1629

Discussion
Papers

2016

Interconnection and Prioritization

Pio Baake and Slobodan Sudaric

Opinions expressed in this paper are those of the author(s) and do not necessarily reflect views of the institute.

Interconnection and Prioritization

Pio Baake   Slobodan Sudaric

DIW Berlin[*]   *Humboldt-University of Berlin*[†]

December 15, 2016

Abstract

We analyze pricing and competition under paid prioritization within a model of interconnected internet service providers (ISPs), heterogeneous content providers (CPs) and heterogeneous consumers. We show that prioritization is welfare superior to a regime without prioritization (network neutrality) and yields higher incentives for investment in network capacities. As ISPs price discriminate between on-net and off-net CPs, their bottleneck property is propagated and competition for consumers increases resulting in a potential prisoner's dilemma when deciding whether to offer prioritization. We show that peering for prioritized traffic emerges as a collusive outcome and present off-net prices as a further collusive instrument.

JEL–classification numbers: L13, L51, L96

Keywords: interconnection, investment, network neutrality, prioritization

1 Introduction

In November 2015 the *European Union* passed regulation on open internet access (EP and Council of the EU, 2015) providing a legislative answer to the European side of the ongoing debate on "network neutrality" - a concept that broadly requires that all traffic should be treated equally (Wu, 2003). This regulation was accompanied in August 2016 by guidelines set up by the *Body of European Regulators for Electronic Communications (BEREC)* which provide a framework for national regulators on how to execute the regulation in domestic markets (BEREC, 2016). While neutral treatment of internet traffic is a central pillar of the new regulation, internet service providers (ISP) may still offer differentiated Quality-of-Service (QoS) to content providers (CP) under certain conditions.[1]

In this paper we want to focus on the QoS practice of "paid prioritization", where CPs can pay ISPs to be prioritized over remaining traffic in order to reach consumers at a higher throughput rate. Considering that content types are typically highly heterogeneous with respect to transmission quality sensitivity (think of e-mails as opposed to live streaming), a tiered transmission quality scheme might be beneficial from a welfare perspective. An aspect we want to shed light on in addition is pricing and how pricing of prioritization can affect competition. While multi-homing of CPs and the extensive use of Content Delivery Networks (CDNs) is commonly observed with the heavy-weights of the content industry (Netflix, Youtube/Google, etc.), smaller CPs typically single-home and rely on interconnected networks.[2] Considering the sequential structure, where signals possibly pass through different networks when being transported from emitters (here CPs) to receivers (here consumers) the possibility of double marginalization has to be taken into account. At the same time, emitter and receiver might belong to the same or to different networks which could motivate price discrimination between on-net and off-net traffic when offering prioritization. Lastly, we want to analyze the role of interconnection charges in prioritization schemes.

---

[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]Radware Ltd. (2013) report that CDNs are used by 21 percent of European e-commerce websites and 23 percent for the US respectively.

We present a model with two symmetric and interconnected ISPs connecting a continuum of heterogeneous CPs to differentiated consumers. Content is differentiated with respect to connection quality sensitivity and we assume that in a regime of network neutrality ISPs offer a single quality level to all CPs free of charge whereas in a discriminatory regime networks operators offer two quality classes where the "priority lane" is associated with some fee structure, while the "non-priority" class remains free of charge.

Using this framework we show that prioritization propagates the bottleneck property of the ISPs such that competition for consumers intensifies heavily. At the same time prioritization allocates existing capacity more efficiently resulting in a higher welfare compared to a neutral regime. Strengthened competition combined with higher network quality leaves consumers as the main beneficiaries from prioritization. This increase in competition outweighs any profits network operators can make on the CP side in a discriminatory regime. In fact ISPs face a prisoner's dilemma: they have a unilateral incentive to introduce prioritization offers, but are worse off once both networks do so. The fee structure plays thereby an important role. As long as networks charge each other for terminating prioritized traffic, ISPs have an incentive to increase incoming prioritization traffic from CPs connected to the other network (off-net). Therefore, they lower consumer prices and make prioritization for off-net CPs more attractive. Networks therefore have a strong incentive to lower the interconnection fee in order to reduce competition for consumers. We furthermore show that charges for incoming off-net traffic can be used as a potential collusive instrument to overcome the prisoner's dilemma and show that the simplest fee structure, where ISPs only charge CPs connected to their own network (on-net) for prioritization can be established as a collusive outcome. Lastly, in terms of investment incentives we show that capacity investment is higher under prioritization as a discriminatory regime allows for higher increases in consumer prices.

Pricing in telecommunication networks is not a new area of interest for economists. Starting with pricing in telephone networks Armstrong (1998) and Laffont et al. (1998a) analyze how networks can use access charges as means of collusion. Since access charges increase each network's effective marginal cost, reciprocal charging can lead to higher consumer prices. In their accompanying article Laffont et al. (1998b) focus on discriminatory pricing for the termination of calls and conclude that this creates externalities between operators

even though their networks are interconnected. Dessein (2003) elaborates on those results and shows that access charges are difficult to maintain as an instrument for collusion when consumers are heterogenous. Later work explicitly focused on interconenction of internet networks. Adopting a two-sided market approach with content providers and consumers as distinct market sides Laffont et al. (2003) show that operators will charge their customers as if the entire customers' traffic were off-net. We will build on this stream of literature in a sense that we consider differences in pricing when offering prioritization of traffic which occurs entirely on-net as opposed to traffic which originated off-net.[3] Also, the interconnection aspect is reflected in the sense that we allow networks to charge CPs for simply passing traffic on (double marginalization) and to charge each other for exchanging traffic (interconnection fee).

In the traditional telecommunication literature the capacities of networks and eventual congestion problems played a minor role and therefore transmission quality and quality differentiation were largely disregarded. However, with the upcoming availability of broadband connections during the 2000s, new bandwidth intensive internet applications emerged which pushed network capacities especially in the last-mile segment to their limits. The resulting congestion externalities motivated CPs to strike deals for better QoS as connection quality is a crucial input factor in their competition for consumers. This spurred a general debate about whether or not those practices violate network neutrality. 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 modelling bandwidth limits where different qualities could introduce welfare gains in light of congested networks. Another stream of literature tackles the congestion problem using a queueing approach. Choi

---

[3]However, there is a crucial difference to price discrimination in telephone networks. While some phone plans discriminate in prices between on-net and off-net calls, the customer only gets billed by a single phone company. In analogy to our setting, however, the customer would pay her provider the price for on-net calls and each provider who terminated off-net calls seperately. This reflects the idea that e.g. a CP based in the United States might pay a European ISP for prioritization even though his internet access point is in the US.

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 heterogenous 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 as it allocates existing capacity more efficiently and provides higher investment incentives in the long run.

While we follow the same direction in terms of CP heterogeneity and the use of queueing, our model differs substantially as we model two competing and interconnected platforms. 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. Choi et al. (2013) present a similar setup but also account for the interconnection aspect between networks which makes it more comparable to our model. However, the most closely related work is the model by Bourreau et al. (2015) where competing ISPs offer queueing based prioritization to differentiated CPs and ISPs face a prisoner's dilemma when deciding whether to introduce prioritization, a result which recurs in our own analysis. Still, the focus of their model is slightly different. Bourreau et al. (2015) and also Njoroge et al. (2013) consider an elastic number of active CPs in the market and interpret the resulting difference as an effect on innovation in addition to the comparison in terms of welfare and investment incentives. While we acknowledge the importance of this issue, we do not allow for explicit exclusion of CPs in our model. At the same time Bourreau et al. (2015) disregard the issue of interconnection as they assume that the competing networks do not provide internet access to the CPs in the model. In terms of investment incentives we support the results of Njoroge et al. (2013) and Bourreau et al. (2015) as we also show that capacity investment is higher under prioritization. It that sense our work complements the existing literature as it is to our knowledge the only model which combines the problems of interconnection with competing platforms and queueing based prioritization of heterogeneous content types.

The paper is structured as follows. Section 2 introduces the model. Section 3 presents the case of network neutrality as benchmark. Section 4 characterizes the equilibrium with prioritization. Section 5 compares the two regimes regarding welfare aspects and investment

incentives. In Section 6 we consider collusion between ISPs. Section 7 concludes.

2 Model

We analyze a basic Hotelling model with two identical ISPs $i = 1, 2$ located at the ends of the Hotelling spectrum ($\lambda_i = 0$ for $i = 1$ and $\lambda_i = 1$ for $i = 2$). There is a continuum of differentiated consumers and content providers (masses normalized to one). Consumers have a uniformly distributed location $x \in [0, 1]$ and obtain utility u_i from consuming one unit of content from each CP of type θ delivered with quality q_i

$$u_i = \theta q_i \tag{1}$$

Overall utility (2) then depends on the aggregate monetary utility Eu_i from consuming content, the price p_i for internet access and the location of the consumer. Eu_i will be defined in detail later but can be thought of as the aggregate utility from content consumption across all CPs and all content types. 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.

$$U_i = \underline{u} + Eu_i - p_i - |\lambda_i - x| \tag{2}$$

CPs are differentiated with respect to their quality sensitivity θ which we assume to be uniformly distributed on $[0, 1]$. 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 purely advertisement based and that the interaction between a CP of type θ and one consumer connected to ISP i generates advertisement revenues

$$r_i = \theta q_i \tag{3}$$

for the CP.[4] Furthermore, we assume that CPs are differentiated with respect to their location $\mu \in [0,1]$ such that CPs incur some cost $t(\lambda_i) = \tilde{c}|\lambda_i - \mu|$ when engaging with an ISP where

$$\tilde{c} = \begin{cases} c \text{ with prioritization} \\ 0 \text{ without prioritization} \end{cases} \tag{4}$$

with $c > 0$. Note, (4) implies that only engaging in a prioritization contract with an ISP is costly for CPs. We follow the argument made in Bourreau et al. (2015) and assume that there is ubiquitous interconnectivity in the economy, such that CPs can reach consumers at ISP i at non-priority ("best-effort") quality irrespective of whether they are conencted to ISP i or j, while prioritization requires some direct agreement between the CP and ISP i. Costs $t(\lambda_i)$ can then be interpreted to capture geographical or contractual differences when engaging with one of the ISPs while the cost parameter c pins down the absolute level. Note, this structure implies that for each content type θ there is a full spectrum of μ. This reflects the idea that there is heterogeneity of CPs within the same content type. Their overall mass is normalized to one such that $\int_0^1 \int_0^1 d\theta d\mu = 1$.

Considering the quality of traffic, we distinguish between the backbone networks and the last-mile networks of the ISPs. We assume that the bottlenecks are the last-mile networks of the ISPs and that the traffic quality perceived by the consumers and the CPs depends on the capacities of the last-mile networks the consumers are connected to. Furthermore, we use a M/M/1 queueing model with an arrival rate of content requests equal to one and define quality as 1 over the waiting time. More specifically, let N_i denote the number of consumers connected to ISP i and let Y_i^P denote the number of CPs who purchased prioritization in network i. The quality q_i^p of prioritized traffic in network i and the quality q_i^0 of non-prioritized traffic in network i are given by

$$q_i^p = k_i - N_i Y_i^P \text{ with prioritiation} \tag{5}$$

$$q_i^0 = \frac{k_i - N_i}{k_i} q_i^p \text{ without prioritiation} \tag{6}$$

---

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

where k_i denotes the network capacity of ISP i.[5] Regarding notation we introduce superscripts p for "priority" and 0 for "no priority" in a regime of paid prioritization, and leave values under network neutrality without superscript. For the timing of the game we assume the following:

1. ISPs invest in network capacity (relevant in section 5.2)

2. Collusion (relevant in section 6).

3. ISPs agree on a symmetric interconnection fee. ISPs then set their connection prices for consumers and prioritization fees for CPs.

4. Consumers and CPs decide to which network they connect to and CPs decide whether they choose prioritization or not.

The solution concept is sub-game perfection and we solve the game by backward induction.

3 Benchmark case: No prioritization

In this section stages 1 and 2 of the game structure are omitted and the game starts at stage 3. We first establish a simple benchmark scenario without prioritization ("network neutrality"). In this scenario ISPs can not sell prioritization to CPs and their only sources of revenue are selling internet access to consumers.

Starting with CPs using (3)-(6) the profit $\pi(\theta, \mu, \cdot)$ without prioritized traffic can be written as

$$\pi_i(\theta, \mu, \cdot) = \theta \left(q_i N_i + q_j N_j \right) \tag{7}$$

where $q_i = q_i^0(Y_i^p = 0) = k_i - N_i$ and $q_j = q_j^0(Y_j^p = 0) = k_j - N_j$.[6] Since non-prioritized traffic is free for CPs (see section 2 for a motivation), CPs are indifferent between the two ISPs and the resulting allocation of CPs to ISPs is irrelevant for the equilibrium.

---

[5]We assume that each network could shoulder the whole traffic by itself, i.e. $k_i > 1 \forall i$ such that the non-negativity constraints for the qualities can be ignored. Also this assures that there are not any purely allocative reasons behind our setup.

[6]Here and in the following we will omit function arguments where this does not lead to any confusion.

Turning to consumers the average monetary utility without prioritization Eu_i can then be written as:

$$Eu_i = q_i \int_0^1 \int_0^1 \theta d\theta d\mu \qquad (8)$$

Eu_i aggregates over all content types θ at each location μ. Since there is only one quality level in the neutral regime all content types arrive at the uniform quality level q_i. The market shares of both ISPs are then given by the indifferent consumer on the Hotelling line

$$N_1 = \frac{1}{2} + \frac{1}{2}\left[(Eu_1 - p_1) - (Eu_2 - p_2)\right] \text{ and } N_2 = 1 - N_1. \qquad (9)$$

which gives rise to the following maximization problems for platform i:

$$\max_{p_i} \Pi_i = p_i N_i(\cdot), \ \ i = 1, 2$$

Note, that even though this is a very simple maximization problem it is not the standard Hotelling problem. The endogeneity of q_i leads to a less elastic market share $N_i(\cdot)$ than in the standard Hotelling model. Still, equilibrium prices are directly obtained from intersecting the respective reaction functions

$$p_i = \frac{1}{6}\left(9 + k_i - k_j\right); \ \ i \neq j \qquad (10)$$

and utilizing (9) gives rise to equilibrium profits of

$$\Pi_i = \frac{1}{108}\left(9 + k_i - k_j\right)^2; \ \ i \neq j \qquad (11)$$

4 Prioritization: Consumers, Content Providers and ISPs

We start by characterizing the behavior of consumers and content providers. Turning to the decisions of the ISPs we characterize properties of the ISPs' charges for prioritized traffic.

4.1 Content Providers and Consumers

Consider first the decisions of the CPs to choose prioritization and to which ISP to connect. Without prioritization profits are determined by best-effort quality levels q_i^0, q_j^0 as in (6).

$$\pi_i^0(\theta, \mu, \cdot) = \theta \left(q_i^0 N_i^p + q_j^0 N_j^p \right) \tag{12}$$

The profit with prioritized traffic is given by (14) where $t(\lambda_i)$ denotes the costs of prioritization which we introduced in section 2. Furthermore, we assume that if CPs decide to prioritize they will prioritize at both networks such that total prioritized traffic in a network Y_i^p is given by

$$Y_i^p = Z_i^p + Z_j^p \tag{13}$$

and consists of traffic originating in the same network (on-net) Z_i^p and traffic originating in the other network (off-net) Z_j^p.[7] If prioritization and connection to ISP i is chosen, profits are given by

$$\pi_i^p(\theta, \mu, \cdot) = (\theta q_i^p - f_{ii}) N_i^p + \left(\theta q_j^p - f_{ij} - f_{jt} \right) N_j^p - t(\lambda_i) \tag{14}$$

Regarding the fee structure we allow a priori for discrimination between prioritizatin of on-net and off-net traffic and for double marginalization. ISPs can therefore charge a different price for prioritization if the incoming traffic is on-net (f_{ii}) or off-net (f_{it}). Also, ISPs are able to charge a fee for prioritization even if termination occurs in different networks (f_{ij}), a fee for simply passing on traffic.

Figure 1 summarizes this fee structure with an exemplary CP who engages in prioritization contracts with ISP i and ISP j. Regarding on-net charges (1a) ISP i charges a fee f_{ii} for reaching a household connected to his network at quality q_i^p. The household in turn pays the internet access fee p_i. If the same CP wants to reach a household connected to ISP j at prioritized quality q_j^p we allow for two additional payments streams. First, the CP might have to pay f_{ij} to ISP i for passing traffic on (double marginalization), and secondly, ISP

---

[7]Note that we could guarantee prioritization at both networks by introducing a fixed component F in addition to the location specific component to the cost function. Given this amended cost function $t(\lambda_i) = c|\lambda_i - \mu| + F$ CPs would only be able to recover the fixed costs F by prioritizing at both networks. For tractability reasons we leave out the fixed component and focus on symmetric solutions where CPs prioritize at both networks by assumption instead.
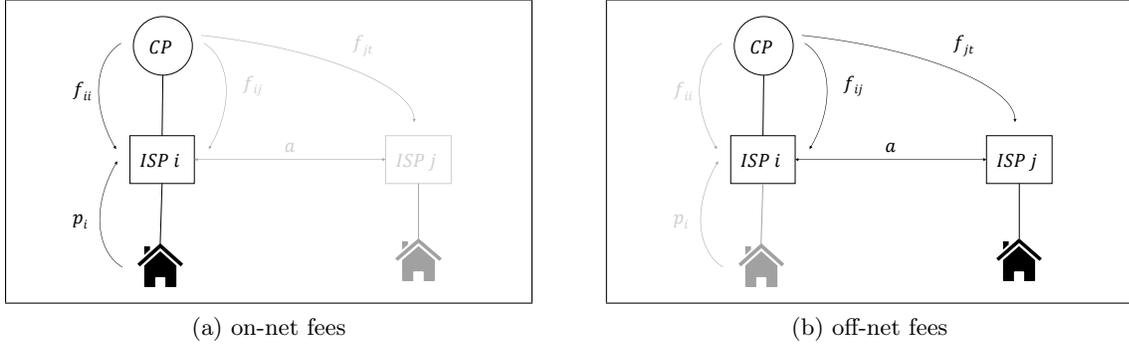
(a) on-net fees   (b) off-net fees

Figure 1: Fee structure

j might charge a termination fee f_{jt} which potentially differs from ISP j's on-net fee f_{jj} (discrimination). Note, that the linear fee structure could also be rewritten as a system of fixed payments or a combination of both (two-part tariffs).[8]

Comparing the profits of the CPs we define critical values of θ such that CPs are indifferent between prioritization and the free "best-effort" quality. First, we define functions $\theta_i^k(\cdot)$ such that $\pi^0(\theta_i^k(\cdot), \cdot) = \pi_i^p(\theta_i^k(\cdot), \cdot)$:

$$\theta_i^k(\cdot) = \frac{f_{ii}N_i^p + (f_{ij} + f_{jt})N_j^p + t(\lambda_i)}{N_i^p(q_i^p - q_i^0) + N_j^p(q_j^p - q_j^0)} \tag{15}$$

pinning down the content class $\theta_i^k(\cdot)$ which is indifferent between prioritization and no prioritization for a given location μ. Secondly, we define $\mu^k(\cdot)$ such that $\theta_1^k(\mu^k(\cdot), \cdot) = \theta_2^k(\mu^k(\cdot), \cdot)$:

$$\mu^k(\cdot) = \frac{1}{2} + \frac{1}{2c}\left[(f_{22}N_2^p + (f_{21} + f_{1t})N_1^p) - (f_{11}N_1^p + (f_{12} + f_{2t})N_2^p)\right] \tag{16}$$

yielding market shares of the ISPs on the CP prioritzation market. Using these critical values and focusing on interior solutions, i.e., $\mu^k(\cdot) \in (0,1)$ and $0 \leq \theta_i^k(\cdot) \leq 1$ for all $\mu \leq \mu^k(\cdot)$, we can calculate the numbers of CPs who connect either to ISP 1 or 2 and

---

[8]Remember that the quantity of transmitted content is fixed in our model (unit demand) and therefore a distinction between linear and fixed fees is irrelevant. Non-linear tariffs are not applicable to this setup for the same reason. This allows us to focus on the linear fees presented above.

choose prioritization as:

$$Z_1^p = \int_0^{\mu^k(\cdot)} \left(1 - \theta_1^k(\cdot)\right) d\mu \tag{17}$$

$$Z_2^p = \int_{\mu^k(\cdot)}^1 \left(1 - \theta_2^k(\cdot)\right) d\mu \tag{18}$$

A CP will therefore engage in a prioritization contract if she offers sufficiently quality-sensitive content ($\theta \geq \theta_i^k$) and/or it is relatively easy for her to engage in such contracts ($\mu \leq \mu^k$).

The average monetary utility with prioritization Eu_i^p perceived by consumers connected to ISP i is then given by

$$Eu_i^p = q_i^p \left[\int_0^{\mu^k(\cdot)} \int_{\theta_1^k(\cdot)}^1 \theta d\theta d\mu + \int_{\mu^k(\cdot)}^1 \int_{\theta_2^k(\cdot)}^1 \theta d\theta d\mu \right]$$
$$+ q_i^0 \left[\int_0^{\mu^k(\cdot)} \int_0^{\theta_1^k(\cdot)} \theta d\theta d\mu + \int_{\mu^k(\cdot)}^1 \int_0^{\theta_2^k(\cdot)} \theta d\theta d\mu \right] \tag{19}$$

To gain some intuition for Eu_i^p one should note that there is content which arrives at prioritized quality q_i^p while the remaining content arrives at best-effort quality q_i^0. Both 'quality groups', however, consist of content provided by CPs coming from both networks, therefore for each quality group there is content provided by CPs connected to the same network (on-net) and content provided by CPs connected to the other network (off-net). Finally, turning to the consumers' decisions the ISPs market shares can be written as before (see (9)) by pinning down the indifferent consumer

$$N_1^p = \frac{1}{2} + \frac{1}{2} \left[(Eu_1^p - p_1^p) - (Eu_2^p - p_2^p) \right] \text{ and } N_2^p = 1 - N_1^p \tag{20}$$

4.2   Internet Service Providers

Employing the functions $Z_i^p(\cdot)$ and $N_i^p(\cdot)$ the maximization problem of the ISPs with respect to their prices can be written as

$$\max_{p_i, f_{ii}, f_{ij}, f_{it}} \Pi_i^p = \underbrace{p_i N_i^p + f_{ii} Z_i^p N_i^p}_{on-net} + \underbrace{f_{ij} Z_i^p N_j^p + f_{it} Z_j^p N_i^p}_{off-net} + \underbrace{a \left[Z_j^p N_i^p - Z_i^p N_j^p \right]}_{interconnection} \tag{21}$$

12

The maximization problem reflects the fee structure depicted in figure 1. While the *on-net* and *off-net* parts concern payments from consumers to ISPs and from CPs to ISPs the *interconnection* part reflects the settlement for inter-network traffic. Depending on the balance between incoming prioritized traffic $Z_j^p N_i^p$ and outgoing prioritized traffic $Z_i^p N_j^p$ ISPs pay or receive a symmetric fee a. Analyzing the respective first order conditions given in the appendix we obtain the following result.[9]

Lemma 1 *The optimal charges f_{ij}^* can be normalized to zero.*

Proof. To prove the result we show that the first order conditions $\partial \Pi_i / \partial f_{ij} = 0$ and $\partial \Pi_i / \partial f_{ii} = 0$ are linearly dependent. We start by rewriting the indifference conditions for the consumers and the content providers and define

$$\Delta_1 := Eu_1^p - p_1^p - N_1^p - (Eu_2^p - p_2^p - (1 - N_1^p))$$

$$\Delta_2 := Z_1^p - \int_0^{\mu^k(\cdot)} \left(1 - \theta_1^k(\cdot)\right) d\mu \text{ and } \Delta_3 := Z_2^p - \int_{\mu^k(\cdot)}^1 \left(1 - \theta_2^k(\cdot)\right) d\mu.$$

Note that $\Delta_1 \equiv \Delta_2 \equiv \Delta_3 \equiv 0$ allows us to define N_1^p, Z_1^p and Z_2^p as functions of the ISPs' connection prices for consumers and prioritization fees for CPs. Starting with the partial derivatives of Δ_l with $l = 1, 2, 3$ we obtain

$$\frac{\partial \Delta_l}{\partial f_{ij}} = \frac{N_j^p}{N_i^p} \frac{\partial \Delta_l}{\partial f_{ii}}.$$

Using these results as well as $\Delta_l = 0$ and the implicit function theorem, standard comparative statics show that the reactions of N_1^p, Z_1^p and Z_2^p to changes in f_{ii} and f_{ij} satisfy

$$\frac{\partial N_1^p}{\partial f_{ij}} = \frac{N_j^p}{N_i^p} \frac{\partial N_1^p}{\partial f_{ii}}, \ \frac{\partial Z_i^p}{\partial f_{ij}} = \frac{N_j^p}{N_i^p} \frac{\partial Z_i^p}{\partial f_{ii}}, \frac{\partial Z_j^p}{\partial f_{ij}} = \frac{N_j^p}{N_i^p} \frac{\partial Z_j^p}{\partial f_{ii}}.$$

Substituting in the first order conditions $\partial \Pi_i^p / \partial f_{ii}$ and $\partial \Pi_i^p / \partial f_{ij}$ leads to

$$\frac{\partial \Pi_i^p}{\partial f_{ij}} = \frac{N_j^p}{N_i^p} \frac{\partial \Pi_i^p}{\partial f_{ii}}.$$

Hence, f_{ij}^* can be normalized to zero. ∎

---

[9]At this point we assume that the maximization problem is well-behaved. However, numerical analysis shows that our equilibrium results indeed characterize a global maximum.

Recall from the definition of CP profits (14) that when chosing ISP i payments going to ISP i are given by $f_{ii}N_i^p + f_{ij}N_j^p$ whereas $f_{jt}N_j^p$ is the payment to ISP j. Since CPs make a binary decision of whether to accept the prioritization offer or not, only the aggregate level of payments matters. In other words, from a CP perspective the composition of payments is irrelevant. In particular regarding the fraction going to ISP i, $f_{ii}N_i^p + f_{ij}N_j^p$, only the total value matters; splitting it into components depending on f_{ii} and f_{ij} is just 'book keeping'. From a regulatory point of view this result has an interesting implication: an isolated regulation of double-marginalization charges is unlikely to be effective as ISPs can reshuffle remaining fees such that the resulting market outcome is unchanged. Any regulation concerning double-marginalization charges f_{ij} must therefore be accompanied with a regulation of on-net charges f_{ii} in order to be effective.

Using $f_{ij}^* = f_{ji}^* = 0$ and analyzing the ISPs' first order conditions with respect to f_{ii} and f_{it} we then obtain

Lemma 2 *With $f_{ij}^* = f_{ji}^* = 0$ and*

$$f_{it} = f_{ii} - \frac{a}{N_i^p}$$

the solution f_{ii}^* of $\partial \Pi_i^p / \partial f_{ii} = 0$ also solves $\partial \Pi_i^p / \partial f_{it} = 0$ and implies $Z_1^p = Z_2^p$.

Proof. Using $\Delta_2 = 0$ and $\Delta_3 = 0$ (see the proof of Lemma 1) $f_{it} = f_{ii} - a/N_i^p$ leads to $Z_1^p = Z_2^p$ as well as

$$\frac{\partial \Delta_3}{\partial f_{it}} = \frac{\partial \Delta_3}{\partial f_{ii}} \text{ and } \frac{\partial \Delta_3}{\partial Z_1^p} = \frac{\partial \Delta_3}{\partial Z_2^p}$$

and ($i = 1, 2$)

$$\frac{\partial \Delta_i}{\partial f_{it}} = \frac{\partial \Delta_j}{\partial f_{ii}}, \ \frac{\partial \Delta_i}{\partial f_{jt}} = \frac{\partial \Delta_j}{\partial f_{jj}} \text{and } \frac{\partial \Delta_i}{\partial Z_i^p} = \frac{\partial \Delta_j}{\partial Z_j^p}, \ \frac{\partial \Delta_i}{\partial Z_j^p} = \frac{\partial \Delta_j}{\partial Z_i^p}.$$

Employing these equations, simple calculations show that solving $\partial \Pi_i^p / \partial f_{ii} = 0$ for f_{ii} and substituting in $\partial \Pi_i^p / \partial f_{it}$ also implies $\partial \Pi_i^p / \partial f_{ii} = 0$. ∎

This result has two main implications. First of all we see that the fee for prioritization of off-net traffic f_{it} is lower than the fee for on-net traffic f_{ii} as long as as interconnection fees a are positive. Since ISPs can not attract all CPs for their direct prioritization

14

business through f_{ii} they want to increase the balance of incoming traffic by lowering f_{it} such that the settlement balance $a\left[Z_j^p N_i^p - Z_i^p N_j^p\right]$ moves in their favour. Vice versa, we have $f_{it} = f_{ii}$ for $a = 0$. That means agreeing on seetlement free interconnection implies non-discrimination between prioritization of on-net and incoming off-net traffic as the settlement balance is irrelevant. In section 6 we elaborate on this result.

Using Lemmata 1 and 2 and imposing symmetry ($k_i = k_j = k$) we can then simplify the maximization problem in (21) to obtain equilibrium prices of

$$f_{ii}^* = \frac{1}{4}\left(2 + 4a - c - 4Z_i^p\right) - \frac{Z_i^p}{2k}\left(1 - 2Z_i^p\right) \tag{22}$$

$$f_{ij}^* = 0 \tag{23}$$

$$f_{it}^* = f_{ii}^* - 2a \tag{24}$$

$$p_i^{p*} = \frac{1}{6(k - Z_i^p)^2}\left[6Z_i^{p2}\left(1 + Z_i^p - 2f_{ii}^* Z_i^p\right)\right.$$
$$\left. + \left(k^2 - 2kZ_i^p\right)\left(12a^2 + c^2 + 6cf_{ii}^* - 6a\left(c + 4f_{ii}^*\right) + 6\left(1 + 2f_{ii}^*\left(f_{ii}^* - Z_i^p\right) + Z_i^p\right)\right)\right] \tag{25}$$

and equilibrium consumer market shares and number of prioritization contracts of

$$N_i^p = N_j^p = \frac{1}{2} \tag{26}$$

$$Z_i^p = Z_j^p = \frac{1}{20}\left(3 + 6k - \sqrt{9 + 4k(9k + 5c - 1)}\right) \tag{27}$$

Let us look at the equilibrium number of prioritization contracts in more detail. Note, that we have $\frac{dZ_i^p}{da} = 0$, i.e. the share of prioritizing CPs does not depend on the interconnection fee a. Consider the total payments a CP who opts for prioritization is charged $f_{total}^* = f_{ii}^* N_i^p + f_{it}^* N_j^p$. In equilibrium we have $\frac{df_{total}^*}{da} = 0$ and therefore $\frac{dZ_i^p}{da} = 0$. Furthermore, we have $\frac{dZ_i^p}{dc} < 0$. As engaging in prioritization contracts becomes more costly the share of CPs who can afford to do so decreases. Lastly, let us analzye the effect of existing capacity k. Remember we focus on interior solutions in a sense that $0 \le \theta_i^k(\cdot) \le 1$ for all μ, such that it is sufficient to check whether it holds for the indifferent CP $0 \le \theta_i^k(\mu^k) \le 1$. Using our equilibrium results above in equations (15) and (16) we can reduce this condition to an upper bound $\bar{c}$ for our cost parameter c.[10] Regarding our comparative static result we

---

[10]The upper bound is given by $\bar{c} = \frac{1}{72k}\left(k\left(-\sqrt{\frac{(4k-3)^2\left(64k^2+9\right)}{k^2}} + 32k + 12\right) - 9\right)$.

then have $\frac{dZ_i^p}{dk} > 0$ for $0 < c \leq \bar{c}$ implying that the number of prioritization contracts increases in the level of capacity k.

4.3   Interconnection fee

Finally we want to analyze which level of settlement fees a would be optimal if ISPs maximize joint profits. We follow the tradition in the literature (see e.g. Laffont et al., 2003) and assume that networks agree on a symmetric interconnection fee a prior to competing in CP and consumer prices. Utilizing our equilibrium results and in particular $Z_j^p = Z_i^p$ and $N_i^p = N_j^p = \frac{1}{2}$ we obtain ISP industry profits of

$$\Pi_{ISP}^p = \Pi_i^p + \Pi_j^p = p_i^{p*} + f_{ii}^* Z_i^p + f_{it}^* Z_i^p \tag{28}$$

Taking the derivative with respect to a evaluated at equilibrium values and remembering that $f_{it}^* = f_{ii}^* - \frac{a}{1/2}$ and hence $\frac{df_{it}^*}{da} = \frac{df_{ii}^*}{da} - 2 = -1$ yields

$$\frac{d\Pi_{ISP}^p}{da} = \underbrace{\frac{dp_i^{p*}}{da}}_{=-2Z_i^p} + \underbrace{\frac{df_{ii}^*}{da}}_{=1} Z_i^p + \underbrace{\frac{dZ_i^p}{da}}_{=0} f_{ii}^* + \underbrace{\frac{df_{it}^*}{da}}_{=-1} Z_i^p + \underbrace{\frac{dZ_i^p}{da}}_{=0} f_{it}^* = -2Z_i^p < 0 \tag{29}$$

which gives rise to the following result:

Proposition 1 *Settlement-free interconnection $a^* = 0$ maximizes joint profits.*

Consider increasing the interconnection fee to $a > a^*$. The increase in fees for on-net prioritization $\frac{df_{ii}^*}{da} = 1$ is completely offset by the decrease in fees for off-net prioritization $\frac{df_{it}^*}{da} = -1$. However, ISPs want to generate a positive net inflow of settlement charges $a\left[Z_j^p N_i^p - Z_i^p N_j^p\right]$ and therefore want to increase their share of the market on the consumer side by lowering prices, $\frac{dp_i^{p*}}{da} = -2Z_i^p < 0$. Since the number of prioritization contracts does not depend on the interconnection fee $\frac{dZ_i^p}{da} = 0$, a can be seen as an instrument to shift rents from consumers to ISPs, i.e. $-\frac{dCS^p}{da} = -2Z_i^p = \frac{d\Pi_{ISP}^p}{da}$.

Since the industry profit is decreasing in a, ISPs have an incentive to agree on settlement-free interconnection and this corresponds nicely to the peering agreements we observe in the industry. However, we want to stress that peering as an equilibrium outcome can in our model only be supported through an industry structure with symmetric ISPs. Once

we consider an asymmetric industry structure, settlement free interconnection seems to be unlikely. Note further, that changes in a are welfare-neutral and just shift rents from ISPs to consumers, which makes the interconnection fee an interesting instrument for policy makers.[11]

5 Comparison

This section compares the neutral regime characterized in section 3 to the prioritization regime from 4. 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 in both regimes. Proofs are delegated to the Appendix.

5.1 Welfare

Turning to equilibrium values given in equations (22)-(27), we can now compare the prioritization scheme to the bechmark scenario. Starting with the definition of consumer utility (2) and utilizing symmetry we can calculate a simplified measure for consumer surplus under paid prioritization

$$CS^p = 2 \left[\int_0^{\frac{1}{2}} \underline{u} + Eu_i^p - p_i^p - |\lambda_i - x| \, dx \right] \tag{30}$$

and under network neutrality accordingly

$$CS = 2 \left[\int_0^{\frac{1}{2}} \underline{u} + Eu_i - p_i - |\lambda_i - x| \, dx \right] \tag{31}$$

and Eu_i^p and Eu_i are given by (19) and (8) evaluated at respecive equilibrium values. Also, let $\Delta CS \equiv CS^p - CS$

---

[11] We a priori assume $a \geq 0$. Let us consider a lift of the non-negativity constraint. $a < 0$ implies a negative net inflow of settlement charges $a \left[Z_j^p N_i^p - Z_i^p N_j^p \right]$ when attracting consumers. To dampen this effect ISPs will increase consumer prices. However, this mechanism would be very prone to fraudulent and manipulative behavior. For example, if $a < 0$, network i could generate money by setting up a machine which requests prioritized content from network j. This leads us to the presumption that in practical applications the non-negativity contraint on the interconnection charge would be binding and indeed we do not observe negative interconnection charges in the industry. The assumption $a \geq 0$ will therefore be maintained for the subsequent analysis.

Proposition 2 *Consumers benefit from prioritization $CS^p > CS$ as consumer prices decline $p_i^p < p_i$ and average connection quality increases $Eu_i^p > Eu_i$.*

Competition for consumers is now propagated. Attracting consumers not only increases market share on the consumer side, which in turn shifts the settlement balance between networks in favor of the attracting network. Also, it makes the ISP's prioritization offering more appealing to CPs increasing the network's share of CP business which in turn has a negative effect on the settlement balance. However, the attraction of CPs is less intense since ISPs can opt for off-net prioritization instead of switching networks. The increase in importance of the consumer market is reflected in the fact that consumer prices decrease $p_i^p < p_i$ and average connection quality increases $Eu^p > Eu$ which in the end is beneficial for consumers.

Turning to CPs we firstly define industry profits

$$\Pi_{CP}^p = 2\left[\int_0^{\frac{1}{2}}\int_0^{\theta_i^k}\pi_i^0 d\theta d\mu + \int_0^{\frac{1}{2}}\int_{\theta_i^k}^1\pi_i^p d\theta d\mu\right] \tag{32}$$

$$\Pi_{CP} = \int_0^1\int_0^1\pi_i^0 d\theta d\mu \tag{33}$$

where π_{CP} denotes the industry profit without prioritization (7) and π_{CP}^p with prioritization contract (14) evaluated at respective equilibrium values. Furthermore, let $\Delta\Pi_{CP} \equiv \Pi_{CP}^p - \Pi_{CP}$.

Proposition 3 *The aggregated CP industry is worse off under prioritization $\Pi_{CP}^p < \Pi_{CP}^0$ while CPs with high quality sensitivity potentially benefit from prioritization.*

Since CPs who choose to prioritize now incur costs $t(\lambda_i)$ and pay fees $\{f_{ii}^*, f_{ij}^*, f_{it}^*\}$, the aggregated benefits of higher quality provision for content sensitive providers can not prevail. Since we have $q_i^0 < q_i < q_i^p$ there is a fraction of CPs who do not prioritize $1 - (Z_i^p + Z_j^p)$ and are worse off under the prioritzation scheme. Also, a fraction of CPs with prioritization are also worse off as we can always find CPs who are indifferent between purchasing prioritization and sticking to the best-effort quality, leaving CPs with very quality sensitive content as the only group of CPs who potentially benefit from prioritization.

Finally, we can compare ISP industry profits under prioritization (see (28))

$$\Pi^p_{ISP} = \Pi^p_i + \Pi^p_j = p^{p*}_i + f^*_{ii} Z^p_i + f^*_{it} Z^p_i$$

and under net neutrality

$$\Pi_{ISP} = \Pi_i + \Pi_j = \frac{3}{2} \tag{34}$$

in the symmetric equilibrium and define $\Delta \Pi_{ISP} \equiv \Pi^p_{ISP} - \Pi_{ISP}$.

Proposition 4 *ISP profits are lower under prioritization* $\Pi^p_{ISP} < \Pi_{ISP}$. *If costs are low* $c \leq \tilde{c} \leq \overline{c}$, *ISPs face a prisoner's dilemma.*

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. This result also holds if ISPs exchange traffic for free $a = 0$. However, this leads to the question whether ISPs face a prisoner's dilemma. If given that the other ISP does not offer prioritization, introducing prioritization leads to higher profits than given by (11), a coordination failure may arise where both ISPs end up with lower profits. In fact one can show that if the costs for CPs to engage in prioritization are low enough $c < \tilde{c} < \overline{c}$ the coordination problem will arise. If costs are high on the other hand ISPs will not offer prioritization. This result also allows an interpretation regarding the competitive structure of the market: if the CP market is very competitive (c is very low, i.e. networks are not very differentiated) the dilemma will arise. The critical level $\tilde{c}$ is characterized in the appendix.

Finally we combine our results into one welfare measure

$$\Delta W = \Delta CS + \Delta \Pi_{CP} + \Delta \Pi_{ISP}$$

and conclude this section with the welfare comparison between both regimes.

Proposition 5 *Welfare is higher under prioritization* $\Delta W > 0$.

Consumers face lower prices for internet access and higher average quality. While some CPs benefit from an increase in quality due to prioritization of high-sensitivity traffic, the aggregate of CPs suffers from prioritization in light of setup costs and fees charged by

network operators. On the ISP side the additional revenue stream is depleted by tightened competition on the consumer side to a point, where ISPs have lower profits under prioritization. However, the net effect is positive.

5.2 Investment incentives

In this section we want to compare investment incentives between the regime with and without paid prioritization from a symmetric equilibrium perspective. We assume that investment costs $C_i(k_i)$ with $C_i'(k_i) > 0$, $C_i''(k_i) > 0$ are identical in both regimes and therefore we can restrict our analysis to the comparison of marginal profits gross of investment costs. The marginal profit in the regime without prioritization $\frac{d\Pi_i}{dk_i}$ follows from (11)

$$\left.\frac{d\Pi_i}{dk_i}\right|_{k_i=k_j} = \left.\left(p_i\frac{dN_i}{dk_i} + N_i\frac{dp_i}{dk_i}\right)\right|_{k_i=k_j} = \frac{1}{6} \tag{35}$$

whereas marginal profits under prioritization $\frac{d\Pi_i^p}{dk_i}$ can be written as

$$\left.\frac{d\Pi_i^p}{dk_i}\right|_{k_i=k_j} = \left(p_i^{p*}\frac{dN_i^p}{dk_i} + N_i^p\frac{dp_i^{p*}}{dk_i} + N_i^p f_{ii}^*\left[\frac{dZ_i^p}{dk_i} + \frac{dZ_j^p}{dk_i}\right]\right.$$
$$\left.\left. + (Z_i^p + Z_j^p)\left[f_{ii}^*\frac{dN_i^p}{dk_i} + N_i^p\frac{df_{ii}^*}{dk_i}\right]\right)\right|_{k_i=k_j} \tag{36}$$

where we use our results from Lemmata 1 and 2 in the definition of ISP profits under prioritization (14) and our result from section 4.3 such that interconnection is settlement-free ($a = 0$). Figure 2 compares investment incentives for different levels of initial symmetric investment network capacity $k_i = k_j = k$. The solid line depicts investment incentives under prioritization $\frac{d\Pi_i^p}{dk_i}$ and the dashed line the case without prioritization $\frac{d\Pi_i}{dk_i}$. Proposition 6 summarizes the illustrated results.

Proposition 6 *Investment incentives are higher in a regime with prioritization $\Delta_i^k \equiv \frac{d\Pi_i^p}{dk_i} - \frac{d\Pi_i}{dk_i} > 0$. The difference is less pronounced if initial capacity levels are high $\frac{d\Delta_i^k}{dk_i} < 0$ or if prioritization is costly $\frac{d\Delta_i^k}{dc} < 0$.*

We should note that the difference $\Delta_i^k > 0$ is mainly driven by the difference in marginal profits made from consumers $\left(p_i^{p*}\frac{dN_i^p}{dk_i} + N_i^p\frac{dp_i^{p*}}{dk_i}\right) - \left(p_i\frac{dN_i}{dk_i} + N_i\frac{dp_i}{dk_i}\right) > 0$. In both regimes
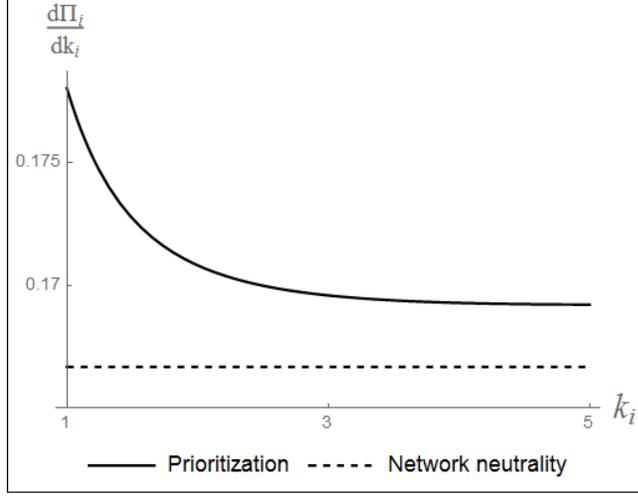
Figure 2: Comparison of investment incentives

an increase in quality allows networks to increase consumer prices. However, with prioritization consumers additionally benefit from the presence of prioritized content classes. The positive effect in terms of consumers is especially pronounced if we consider that the marginal profits made from the CP business are negative $N_i^p f_{ii}^* \left[\frac{dZ_i^p}{dk_i} + \frac{dZ_j^p}{dk_i} \right] + (Z_i^p + Z_j^p) \left[f_{ii}^* \frac{dN_i^p}{dk_i} + N_i^p \frac{df_{ii}^*}{dk_i} \right] < 0$. Utilizing Lemma 2 this inequality implies

$$f_{ii}^* \left[N_i^p \frac{dZ_i^p}{dk_i} + Z_i^p \frac{dN_i^p}{dk_i} \right] < -N_i^p Z_i^p \frac{df_{ii}^*}{dk_i} \tag{37}$$

with $\frac{dZ_i^p}{dk_i} > 0$, $\frac{dN_i^p}{dk_i} > 0$ and $\frac{df_{ii}^*}{dk_i} < 0$. Even though the capacity investment would increase the market share on the consumer side $\frac{dN_i^p}{dk_i} > 0$ and the number of prioritization contracts $\frac{dZ_i^p}{dk_i} > 0$, the adverse effect on prioritization fees prevails $\frac{df_{ii}^*}{dk_i} < 0$. Note that increasing the capacity level has a positive impact on both quality levels, which in turn makes purchasing prioritization less appealing.

Similarly, the higher the initial capacity level, the higher the free best-effort quality. This in turn lowers the benefit of purchasing prioritization and therefore $\frac{d\Delta_i^k}{dk_i} < 0$. The higher the cost for engaging in a prioritization contract c, the smaller the number of CPs who can afford to purchase prioritization, implying $\frac{d\Delta_i^k}{dc} < 0$.

6   Collusion

This section analyzes how a symmetric fee for prioritization of off-net traffic $f_{it} = f_{jt} = f_t$ can be used as an instrument for collusion. In particular we are interested in the cases where f_t deviates from the optimal level of the non-collusive case towards zero. The idea is that ISPs offer a package of prioritization (on-net and off-net) but CPs only pay one fee f_{ii} to the ISP they are connected to. The panel in figure 3 depicts equilibrium values of the collusive game (solid line) for different levels of f_t where the right bound is given by f_{it}^* such that the depicted range shows the movement from f_{it}^* towards zero. The gray horizontal lines indicate reference values from the non-collusive game as characterized in section 4. The dashed line shows reference values from the regime without prioritization as shown in section 3.[12]

The top left graph depicts ISP profits for different levels of f_t. As can be seen collusion regarding the pricing of off-net prioritization is a potential remedy to the prisoner's dilemma the ISPs are facing. Profits can be increased above the non-collusive level by lowering the level of f_t towards zero.

Turning to the mid left graph we see as f_t approaches zero the optimal f_{ii} decreases as well which makes opting for prioritization very cheap for CPs increasing the overall number of prioritization contracts (top right graph). The complementarity between f_t and f_{ii}^p is, however, not straightforward. If we disentangle f_t into f_{it} and f_{jt} one can show that $\frac{\partial f_{ii}^p}{\partial f_{it}} > 0$, so that prioritization fees are complementary on-net. However, we also have $\frac{\partial f_{ii}^p}{\partial f_{jt}} < 0$, which implies that on-net fees f_{ii}^p and off-net fees f_{jt} are strategic substitutes. What we observe in figure 3, however, is $\frac{df_{ii}^p}{df_t} > 0$ which implies $|\frac{\partial f_{ii}^p}{\partial f_{it}}| > |\frac{\partial f_{ii}^p}{\partial f_{jt}}|$ using our separation $\frac{df_{ii}^p}{df_t} = \frac{\partial f_{ii}^p}{\partial f_{it}} \frac{df_{it}^p}{df_t} + \frac{\partial f_{ii}^p}{\partial f_{jt}} \frac{df_{jt}^p}{df_t} > 0$ with $\frac{df_{it}^p}{df_t} = \frac{df_{jt}^p}{df_t} = 1$.

As CP payments per consumer are now very low, competition for consumers becomes less intense which pushes consumer prices up (mid right graph). Also, it is easy to show that for $Z_i^p \to \frac{1}{2}$ we have $q_i^p \to q_i$: prioritization of every CP is equivalent to no prioritization at all. Unsurprisingly, the aggregate quality level Eu_i decreases towards the level under net neutrality (bottom graph).[13] This implies that by pushing f_t towards zero ISPs can

---

[12]For the purpose of illustration equilibrium values are evaluated at $c = 0.02$, $k = 2$ and $a = 0$.

[13]Note, however that full prioritization, i.e. $Z_i^p = \frac{1}{2}$ can not be reached since we assume $c > 0$, i.e. prioritzation is costly. When we argue that values approach levels of net neutrality we are strictly speaking
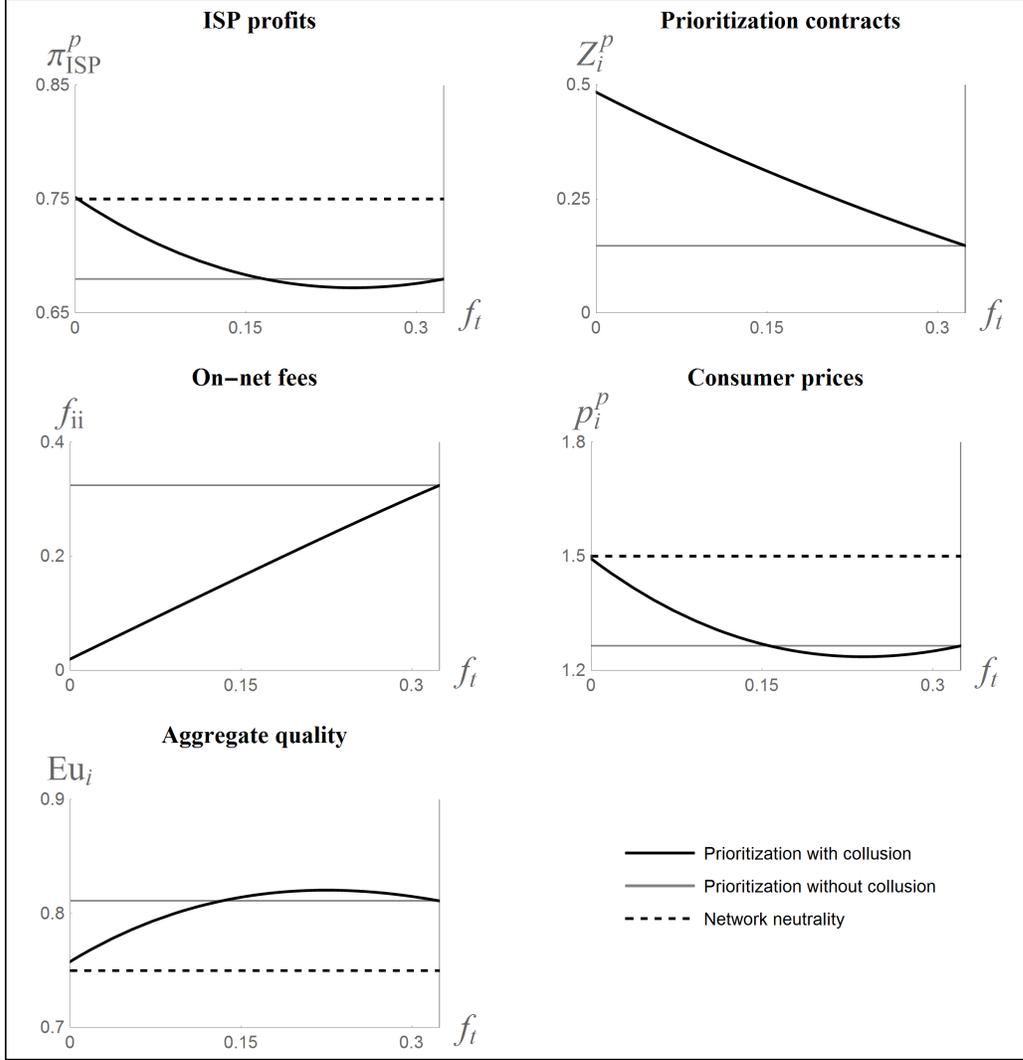
Figure 3: Collusion

simply coordinate to go back towards the neutral regime. Consumer prices approach the level observed in the case without prioritzation $p_i \to p_i^0$ (bottom right graph) and so do ISP profits $\Pi_{ISP}^p \to \Pi_{ISP}$ (top left graph). A similar argument can be made for increasing f_t above the competitive level f_{it}^*.[14]

However, this suggests the presence of opposing opposing effects as can be seen in the

---

about closing in on levels of net neutrality up to some margin ϵ where ϵ is determined by c. This, however, is omitted for the sake of brevity.

[14] Prioritization gets increasingly more costly which implies that the total number of prioritization contracts decreases. If the number of prioritization contracts is very low we have $q_i^0 \to q_i$ which again relates to a regime without prioritization. However, consumer prices increase in this case in f_t. Since consumer demand in our model is inelastic by design, however, this increase is theoretically unbounded. This makes a reasonable analysis in ranges of high f_t values very hard and we therefore stick to the presented case of $f_t \to 0$.

non-monotonicity of e.g. the consumer price graph. Consider lowering f_t marginally from the competitive level f_{it}^*. Since f_{ii} is complementary, the direct payment stream from CPs to ISPs drops initially. ISPs try to offset the reduced fee level by attracting new consumers, so consumer prices decrease even further. However, after a certain threshold substituting the reduced prioritization fees through attracting new consumers becomes too costly and consumer prices begin to increase again.

This provides a good narrative for the 'consumer sweet spot' in the non-collusive case. For $f_t \rightarrow f_{it}^*$ attracting consumers becomes crucial and unsurprisingly consumer prices are low. At the same time utility from consuming content is high as existing capacity is allocated efficintly through the queueing mechanism. This, however, leads to low ISP profits. By colluding with respect to off-net prices network operators can avoid this pitfall. In particular our results show that a regime where ISPs only charge positive prices for on-net CPs (and zero prices for off-net CPs, $f_t = 0$) can be an outcome of collusive behavior. In this case ISPs can approach the profit level under net neutrality and CPs face very low fees for prioritization (f_{ii} is still positive), however consumers face higher prices and lose the benefit of prioritized content classes.

We would like to end this section by briefly discussing the stability of the collusive outcome. Obviously, every network could profitably deviate from the collusive fee $f_t = 0$ by marginally increasing its fees above zero $f_{it} > 0$. This would render the collusion unstable since in our model there is no contractual punishment for deviation or any other more advanced mechanism ensuring collusion. This is beyond the scope of this work. Nevertheless, we would like to make an intuitive argument in favour of the collusive outcome. Let us imagine a CP who signed a contract with ISP i at price $f_{it}(f_t) > 0$ and $f_t = 0$ promising prioritization in both networks. In order to be able to make this offer, there must also be some agreement between ISP i and ISP j in place. If ISP j now introduces a positive additional fee in exchange for prioritization in network j, ISP i can not fulfill the promise made in the contract with the CP anymore invoking potential legal repercussions between the CP and ISP i and consequently between ISP i and ISP j. The fact that deviating from the collusive price setting (i.e. $f_{it} > f_t = 0$) is not in the interest of CPs could therefore make the collusion more stable.

7 Conclusion

We characterized an equilibrium with competing and interconnected ISPs offering queueing-based prioritization to heterogeneous CPs. Prioritization leads to increased competition for consumers and to a more efficient use of existing network capacity, implying a higher welfare level compared to a regime of network neutrality. Comparing investment incentives we also conclude that capacity investment is higher under prioritization.

We showed that details in the fee structure matter when it comes to collusion of ISPs. Since a positive interconnection fee produces a strong incentive to generate a positive inter-network balance by 1) attracting consumers through low consumer prices and 2) discriminating in favour of incoming off-net traffic, networks have a strong incentive to decrease interconnection fees. This is especially true since networks are likely to face a prisoner's dilemma when it comes to the introduction of prioritization and would therefore want to reclaim some lost rents. We motivate collusion with respect to pricing of off-net traffic as a potential remedy to the dilemma and show that the simplest fee structure where only on-net CPs are charged for prioritization can be a collusive industry outcome. In light of the ongoing network neutrality debate our analysis favours a non-neutral regime and highlights once more the importance of competition for consumers within this context. However, there are some clear limitations to our model. We restrict our analysis to the case where CPs single-home when it comes to prioritization and rely on interconnection agreements to reach different networks. If CPs are able to multi-home for prioritization, the bottleneck property of ISPs is likely to be weakened and the role of interconnection is less clear. Our results also rely on a symmetric setup. However, having in mind regional ISPs which partly rely on infrastructure provided by larger ISPs, we expect our results to change significantly. Lastly, one could also imagine the content industry to actively react to ISP investments into network capacity. Since capacity investment affects all quality levels it is not clear to what extent ISP and CP investments are complementary. These questions could provide some direction for future research.

Appendix

A   Equilibrium conditions

The following system of equations describes the first order conditions to the maximization problem presented in (21).

$$
\begin{aligned}
\frac{\partial \Pi_i}{\partial p_i} &= N_i + p_i \frac{\partial N_i}{\partial p_i} + f_{ii}\left[Z_i\frac{\partial N_i}{\partial p_i} + N_i\frac{\partial Z_i}{\partial p_i}\right] + (f_{ij}-a)\left[Z_i\frac{\partial N_j}{\partial p_i} + N_j\frac{\partial Z_i}{\partial p_i}\right] \\
&\quad + (f_{it}+a)\left[Z_j\frac{\partial N_i}{\partial p_i} + N_i\frac{\partial Z_j}{\partial p_i}\right] = 0 \\
\frac{\partial \Pi_i}{\partial f_{ii}} &= Z_i N_i + p_i \frac{\partial N_i}{\partial f_{ii}} + f_{ii}\left[Z_i\frac{\partial N_i}{\partial f_{ii}} + N_i\frac{\partial Z_i}{\partial f_{ii}}\right] + (f_{ij}-a)\left[Z_i\frac{\partial N_j}{\partial f_{ii}} + N_j\frac{\partial Z_i}{\partial f_{ii}}\right] \\
&\quad + (f_{it}+a)\left[Z_j\frac{\partial N_i}{\partial f_{ii}} + N_i\frac{\partial Z_j}{\partial f_{ii}}\right] = 0 \\
\frac{\partial \Pi_i}{\partial f_{ij}} &= Z_i N_j + p_i \frac{\partial N_i}{\partial f_{ij}} + f_{ii}\left[Z_i\frac{\partial N_i}{\partial f_{ij}} + N_i\frac{\partial Z_i}{\partial f_{ij}}\right] + (f_{ij}-a)\left[Z_i\frac{\partial N_j}{\partial f_{ij}} + N_j\frac{\partial Z_i}{\partial f_{ij}}\right] \\
&\quad + (f_{it}+a)\left[Z_j\frac{\partial N_i}{\partial f_{ij}} + N_i\frac{\partial Z_j}{\partial f_{ij}}\right] = 0 \\
\frac{\partial \Pi_i}{\partial f_{it}} &= Z_j N_i + p_i \frac{\partial N_i}{\partial f_{it}} + f_{ii}\left[Z_i\frac{\partial N_i}{\partial f_{it}} + N_i\frac{\partial Z_i}{\partial f_{it}}\right] + (f_{ij}-a)\left[Z_i\frac{\partial N_j}{\partial f_{it}} + N_j\frac{\partial Z_i}{\partial f_{it}}\right] \\
&\quad + (f_{it}+a)\left[Z_j\frac{\partial N_i}{\partial f_{it}} + N_i\frac{\partial Z_j}{\partial f_{it}}\right] = 0
\end{aligned}
$$

Adding the system of auxillary equations presented in the proof of Lemma 1

$$
\begin{aligned}
\Delta_1 &:= Eq_1^p - p_1 - N_1^p - (Eq_2^p - p_2 - (1-N_1^p)) && = 0 \\
\Delta_2 &:= Z_1^p - \int_0^{\mu^k(\cdot)} \left(1 - \theta_1^k(\cdot)\right) d\mu && = 0 \\
\Delta_3 &:= Z_2^p - \int_{\mu^k(\cdot)}^1 \left(1 - \theta_2^k(\cdot)\right) d\mu && = 0
\end{aligned}
$$

and normalizing $f_{ij} = 0$ (see Lemma 1) we can apply the implicit function theorem to determine the remaining elasticities. Imposing symmetry we obtain closed form solutions for our equilibrium values.

B   Omitted proofs

Proposition 1

Proof. See (29). ∎

Proposition 2

Proof. In a first step we show that $p_i^{p*} < p_i$. Note, that $\frac{dp_i^{p*}}{da} \leq 0$ such that it is sufficient to check whether $p_i^{p*} < p_i$ at $a = 0$. We can then see that $\frac{d(p_i - p_i^{p*})}{dc}\Big|_{a=0} < 0 \ \forall \ c \in [0, \bar{c}]$ and therefore it suffices to show that $(p_i - p_i^{p*})\big|_{a=0, c=\bar{c}} > 0$ which is true $\forall \ k \geq 1$ and therefore implying $p_i^{p*} < p_i$. Similarly, we show that $Eu_i < Eu_i^p$. Note, that $\frac{dEu_i^p}{dc} < 0$ and obviously $\frac{dEu_i}{dc} = 0$. Since we have $(Eu_i^p - Eu_i)\big|_{c=\bar{c}} > 0$ it follows by monotonicity that $Eu_i < Eu_i^p \ \forall \ c \in [0, \bar{c}], 1 \leq k$. Finally, $p_i^{p*} < p_i$ and $Eu_i < Eu_i^p$ implies $CS < CS^p$ from the definitions in (30) and (31). $\blacksquare$

Proposition 3

Proof. In equilibrium we have $q_i^0 < q_i < q_1^p$ implying that profits without prioritization (7) decrease. Since there are CPs of type $\theta_i^k(\cdot)$ who are indifferent between prioritization and best-effort quality (15), some CPs are also worse off under prioritization. Since $\frac{\partial \Pi_i^p}{\partial \theta} > 0$ only CPs at the high end of the type spectrum potentially benefit. For the aggregate effect we should note that $\frac{d\Pi_{CP}^p}{dc} > 0$ and $\frac{d\Pi_{CP}}{dc} = 0$. Since $\Pi_{CP}^0 - \Pi_{CP}^p\big|_{c=\bar{c}} > 0$ we have $\Pi_{CP}^p < \Pi_{CP}^0 \ \forall c \in [0, \bar{c}], 1 \leq k$. $\blacksquare$

Proposition 4

Proof. We start by comparing profits Π_{ISP}^p and Π_{ISP}. Note, that $\frac{d(\Pi_{ISP}^p - \Pi_{ISP})}{dc} < 0$ as $\frac{d\Pi_{ISP}^p}{dc} > 0$ and $\frac{d\Pi_{ISP}}{dc} = 0$ such that it is sufficient to show that $\Pi_{ISP}^p - \Pi_{ISP} > 0\big|_{c=\bar{c}}$ which holds for all $1 \leq k$.

For the second part of the proposition we characterize an equilibrium where only one ISP offers prioritization. Assume without loss of generality that ISP 2 does not offer prioritization. Let $\{p_1^d, p_2^d, f_{11}^d\}$ denote equilibrium prices derived from the profit maximization problem of ISP 1 $\max_{p_1, f_{11}} \Pi_1^d = p_1 N_1 + f_{11} Z_1^p N_1$ and ISP 2 $\max_{p_2} \Pi_2^d = p_2 N_2$ where the share of CPs with prioritization is given by $Z_1^p = \int_0^1 \left(1 - \theta_1^k(\cdot)\right) d\mu$ and consumer market shares are obtained as in (9). Comparing profits of ISP 1 evaluated at equilibrium values $\Pi_1^d(\cdot)$ to the symmetric solution without prioritization given in (7) implicitly defines the critical cost level $\tilde{c} : \Pi_1^d(\cdot) = \Pi_1^0$ with $\tilde{c} < \bar{c}$ such that $\Pi_1^d(\cdot) > \Pi_1^0$ for $0 < c < \tilde{c}$ and $\Pi_1^d(\cdot) \leq \Pi_1^0$ for $\bar{c} \leq c < \bar{c}$. $\blacksquare$

Proposition 5

Proof. Note, that all monetary payments are transfers between consumers and ISPs and CPs and ISPs respectively and therefore do not affect overall welfare. The difference in welfare can then be expressed as $\Delta W = 2\left(Eu_i^p - Eu_i - \int_0^{\frac{1}{2}} \int_{\theta_i^k(\cdot)}^1 t(\lambda_i) d\theta d\mu\right)$ such that the sign of the difference will depend critically on the cost level c. One can show that ΔW is decreasing for $c \in [0, 2]$ with $c = 2$

being the smallest positive root to $\Delta W = 0$. However, it is easy to verify that $\bar{c} < 2$ and therefore $\Delta W > 0$ for all $c \in [0, \bar{c}], 1 \leq k$. ∎

Proposition 6

Proof. As $\frac{d\Delta_i^k}{dc} < 0$ one can verify that $\Delta_i^k > 0\big|_{c=\bar{c}}$ for all $1 \leq k$ and therefore $\Delta_i^k > 0$ for all $c \in [0, \bar{c}], 1 \leq k$. ∎

References

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

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. (2013). Net neutrality, business models, and internet interconnection. *American Economic Journal: Microeconomics*.

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

Dessein, W. (2003). Network competition in nonlinear pricing. *RAND Journal of Economics*, pages 593–611.

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

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).

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.

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.

Radware Ltd. (2013). State of the union - european ecommerce page speed and web performance.

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