioritization the investment incentives only depend on the initial symmetric capacity level k. As investment incentives in the CDN case partly depend on the asymmetric solution outlined in the proof of Lemma 2), we obtain a final form depending on k and an asymmetric market share N_i^A which is implicitly defined by $\Delta_{N^A} = 0$ in (54). We denote investment incentives in the three regimes therefore as $\kappa^n(k)$, $\kappa^p(k)$

and $\kappa^c(k, N_i^A)$ but refrain from stating explicit expressions at this point. An implementation in the Wolfram Mathematica environment is available on request.

Comparing $\kappa^n(k)$ and $\kappa^p(k)$ reveals a critical level $\overline{k} \approx 6.45$ such that $\kappa^n(k) \leq \kappa^p(k)$ for $k \leq \overline{k}$ and $\kappa^n(k) > \kappa^p(k)$ for $k > \overline{k}$. For the comparison to the CDN case we rely on a numerical approach where we evaluate $\kappa^c(k, N_i^A))$ for a given level k at the root N_i^A to $\Delta_{N^A} = 0$. The performed analysis reveals $\kappa^c(k, N_i^A) > \kappa^n(k)$ and $\kappa^c(k, N_i^A) > \kappa^p(k)$ while $\lim_{k\to\infty}\big(\kappa^c(k, N_i^A) - \kappa^n(k)\big) = \lim_{k\to\infty}\big(\kappa^c(k, N_i^A) - \kappa^p(k)\big) = 0$. ■

References

Armstrong, M. (1998). Network interconnection in telecommunications. *Economic Journal*, pages 545–564.

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

BEREC (2016). Berec guidelines on the implementation by national regulators of european net neutrality rules.

Bourreau, M., Kourandi, F., and Valletti, T. (2015). Net neutrality with competing internet platforms. *The Journal of Industrial Economics*, 63(1):30–73.

Cheng, H. K., Bandyopadhyay, S., and Guo, H. (2011). The debate on net neutrality: A policy perspective. *Information systems research*, 22(1):60–82.

Choi, J. P., Jeon, D.-S., and Kim, B.-C. (2015). Net neutrality, business models, and internet interconnection. *American Economic Journal: Microeconomics*, 7(3):104–41.

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

Economides, N. and Hermalin, B. E. (2012). The economics of network neutrality. *The RAND Journal of Economics*, 43(4):602–629.

Economides, N. and Hermalin, B. E. (2015). The strategic use of download limits by a monopoly platform. *The RAND Journal of Economics*, 46(2):297–327.

Economides, N. and Tåg, J. (2012). Network neutrality on the internet: A two-sided market analysis. *Information Economics and Policy*, 24(2):91–104.

EP and Council of the EU (2015). Regulation (eu) 2015/2120 of the european parliament and of the council of 25 november 2015 laying down measures concerning open internet access and amending directive 2002/22/ec on universal service and users' rights relating to electronic communications networks and services and regulation (eu) no 531/2012 on roaming on public mobile communications networks within the union (text with eea relevance).

Greenstein, S., Peitz, M., and Valletti, T. (2016). Net neutrality: A fast lane to understanding the trade-offs. *Journal of Economic Perspectives*, 30(2):127–50.

Guo, H. and Easley, R. F. (2016). Network neutrality versus paid prioritization: Analyzing the impact on content innovation. *Production and Operations Management*, 25(7):1261–1273.

Harsanyi, J. C. and Selten, R. (1988). A general theory of equilibrium selection in games. *MIT Press Books*, 1.

Hau, T., Burghardt, D., and Brenner, W. (2011). Multihoming, content delivery networks, and the market for internet connectivity. *Telecommunications Policy*, 35(6):532–542.

Hermalin, B. E. and Katz, M. L. (2007). The economics of product-line restrictions with an application to the network neutrality debate. *Information Economics and Policy*, 19(2):215–248.

Hosanagar, K., Chuang, J., Krishnan, R., and Smith, M. D. (2008). Service adoption and pricing of content delivery network (cdn) services. *Management Science*, 54(9):1579–1593.

Kourandi, F., Krämer, J., and Valletti, T. (2015). Net neutrality, exclusivity contracts, and internet fragmentation. *Information Systems Research*, 26(2):320–338.

Krämer, J. and Wiewiorra, L. (2012). Network neutrality and congestion sensitive content providers: Implications for content variety, broadband investment, and regulation. *Information Systems Research*, 23(4):1303–1321.

Laffont, J.-J., Marcus, S., Rey, P., and Tirole, J. (2003). Internet interconnection and the off-net-cost pricing principle. *RAND Journal of Economics*, pages 370–390.

Laffont, J.-J., Rey, P., and Tirole, J. (1998a). Network competition: I. overview and nondiscriminatory pricing. *The RAND Journal of Economics*, pages 1–37.

Laffont, J.-J., Rey, P., and Tirole, J. (1998b). Network competition: Ii. price discrimination. *The RAND Journal of Economics*, pages 38–56.

Njoroge, P., Ozdaglar, A., Stier-Moses, N. E., and Weintraub, G. Y. (2013). Investment in two-sided markets and the net neutrality debate. *Review of Network Economics*, 12(4):355–402.

Wu, T. (2003). Network neutrality, broadband discrimination. *Journal of Telecommunications and high Technology law*, 2:141.