TSE working paper series

12-355

November 2012

"Internet Interconnection and Network Neutrality"

Jay Pil Choi, Doh-Shin Jeon

and

Byung-Cheol Kim

Internet Interconnection and Network Neutrality*

Jay Pil Choi[†] Doh-Shin Jeon[‡] Byung-Cheol Kim[§]

November 12, 2012

Abstract

We analyze competition between interconnected networks when content is heterogeneous in its sensitivity to delivery quality. In a two-sided market framework, we characterize the equilibrium in a neutral network constrained to offer the same quality and assess the impact of such a constraint vis-à-vis a non-neutral network where Internet service providers (ISPs) are allowed to engage in second degree price discrimination with a menu of quality-price pairs. We find that the merit of net neutrality regulation depends crucially on content providers' business models. More generally, our analysis can be considered as a contribution to the literature on second-degree price discrimination in two-sided platform markets.

JEL Codes: D4, L1, L5

Key Words: Net neutrality, Internet interconnection, Two-sided markets, Second-degree price discrimination, Access (Termination) charges, CPs' business models

*We thank Yuk-Fai Fong, Dominik Grafenhofer, Bruno Jullien, Teddy Kim, David Laband, Wooyoung Lim, Qihong Liu, Juan Moreno-Cruz, Hyunwoo Park, Yossi Spiegel, and seminar audiences at 2012 ICT Conference in Paris, 2012 MaCCI Law and Economics Conference on "The Future of the Internet," Spring 2012 Midwest Economic Theory Conference, 2012 Korean Econometric Society Summer Meeting, Hong Kong University of Science and Technology, Georgia Institute of Technology, University of New South Wales and University of Seoul for helpful comments.

†School of Economics, University of New South Wales, Sydney, NSW 2052,Australia and Department of Economics, Michigan State University, 220A Marshall-Adams Hall, East Lansing, MI 48824 -1038. E-mail: choijay@msu.edu.

‡Toulouse School of Economics and CEPR, Manufacture de Tabacs, 21 allees de Brienne - 31000 Toulouse, France. E-mail: dohshin.jeon@gmail.com.

§School of Economics, Georgia Institute of Technology, 221 Bobby Dodd Way, Atlanta, GA 30332-0225. Email: byung-cheol.kim@econ.gatech.edu.

1 Introduction

The Internet is a system of interconnected computer networks that is often characterized as a "network of networks." The universal connectivity that enables any computers connected to the Internet to communicate with each other is ensured by cooperative interconnection arrangements among network operators. The current state of the Internet is also governed by a more or less implicit principle of "net neutrality" that treats all packets equally and delivers them on a first-come-first-served basis without blocking or prioritizing any traffic based on types of Internet content, services or applications.

However, with the emergence of various on-line multi-media services that demand a significant amount of network bandwidth, network congestion and efficient management of network resources have become important policy issues. In particular, content and applications differ in their sensitivity to delay in delivery. For instance, data applications such as E-mail can be relatively insensitive to moderate delivery delays from users' viewpoints. By contrast, streaming video/audio or Voice over Internet Protocol (VoIP) applications can be very sensitive to delay, leading to jittery delivery of content that provides unsatisfactory user experiences. With such heterogeneity concerning delay costs, one may argue that network neutrality treating all packets equally regardless of content is not an efficient way to utilize the network.

Even if there is an agreement concerning the desirability of offering multi-tiered Internet services, implementation of such a system is not a simple matter with interconnected networks. Guaranteeing a specified quality (speed) of content delivery requires cooperation from other networks when content providers and end users belong to different networks. Though interconnected Internet service providers (hereafter ISPs) agree on the provision of delivery quality, they may compete in the two groups of end users, consumers who subscribe to the Internet access and content/application providers who want to deliver their content for consumption.

In this paper, we develop a theoretical model of interconnection to reflect these key features of the Internet ecosystem and highlight the importance of content providers' business models in assessing the effects of net neutrality. More specifically, we adopt a two-sided market framework in which ISPs serve as platforms that connect content providers (hereafter CPs) and end consumers. On the CP side, there is a continuum of heterogeneous content/application providers who can multi-home, i.e., subscribe to multiple ISPs. CPs' contents differ in their sensitivity to delivery quality: for a clear exposition, we consider two types of CPs. This justifies the need to provide multiple lanes of different delivery qualities. On the consumer side, we assume that consumers single-home and constitute competitive bottlenecks in the market.[1] To model competition on the consumer side of the market we employ a Hotelling model with hinterlands (Armstrong and Wright, 2009) that allows us to represent elastic subscription demand by consumers and ISPs' market power vis-a-vis consumers. By contrast, we assume Bertrand competition without friction on the content side, which simplifies our analysis. These assumptions are to reflect a typical real world environment in which ISPs have strong market power against consumers because of the lack of competition for "the last mile" delivery, while their market power is limited with respect to content providers who can choose among multiple ISPs to distribute their content.

When both CPs and consumers belong to the same ISP, all traffic can be delivered on-net. However, if a CP purchases a delivery service from one ISP and consumers subscribe to another ISP, interconnection between these two ISPs is required for the completion of content delivery. We consider two broad regimes under which packet delivery can take place. Under a neutral regime mandated by net neutrality regulation, all packets are delivered with the same quality. Under a non-neutral regime, in contrast, ISPs are allowed to offer multiple lanes with different delivery quality levels. We assume that the ISPs agree on the delivery quality and reciprocal access charge(s) for the delivery of other ISPs' traffic that terminate on their own networks.

We find that any equilibrium in our model is governed by the so-called "off-net cost pricing principle" on the CP side. The off-net cost pricing principle was discovered by Laffont, Marcus, Rey, and Tirole (hereafter LMRT, 2003) and Jeon, Laffont and Tirole (2004). It means that network operators set prices for their customers as if their customers' traffic were entirely off-net. We generalize the finding of LMRT to a setting of heterogenous content with different delivery qualities across content. We establish that off-net cost pricing on the CP side combined with Hotelling competition with hinterland on the consumer side creates an equivalence between competing ISPs in our model and a hypothetical benchmark case of monopoly with homogenous consumers. Competing ISPs essentially agree on access charges and delivery qualities that would enable them to behave as monopoly bottlenecks against CPs. By using this equivalence, we consider a scenario that would favor price discrimination and thus stack the deck against the neutral regime when the surpluses from interactions between the CPs and end consumers are entirely appropriated by one-side of the market. Nonetheless, we show that a neutral regime can be welfare-enhancing when the CPs' surplus extraction is not such extreme cases.

[1]See Armstrong (2006) for various modes of competition in two-sided markets.

The intuition for this main result is as follows. When choosing the quality for low type CPs whose content is relatively insensitive to delivery speed, the ISPs face a trade-off. A downward distortion in the quality for low type CPs has a *benefit* of extracting more rent of high type CPs and a *cost* of reducing the consumer surplus that the ISPs can extract. This implies that as the ISPs focus on extracting consumer surplus rather than extracting CPs' surplus, there will be less distortion in quality. Because the ISPs have more instruments to extract CPs' surplus in a non-neutral network whereas they are constrained to offer one level of quality in a neutral network, it can happen that they focus on extracting CPs' surplus under non-neutral networks while they focus on extracting consumer surplus under neutral networks. When this arises, the ISPs under neutral networks serve both types' CPs in order to extract more consumer surplus and hence social welfare can be higher under neutral networks than under non-neutral networks.

More generally, the welfare comparison between neutral networks and non-neutral ones possibly may reveal a non-monotonic relationship over the relative allocation of total surplus among two end-user groups, consumers and CPs. Suppose first that consumers take the entire surplus generated by content delivery. Then, ISPs will provide the *first best* quality for each type of CPs in a non-neutral network because this allows them to extract the highest consumer surplus from subscription fees. By contrast, a suboptimal single quality is provided in a neutral network due to the regulatory restriction. Thus, in this case the non-neutral network yields a strictly higher social welfare than the neutral network. As the rent on the CP side increases, however, the social welfare ranking between the two regimes would be reversed due to the accelerated quality distortion against low type CPs in the non-neutral network, provided that the neutral network still serves both types of CPs. Once the exclusion occurs under a single quality provision, the non-neutral network reclaims a higher social welfare. This is because the non-neutral network still serves low type CPs while high type CPs are offered the first best quality in both network regimes.

Our result highlights the importance of the CPs' relative share of total surplus generated by a content delivery—primarily affected by what kinds of business models that CPs would take—in the evaluation of net neutrality regulation. More generally our finding contributes to the literature on second degree price discrimination in platform markets. We show that how an ISP's second-degree price discrimination fares against no discrimination depends on the relative allocation of each group's surplus in a two-sided market.

Our research is closely related to LMRT (2003) who analyze how the access charge allocates communication costs between CPs and end consumers and thus affects competitive strategies of rival networks in an environment of interconnected networks. They show that the principle of off-net-cost pricing prevails in a broad set of environments. Our model builds upon their interconnection model, but focuses on the provision of optimal quality in content delivery services by introducing heterogeneity in CPs' content type. In this setting, we analyze how the quality levels and access charges are determined depending on CPs' business model and on whether there exists net neutrality regulation.

There is a large literature on interconnection in the telecommunication market, initiated by Armstrong (1998) and Laffont, Rey, and Tirole (1998a,b). These researchers show that if firms compete in linear prices, they agree to set interconnection charges above associated costs to obtain the joint profit-maximizing outcome and derive the welfare-maximizing interconnection charge that is lower than the privately negotiated level. They also show that the nature of competition can be altered significantly depending on whether or not two-part tariffs or termination-based price discrimination are employed as price instruments. Their models, however, are devoid of the issue of transmission of quality because all calls are homogeneous. In contrast, we assume heterogeneous types of CPs requiring different transmission qualities and analyze the quality distortion associated with a non-neutral network and the (sub)optimality of net neutrality regulation.[2]

Our research also contributes to the literature on net neutrality. With net neutrality being one of the most important global regulatory issues concerning the Internet, there has been a steady stream of academic papers on various issues associated with net neutrality regulation in recent years.[3] To the best of our knowledge, we are the first to explore implications of net neutrality in the framework of two-sided markets with *interconnected* and *competing* ISPs. Choi and Kim (2010) analyze the effects of net neutrality regulation on investment incentives of a monopoly ISP and CPs. They show that ISPs may invest less in capacity in a non-neutral network than in a neutral network because expanding capacity reduces the CPs' willingness to pay for having a prioritized service. Economides and Hermalin (forthcoming) derive conditions under which network neutrality would be welfare superior to any feasible scheme for prioritized service given a capacity of bandwidth. They show that the ability to price discriminate enhances

[2]Armstrong (1998) and Laffont, Rey, and Tirole (1998a,b) consider inelastic subscription of consumers while we consider elastic subscription demand. In this sense, our model is related to the papers that study elastic subscription demand in the literature on interconnection in telecommunications market, Armstrong and Wright (2009), Dessein (2003), Hurkens and Jeon (2012).

[3]See Lee and Wu (2009) and Schuett (2010) for the surveys about economics literature on network neutrality.

incentives to invest, creating a trade-off between static and dynamic efficiencies. As these papers consider a monopolistic ISP, the interconnection and competition issues do not arise. Hermalin and Katz (HK, 2007) examine a situation in which ISPs serve as platforms to connect CPs with consumers in a framework of two-sided markets. HK consider heterogeneous CPs whose products are vertically differentiated to analyze the effects of net neutrality regulation. Without any restrictions, ISPs can potentially offer a continuum of vertically differentiated services, although the ISPs are required to provide a single tier of Internet service with net neutrality regulation. They compare the single service equilibrium with the multi-service equilibrium. One novelty of our paper with respect to HK is that we analyze how the relative benefit and cost of allowing second-degree price discrimination depends on CPs' business model.

Bourreau, Kourandi, and Valletti (2012) analyze the effect of net neutrality regulation on capacity investments and innovation in the content market with competing ISPs. They show that investments in broadband capacity and content innovation are higher under a non-neutral regime. However, they do not allow interconnection between ISPs by assuming that a CP has access only to the end users connected to the same ISP. Economides and Tåg (2012) also consider both a monopolistic ISP and duopolistic ISPs. Once again the issue of internet interconnection is not considered as they focus on how net neutrality regulation as a zero pricing rule affects pricing schemes on both sides of the market and social welfare.

The remainder of the paper is organized as follows. In section 2, we set up a basic model of interconnected networks with competition where two ISPs play the role of a two-sided market platform that connects CPs on one side and end consumers on the other side. In section 3, we consider two benchmarks of the first best and a monopolistic ISP with homogenous consumers. The latter is intended to introduce some of the key parametric assumptions and establish connections between the monopoly case and competing ISPs later. In section 4, we analyze network competition on the CP side of the market and show that any equilibrium is characterized by the off-net cost pricing principle regardless of net neutrality regulation. In section 5, we analyze network competition in the consumer subscription market and derive a central equivalence result between competing ISPs and a monopolistic ISP: the ISPs that agree to access charges and delivery qualities to maximize their joint profits behave as a monopoly ISP facing homogeneous consumers with inelastic subscription. In sections 6, we analyze ISPs' choice of quality and access charges in each regime and compare them. In particular, we derive conditions under which zero termination fee ("bill and keep") arises endogenously as an equilibrium outcome. In section 7, we conduct a welfare analysis and derive conditions under which the neutral regime can outperform the non-neutral network in social welfare. This result shows the importance of CPs' business models in the evaluation of net neutrality regulation. Section 8 contains our concluding remarks, along with suggestions for possible extensions of our analysis and policy implications with respect to making net neutrality regulation contingent on whether we consider mobile or fixed Internet.

2 A Model of Interconnected Networks with Competition

2.1 ISPs, CPs, and Consumers

We consider two interconnected ISPs denoted by $i = 1, 2$. ISPs serve as platforms in a two-sided market where CPs and end consumers constitute two distinct groups of customers. As pointed out by LMRT (2001, 2003), the traffic between CPs and the traffic between consumers such as E-mail exchanges take up trivial volumes relative to the volume of traffic from CPs to consumers. Thus, we focus on the primary traffic from CPs to consumers who browse web pages, download files, stream multi-media content, etc. As in the literature of interconnected networks, we assume a *balanced traffic pattern* that consumers' interest in a CP is independent of the CP's ISP choice and *reciprocal access pricing* that implies no asymmetry in the access charge for incoming and outgoing traffics.

There is a continuum of CPs whose mass is normalized to one. We consider a simple case of CP heterogeneity. There are two types of CPs: $\theta \in \{\theta_H, \theta_L\}$, with $\Delta\theta = \theta_H - \theta_L > 0$. The measure of θ_k type CP is denoted by ν_k, where $k = H, L$, and $\nu_H = \nu$ and $\nu_L = 1 - \nu$. There is also a continuum of consumers who demand one unit of each content whose value depends on content type θ and its quality q. In our context, quality means speed and reliability of content delivery. Let q_k denote the quality of delivery associated with content of type θ_k. The total surplus generated from interaction between a consumer and a CP of type θ is equal to $\theta u(q)$, where $u' > 0$ and $u'' < 0$ with the Inada condition $\lim_{q \to 0} u'(q) = \infty$. According to our utility formulation, θ reflects the sensitivity of content to delay, with higher valuation content being more time/congestion sensitive. Note that $\theta u(q)$ captures not only a consumer's gross surplus but also a CP's revenue from advertising. We assume that this surplus is divided between a CP and a consumer such that the former gets $\alpha\theta u(q)$ and the latter $(1 - \alpha)\theta u(q)$ with $\alpha \in [0, 1]$. The parameter α reflects the nature of CPs' business model. We have in mind two sources of revenue for CPs: micropayments and advertising revenue.[4] For instance, the parameter α would be higher if CPs can extract surplus from consumers via micropayments in addition to advertising revenues. If CPs' revenue source is limited to advertising, α can be relatively low. We later show that CPs' business models, captured by α, plays an important role in assessing the effects of net neutrality regulations.[5]

The two ISPs are horizontally differentiated on the consumer side. To model elastic participation of consumers, we adopt a "Hotelling model with hinterlands" as in Armstrong and Wright (2009) and Hagiu and Lee (2011). More specifically, let $U_i(\alpha)$ denote the gross utility a consumer derives from the content side by subscribing to ISP i. We have $U_i(\alpha) = \underline{u} + (1-\alpha) \sum_{k=H,L} \nu_{ki}\theta_k u(q_k)$, where $\underline{u}$ is the intrinsic utility associated with the Internet connection and ν_{ki} is the measure of type $k(= H, L)$ CPs whose content can be consumed by subscribing to ISP i. Under interconnection with the reciprocal access pricing described in subsection 2.2, we have $\nu_{k1} = \nu_{k2} \equiv \nu_k^*$ where ν_k^* is the measure of type $k(= H, L)$ CPs subscribed to any of the two ISPs. A consumer's net utility from subscribing to ISP i (gross of the transportation cost), u_i, is given by

$$u_i(\alpha, f_i) = U_i(\alpha) - f_i, \tag{1}$$

where f_i is the subscription price charged by ISP i.

The demand for network i in a Hotelling model with hinterlands is given by

$$n_i = \left[\frac{1}{2} + \frac{u_i - u_j}{2t}\right] + \lambda u_i, \tag{2}$$

where n_i is the measure of consumers subscribing to ISP i, t is the "transportation cost" parameter. The expression in the square bracket in (2) is the standard Hotelling demand specification with inelastic subscription in which consumers of mass one are uniformly located on the Hotelling line, assuming that $\underline{u}$ is sufficiently large relative to t such that the competitive Hotelling market

[4]Our simplifying assumption is that CPs are homogeneous in all dimensions except for their type θ. This implies that they use the same business model; otherwise, CPs have two-dimensional types (θ and business model). In ad-based business model, we can assume that the advertising revenue is proportional to consumer gross utility, which in turn is proportional to $\theta u(q)$. Then, the total surplus is given by $b\theta u(q)$ where b is a positive constant. Hence, by redefining $b\theta$ as θ', we are back to our original formulation.

[5]Casadesus-Masanell and Llanes (2012) also emphasize the importance of user bargaining power vis-a-vis application developers in their analysis of investment incentives in two-sided platforms. See Jullien and Sand-Zantman (2012) for an analysis that endogenizes the choice of business model in the context of congestion pricing and net neutrality. They study a "missing price" problem which arises since consumers do not know each CP's type, which is about the traffic loads on the ISP resulting from the use of the CP's content.

segment is fully covered. The term λu_i represents the demand from ISP i's "hinterlands," with $\lambda \geq 0$ representing the relative importance of market expansion possibilities. More specifically, each ISP faces a downward sloping demand of loyal consumers on each side of the unit interval; consumers in these areas never consider buying from the alternative ISP and have the same transportation cost parameter of t, but are uniformly distributed with a density of h. In such a scenario, the marginal consumer type in the loyal consumer group who is indifferent between not subscribing and subscribing to ISP i is distance x_i^* away (in the hinterlands) from the ISP, where x_i^* is defined by $u_i - tx_i^* = 0$. Thus, the number of consumers in the hinterland is given by $hx_i^* = \lambda u_i$ with $\lambda = \frac{h}{t}$.[6]

2.2 Network Interconnection and Network Neutrality

ISPs provide network services that deliver content from CPs to consumers. The total marginal cost of providing a unit traffic of quality q from CP to end users is assumed to be linear, i.e., $c(q) = cq$ for $q \geq 0$.[7] The total marginal cost has two components, i.e., $c = c_O + c_T$, where $c_O \geq 0$ and $c_T \geq 0$ stand for the cost of origination and that of termination per quality, respectively. We assume that ISPs cannot engage in first-degree price discrimination across content providers depending on content types.

We consider two different regimes under which ISPs can deliver content: neutral regime or non-neutral regime. For simplicity, we consider cooperative choice of quality and access charge under both regimes. Under a non-neutral regime, ISPs can offer multiple classes of services that differ in delivery quality. In other words, they can engage in second degree price discrimination by offering a menu of contracts that charges different prices depending on the quality of delivery. Let q_H be the quality for high type CPs and q_L for low type CPs; let A_H and A_L denote the reciprocal termination charges for each quality class. The termination charge *per unit quality* for type k traffic can be implicitly defined as $a_k = \frac{A_k}{q_k}$. Then, for one unit of off-net traffic of quality $q = q_H$ from ISP j to ISP i (i.e., a consumer subscribed to ISP i asks for content from a CP subscribed to ISP j), the origination ISP j incurs a cost of $c_O q_H$ and pays an access charge of $a_H q_H$ $(= A_H)$ to ISP i, and the termination ISP i incurs a cost of $c_T q_H$ and receives an access

[6] The market expansion possibility parameter λ can also be represented by the same density of consumers in the hinterlands (i.e., $h = 1$), but with a different transportation parameter for consumers in the hinterlands, say t_h.

[7] The assumption of a linear marginal cost in quality can be made without any loss of generality because we can normalize quality to satisfy the assumption of linearity. Suppose that $c(q)$ is nonlinear. By redefining $\widetilde{q}$ as $c(q)/c$, we have a linear marginal cost function $\widetilde{c}(\widetilde{q})= c\widetilde{q}$. Starting from a concave utility function and a convex cost function, after this linealization, the utility function with the normalized cost remains still concave.

charge of $a_H q_H$ from ISP i. Let $\widehat{c}_k \equiv c + a_k - c_T$ $(= c_O + a_k)$ denote the perceived unit quality cost of the off-net content that terminates in the other network for $q = q_k$, where $k = H, L$.[8]

In a neutral regime or in the presence of net neutrality regulation, ISP i is constrained to offer a single uniform delivery quality q.[9] The ISPs jointly choose a single quality level and a single access charge A with $a = \frac{A}{q}$. Let $\widehat{c} \equiv c + a - c_T$ denote the off-net cost per unit quality in the neutral network. Define $\widehat{\mathbf{c}}$ such that $\widehat{\mathbf{c}} \equiv (\widehat{c}_H, \widehat{c}_L)$ in the non-neutral network and $\widehat{\mathbf{c}} \equiv \widehat{c}$ in the neutral network.

We note that because of the interconnection agreement between the two ISPs, a CP can reach any consumer subscribed to either ISP regardless of the ISP it chooses to deliver its content. Figure 1 illustrates the flows of traffic and payment over two interconnected networks.

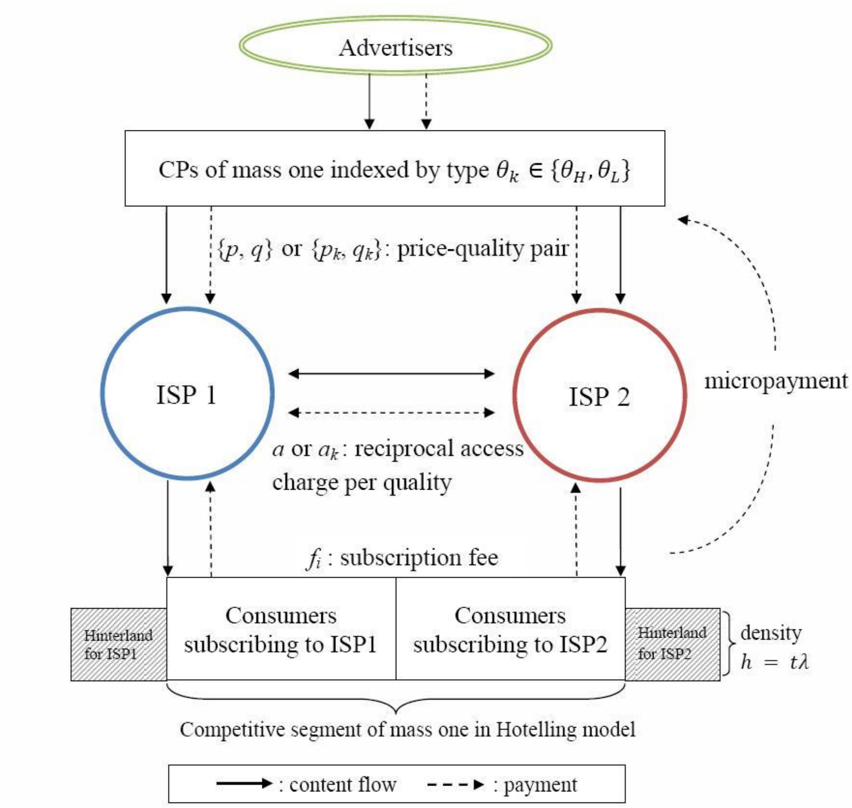

Figure 1: The flows of traffic and payment over interconnected networks

[8]In a non-neutral regime, we can further distinguish two cases depending on whether or not termination-based price discrimination (TPD) is possible. With TPD, ISP i proposes a pricing schedule $\{p_i(q), \widehat{p}_i(q)\}$ for $q \in \{q_H, q_L\}$ such that upon paying $p_i(q)$ (respectively, $\widehat{p}_i(q)$) a CP can obtain delivery of its content with quality q from ISP i for a unit of on-net traffic (respectively, a unit of off-net traffic). In this paper, we do not consider the possibility of TPD, that is, we analyze only the case where $p_i(q) = \widehat{p}_i(q)$. However, the qualitative results do not change when we consider TPD if CPs are allowed to multi-home.

[9]Under the neutral regime, there is no TPD because content cannot be treated differently depending on its destination.

2.3 Timing of Decisions

The game is played in the following sequence.

- Stage 1: The quality levels and the corresponding access charges are negotiated between the ISPs.
- Stage 2: In the non-neutral regime, each ISP i with $i = 1, 2$ simultaneously sets for CPs $\{p_i(q_H), p_i(q_L)\}$ a menu of a unit price for each delivery quality. In the neutral regime, there is only one delivery quality and each ISP sets a price of $p_i(q)$. Given the price schedules, each CP decides whether to participate in the market, and if it participates, decides which ISP to use to deliver its content (and what type of delivery service to purchase in the non-neutral regime).
- Stage 3: Each ISP i with $i = 1, 2$ simultaneously posts its consumer subscription fee f_i and consumers make their subscription decisions.

One main reason to consider this sequential timing rather than two alternative timing scenarios where stages 2 and 3 are reversed or take place simultaneously is that the ISPs have less incentive to deviate from the joint-profit maximizing prices under this sequential timing than under the other ones.[10] We establish that the off-net cost pricing on the content side holds regardless of the timing (in section 4) and show that there is an upper bound on the ISPs' joint profit associated with the off-net cost pricing (in section 5). Hence, the upper bound does not depend on the timing we choose. Finally, we show that the ISPs can achieve this upper bound under the sequential timing specified above (in section 6).

3 Benchmarks

3.1 First-Best

Before analyzing market outcomes under various regimes, we first analyze the first-best outcome as a benchmark. Note first that the access charge plays no role in the first-best since it is a pure transfer and that the social cost of providing q to a pair of consumer and CP is cq regardless of whether they belong to the same ISP or two different ISPs. Therefore, for any given configuration

[10] For instance, under the simultaneous timing, if ISP i deviates in its offer to CPs, it can also adjust its offer to consumers, but ISP j cannot. In contrast, in the sequential timing that we consider, if ISP i deviates in stage 2 by changing its offer to CPs, ISP j can adjust its offer to consumers in stage 3.

of consumers subscribing to the networks, the socially optimal quality should maximize $\theta_k u(q)-cq$ and hence, the first-best quality level for CPs of type θ_k, denoted by q_k^{FB}, is determined by the following condition:

$$\theta_k u'(q_k^{FB}) = c, \text{ where } k = H, L. \tag{3}$$

The marginal benefit of an incremental improvement of delivery quality for the content of type θ must be equal to c, the marginal cost associated with such an adjustment.

Define u^{FB} and c^{FB} as the gross utility from all content providers and its associated content delivery cost for each consumer when the first best delivery qualities are chosen:

$$\begin{aligned} u^{FB} &= \sum_{k=H,L} \nu_k \theta_k u(q_k^{FB}), \\ c^{FB} &= \sum_{k=H,L} \nu_k c q_k^{FB}. \end{aligned}$$

Then, social welfare as a function of the measure of subscribed consumers, denoted by $N(\geq 1)$, is given by

$$W^{FB}(N) = N \times \left(\underline{u} + u^{FB} - c^{FB}\right) - T(N),$$

where $T(N)$ represents the total transportation cost incurred by consumers. Since the number of consumers in the competitive market is normalized to one, the total number of consumers from the hinterlands is given by $(N-1)$. The total transportation cost $T(N)$ is minimized with a symmetric subscription pattern to the two ISPs. Let x be the distance from the marginal consumer in the hinterlands to the closest ISP with $(N-1)/2$ subscribers. Then, $x = \frac{N-1}{2h}$ where $h(=\lambda t)$ is the density for the consumers in the hinterlands. With our "hinterlands" specification, we thus have

$$T(N) = 2\int_0^{\frac{1}{2}} txdx + 2h\int_0^{\frac{N-1}{2h}} txdx = \frac{t}{4} + \frac{(N-1)^2}{4\lambda}.$$

The first-best measure of subscribed consumers, denoted by N^{FB}, is given by the following first order condition:

$$\underline{u} + u^{FB} - c^{FB} = \left.\frac{dT(N)}{dN}\right|_{N=N^{FB}} = \frac{N^{FB}-1}{2\lambda}(= tx^{FB}). \tag{4}$$

The transportation cost of the marginal consumer (tx^{FB}) should be equal to ($\underline{u} + u^{FB} - c^{FB}$) in the first-best outcome.

Proposition 1 *(First-best) The first best outcome characterized by (3) and (4) requires non-neutrality since* $q_H^{FB} > q_L^{FB}$*. Under perfect price discrimination, the first-best outcome can be implemented by a price schedule to CPs* $p(q_k^{FB}) = \alpha\theta_k u(q_k^{FB})$ *and consumer subscription price* $f = c^{FB} - \alpha u^{FB}$*: it requires a subsidy on consumer side if* $c^{FB} < \alpha u^{FB}$*. Then, each CP and each ISP realize zero profit.*

The point of the proposition is simple. With heterogeneous content that differs in sensitivity to delivery quality, the uniform treatment of content mandated by net neutrality in general would not yield a socially optimal outcome. The prices described in Proposition 1 are the unique ones that implement the first-best outcome under the budget constraint of the social planner: each ISP and each CP realizes zero profit and the marginal consumer is indifferent between subscribing and not subscribing.

3.2 Monopoly ISP with Homogeneous Consumers and Key Assumptions

As another benchmark, we consider a hypothetical setting in which a monopoly ISP provides content delivery service from the continuum of CPs (described previously) to *homogeneous* consumers with inelastic subscription.[11] This benchmark serves a crucial role in characterizing the equilibrium with competing ISPs because we establish an equivalence result between the monopolistic outcome of this benchmark and the competitive outcome on the CP side. This benchmark is also useful in introducing our key assumptions. We normalize, without loss of generality, the total measure of consumers to one. We assume that the monopoly simultaneously announces the price-quality pairs for CPs and the fee for consumers.[12]

■ Non-neutral Network

Let $\{(p_H, q_H), (p_L, q_L)\}$ be the menu of contracts offered to CPs which satisfies the incentive and participation constraints of CPs (defined below). Then, each consumer's gross utility is given by $U(\alpha) = \underline{u} + (1-\alpha)\sum_{k=H,L} \nu_k[\theta_k u(q_k)]$, which can be fully extracted by a subscription fee f with homogeneous consumers. The ISP's profit from the content side is $\pi^{CP} = \sum_{k=H,L} \nu_k\ [p_k - cq_k]$. The overall profit for the ISP can be written as $\Pi^M(\alpha) = U(\alpha) + \pi^{CP}$. Thus, the monopolistic

[11] In this setup with homogenous consumers, the monopolistic ISP can extract the whole consumer surplus.

[12] The optimal outcome chosen by the monopoly ISP with simultaneous pricing is the same as the one chosen with sequential pricing in which it first chooses the price-quality pairs for CPs and then the fee for consumers.

ISP's mechanism design problem can be described as:

$$\max_{(p_k,q_k)} \Pi^M(\alpha) = \underline{u} + \sum_{k=H,L} \nu_k[p_k + (1-\alpha)\theta_k u(q_k) - cq_k]$$

subject to

$$\begin{aligned} IC_H &: \alpha\theta_H u(q_H) - p_H \geq \alpha\theta_H u(q_L) - p_L; \\ IC_L &: \alpha\theta_L u(q_L) - p_L \geq \alpha\theta_L u(q_H) - p_H; \\ IR_H &: \alpha\theta_H u(q_H) - p_H \geq 0; \\ IR_L &: \alpha\theta_L u(q_L) - p_L \geq 0, \end{aligned}$$

where IC_k and IR_k refer to type k CPs' incentive compatibility constraint and individual rationality constraint, respectively.

This is a standard mechanism design problem for second-degree price discrimination. As usual, the high-type's incentive compatibility constraint IC_H and the low-type's individual rationality constraint IR_L are binding: we thus have

$$p_H = \alpha\theta_H u(q_H) - \alpha\Delta\theta u(q_L); \qquad p_L = \alpha\theta_L u(q_L). \tag{5}$$

This leads to the following reduced problem

$$\max_{\{q_H,q_L\}} \Pi^M(\alpha, q_H(\alpha), q_L(\alpha)) = \underline{u} + \sum_{k=H,L} \nu_k[\theta_k u(q_k) - cq_k] - \nu \cdot \alpha\Delta\theta u(q_L).$$

The objective in the reduced program shows that the ISP extracts full surplus except for the rent to high type CPs, which is given by $\nu \cdot \alpha\Delta\theta u(q_L)$. From the first order conditions, we find that the optimal quality for the high type is determined by $\theta_H u'(q_H^*) = c$ for any α, which is equal to the first-best level, regardless of α, i.e., $q_H^* = q_H^{FB}$. By contrast, the low type CPs' quality is characterized by

$$\left(\theta_L - \frac{\nu}{(1-\nu)} \cdot \alpha\Delta\theta\right) u'(q_L^*(\alpha)) = c. \tag{6}$$

As in the standard mechanism design problem, there is a downward distortion in quality for the low type, that is, $q_L^*(\alpha) \leq q_L^{FB}$ with the equality holding only for $\alpha = 0$.

We assume that if CPs extract all the surplus from consumers (i.e., $\alpha = 1$), the monopoly

ISP prefers serving both types under second-degree price discrimination.

Assumption 1 $q_L^*(\alpha = 1) > 0$

$q_L^*(\alpha = 1)$ requires $\theta_L > \frac{\nu}{1-\nu}\Delta\theta$. Assumption 1 ensures that $q_L^*(\alpha) > 0$ for any $\alpha \in [0, 1]$ because total differentiation applied to (6) shows that the low-type quality is decreasing in α :

$$\frac{dq_L^*}{d\alpha} = \frac{\nu\Delta\theta u'(q_L^*)}{((1-\nu)\theta_L - \nu\alpha\Delta\theta)\, u''(q_L^*)} < 0. \tag{7}$$

For a given q_L, the rent obtained by a high type CP increases with α. Hence, as the CPs' share of surplus (i.e., α) increases, the ISP has more incentives to distort the quality for the low type. From the envelope theorem, the maximized objective under non-neutral network strictly decreases with α:

$$\frac{d\Pi^M(\alpha, q_H^*(\alpha), q_L^*(\alpha))}{d\alpha} = -\nu \cdot \Delta\theta u(q_L) < 0.$$

■ Neutral Network

Now consider a neutral network where the ISP is constrained to choose only a single price-quality pair (p, q). Given this single quality offer constraint, the ISP decides between serving only the high type CPs with the exclusion of the low type CPs and serving both types of CPs. With the exclusion, it is straightforward that the ISP will choose $q = q_H^{FB}$ and $p = \alpha\theta_H u(q_H^{FB})$, which gives $\widetilde{\Pi}^{EX} = \underline{u} + \nu[\theta_H u(q_H^{FB}) - cq_H^{FB}]$.[13]

If the monopolistic ISP decides to serve both types, then $p = \alpha\theta_L u(q)$ and $f = \underline{u} + (1 - \alpha)\sum_{k=H,L}\nu_k\theta_k u(q)$. Hence, the monopoly ISP chooses a single quality q to solve

$$\max_q \ \widetilde{\Pi}(\alpha) = \underline{u} + (\theta_L + (1-\alpha)\nu\Delta\theta)u(q) - cq.$$

From the first-order condition, we obtain the optimal quality choice when both types of CPs are served:

$$(\theta_L + (1-\alpha)\nu\Delta\theta)u'(\widetilde{q}(\alpha)) = c. \tag{8}$$

Equation (8) indicates that the optimal quality choice lies between the first level qualities for the high and the low types, that is, $q_L^{FB} \leq \widetilde{q}(\alpha) < q_H^{FB}$ to balance the competing needs of the two types with a single quality level, with a higher proportion of the high type CPs (i.e., a higher ν) implying a higher quality provision by the monopolist. By totally differentiating (8), we can

[13]We use a tilde ($\tilde{}$) to denote variables associated with a neutral network.

derive that the quality decreases with α:

$$\frac{d\widetilde{q}(\alpha)}{d\alpha} = \frac{\nu\Delta\theta u'(\widetilde{q})}{(\theta_L + (1-\alpha)\nu\Delta\theta)u''(\widetilde{q})} < 0. \tag{9}$$

In particular, $\widetilde{q}(\alpha) = q_L^{FB}$ when $\alpha = 1$. From the envelope theorem, the monopolistic ISP's profit without exclusion strictly decreases with α as in the non-neutral regime.

$$\frac{d\widetilde{\Pi}(\alpha)}{d\alpha} = -\nu\Delta\theta u(\widetilde{q}(\alpha)) < 0$$

By contrast, the ISP's profit under exclusion, $\widetilde{\Pi}^{EX}$ is independent of α. We assume the following.

Assumption 2 $\widetilde{\Pi}(\alpha = 0) > \widetilde{\Pi}^{EX} > \widetilde{\Pi}(\alpha = 1)$

This assumption, together with the monotonicity of $\widetilde{\Pi}(\alpha)$, implies that there exists a unique threshold level of α denoted by $\alpha^N \in (0,1)$ such that the monopolist ISP serves both types of CPs for $\alpha < \alpha^N$ and excludes the low type CPs $\alpha > \alpha^N$ where α^N is implicitly defined by $\widetilde{\Pi}(\alpha^N) = \widetilde{\Pi}^{EX}$, that is,

$$(\theta_L + (1-\alpha^N)\nu\Delta\theta)u(\widetilde{q}(\alpha^N)) - c\widetilde{q}(\alpha^N) = \nu \cdot (\theta_H u(q_H^{FB}) - cq_H^{FB}). \tag{10}$$

Therefore, the monopolist ISP's profit under the neutral system, $\widetilde{\Pi}^M(\alpha)$, can be written as

$$\widetilde{\Pi}^M(\alpha) = \begin{cases} \widetilde{\Pi}(\alpha) \text{ for } \alpha < \alpha^N \\ \widetilde{\Pi}^{EX} \quad \text{for } \alpha \geq \alpha^N \end{cases}$$

Let $(\widetilde{p}^*(\alpha), \widetilde{q}^*(\alpha))$ represent the ISP's choice under neutral network. Then, the quality chosen by the ISP is given by:

$$\widetilde{q}^*(\alpha) = \begin{cases} \widetilde{q}(\alpha) \text{ for } \alpha < \alpha^N \\ q_H^{FB} \text{ for } \alpha \geq \alpha^N \end{cases}$$

The corresponding retail prices are given by $\widetilde{p}^*(\alpha) = \alpha\theta_L u(\widetilde{q}(\alpha))$ for $\alpha < \alpha^N$ and $\widetilde{p}^*(\alpha) = \alpha\theta_H u(q_H^{FB})$ for $\alpha > \alpha^N$.

■ Assumptions and Social Welfare

Two remarks on our assumptions are in order. First, Assumptions 1 and 2 correspond to a situation analyzed in Hermalin and Katz (2007). Specifically, when a CP extracts the entire surplus, the monopoly ISP prefers excluding the low-type CP without a second degree price discrimination, though it serves both types with it.

Second, from the social welfare point of view, the non-neutral network dominates the neutral network for the extreme cases of $\alpha = 1$ and $\alpha = 0$. Essentially, these two cases can be considered as a representation of one-sided markets. Consider first the case in which CPs capture the whole surplus from interactions with consumers, i.e., $\alpha = 1$. Then, each consumer obtains the basic utility $\underline{u}$ only. So, the monopoly ISP will set $f = \underline{u}$ both under non-neutral and neutral networks. Consequently, we can focus on the monopoly ISP's problem of maximizing profit from CPs, which is a standard problem of one-sided market. In this case, high type CPs consume q_H^{FB} in both regimes, but low types are served only under non-neutral network. This is a standard argument in favor of second-degree price discrimination.

For the other extreme case of $\alpha = 0$, consumers capture all surplus from interactions with CPs. Since consumers are homogeneous, the monopoly ISP can extract full surplus from consumers. The case of $\alpha = 0$ is the same as a standard monopoly in one-sided market with cost function cq. The monopoly will provide services for free to CPs, which means that the ISP bears the entire cost of cq. Under a non-neutral regime, the monopoly ISP provides the first-best quality for each type of CPs and charges the consumer subscription fee $f(\alpha = 0) = \underline{u} + u^{FB}$.[14] The ISP replicates the first-best outcome and earns the profit of $\Pi(\alpha = 0) = \underline{u} + u^{FB} - c^{FB}$. Under a neutral regime, the monopoly ISP is constrained to offer one level of quality and hence can never achieve the first-best outcome. Therefore, the ISP and a social planner prefer a non-neutral network over a neutral network for $\alpha = 0$. In summary, we have:

Proposition 2 *(Monopoly ISP) Consider a monopoly ISP facing homogenous consumers with inelastic subscription.*

(i) If $\alpha = 1$, under Assumptions 1-2, the ISP serves both types of CPs in a non-neutral network but serves only high types in a neutral network. Therefore, social welfare is higher under a non-neutral network than under a neutral network.

(ii) If $\alpha = 0$, the outcome chosen by the ISP coincides with the first-best under a non-neutral network. By contrast, under a neutral network, the first-best can never be realized. Therefore, social welfare is higher under a non-neutral network than under a neutral network.

[14]When $\alpha = 0$, every CP makes zero profit and we can assume that a CP follows the ISP' desire in case of indifference. For any $\alpha > 0$ (hence α can be as close as possible to zero) and $q > 0$, the ISP can exclude low types by charging $p = \alpha\theta_H u(q)$.

Under Assumptions 1-2, we consider a scenario in which the neutral network is always dominated in one-sided market settings, stacking the deck against the neutral network. Later, this result will be contrasted to the case where a neutral network can provide a higher social welfare relative to a non-neutral network.

4 Networks Competition in CP Market and Off-Net Cost Pricing

In this section, we analyze the ISPs' competition in the content market and establish that any equilibrium prices for CPs satisfy the *off-net cost pricing principle*, which we use for subsequent analysis. In the beginning of stage 2, quality levels and access charges are given from stage 1.

LMRT (2003) first showed that in a broad range of environments, network operators set prices for their customers as if their customers' traffic were entirely off-net, which they termed the off-net cost pricing principle. We extend their analysis and confirm that their result is robust to the introduction of heterogenous content types with menu pricing and to alternative timing assumptions.

Lemma 1 *(Off-net cost pricing) Any equilibrium prices that generates positive sales to CPs must satisfy the off-net cost pricing principle. This holds regardless of whether or not networks are neutral.*

Proof. See the Appendix. ■

Lemma 1 shows that off-net cost pricing is a necessary condition that any equilibrium price for CPs generating positive sales must satisfy. This property holds more generally regardless of the timing we consider. In fact, it is straightforward to prove it for the sequential timing of reverse order (i.e., stages 2 and 3 are reversed) or for the simultaneous timing (i.e., stages 2 and 3 take place at the same time). With these alternative timing assumptions, a necessary condition that equilibrium prices on the content side should satisfy is that an ISP should be indifferent between winning a given type of CPs and losing it, given the prices and subscription decisions on the consumer side. For instance, consider a neutral network and let $p(q)$ be an equilibrium price for CPs given that the ISPs previously agreed on (q, a). We normalize the total number of consumers subscribed to one, without loss of generality, and let $s_i \in [0, 1]$ represent ISP i's consumer market share. Suppose first that at $p(q)$ both types of CPs buy connections from ISP i. Then, ISP i's profit from the content side in equilibrium is $p(q) - s_i cq - (1 - s_i)(c + a - c_T) q$.

If it loses the CPs by charging a higher price, its profit from the content side will be $s_i(a - c_T)q$. Therefore, the following inequality must hold in equilibrium:

$$p(q) - s_i cq - (1 - s_i)(c + a - c_T) q \geq s_i(a - c_T)q,$$

which is equivalent to

$$p(q) \geq (c + a - c_T) q = \widehat{c}q.$$

Symmetrically, the condition for ISP j to weakly prefer losing CPs to winning CPs gives the condition

$$p(q) \leq (c + a - c_T) q = \widehat{c}q.$$

Therefore, any equilibrium price should satisfy $p(q) = \widehat{c}q$.

Suppose now that at $p(q)$, only high type CPs buy connection and ISP i wins them. Then, ISP i's equilibrium profit from content is $\nu [p(q) - s_i cq - (1 - s_i)(c + a - c_T) q]$ and its content side profit from losing the CPs is $\nu s_i(a - c_T)q$. The previous logic still applies here again. Therefore, any equilibrium price should satisfy $p(q) = \widehat{c}q$, regardless of whether exclusion of low types occurs or not. The same result holds when we consider a non-neutral network.

We note that the proof of lemma 1 is more involved since under the sequential timing we consider, any ISP's deviation in the content market is followed by both ISPs' reactions in the consumer market and hence we need to explicitly take into account these reactions in our proof.

We also would like to point out that even though off-net cost pricing is a necessary condition for an equilibrium, it is not a sufficient condition for any arbitrary access charge because an ISP may have an incentive to deviate.[15] However, we show in section 6 that it is both necessary and sufficient for the access charge(s) optimally agreed on by the ISPs (to maximize their joint profits) in the first stage.

[15] To illustrate this point, consider a neutral network and the sequential timing of reverse order (i.e., stage 2 and stage 3 are reversed). Suppose that $a = (\alpha\theta_L u(q) + \varepsilon)/q - c_O$ where $\varepsilon > 0$ is infinitesimal. Then, off-net cost pricing leads to $p(q) = \alpha\theta_L u(q) + \varepsilon$. Hence, only high types purchase the quality at off-net cost pricing. Consider now the deviation of ISP i to $p'(q) = \alpha\theta_L u(q)$ such that both types purchase the quality. This deviation is profitable in the content side if and only if

$$\nu s_i(\alpha\theta_L u(q) + \varepsilon - cq) < \alpha\theta_L u(q) - s_i cq - (1 - s_i)(\alpha\theta_L u(q) + \varepsilon),$$

where s_i is ISP i's market share in the content side at the off-net cost. The condition is equivalent to

$$[\nu s_i + (1 - s_i)]\, \varepsilon < (1 - \nu)s_i(\alpha\theta_L u(q) - cq),$$

which holds for $\varepsilon > 0$ small enough as long as $\alpha\theta_L u(q) > cq$.

Now we examine the profit that each ISP obtains on the content side under the off-net cost pricing. Let n_i be the number of consumers subscribed to ISP i for $i = 1, 2$ at stage 3. Consider a given CP who uses quality q under the off-net cost pricing $\widehat{c}q$. If this CP subscribed to ISP i at stage 2, then ISP i's profit from this CP is,

$$\widehat{c}q(n_i + n_j) - cqn_i - (c + a - c_T)\,qn_j = (\widehat{c} - c)qn_i.$$

If this CP subscribed to ISP j at stage 2, then ISP i's profit from this CP is,

$$(a - c_T)qn_i = (\widehat{c} - c)qn_i.$$

Therefore, we derive:

Lemma 2 *(Profit from CPs) Consider any off-net cost pricing equilibrium. Then, ISP i's profit from the CP side is given by $n_i\widehat{\pi}^{CP}$, where n_i is the number of consumers subscribed to ISP i for $i = 1, 2$ and $\widehat{\pi}^{CP} \equiv \sum_{k=H,L} \nu_k^*(a_k - c_T)q_k = \sum_{k=H,L} \nu_k^*(\widehat{c}_k - c_T)q_k$ where ν_k^* is the measure of type $k(= H, L)$ CPs which subscribed to any of the two ISPs. This result holds regardless of whether networks are neutral or not. In the neutral network, if both types are served, it is required that $a_H = a_L = a$.*

Note that the result of this lemma does not depends on the subscription distribution of CPs across the two ISPs because each ISP is indifferent between winning and losing CPs at off-net cost pricing. Note that $\widehat{\pi}^{CP}$ represents the profit *per consumer* that each ISP makes from the content side with off-net cost pricing and does not depend on (n_1, n_2).

5 Networks Competition in Consumer Subscription Market

In the previous section, we showed that off-net cost pricing must be satisfied in any equilibrium and that under off-net cost pricing, each ISP's profit from the content side is given by $n_i\widehat{\pi}^{CP}$. Given these results, let us study the competition between two ISPs in the consumer subscription market. With off-net cost pricing in the content side market, ISP i's total profit can be written as

$$\Pi_i = n_i \cdot (f_i + \widehat{\pi}^{CP}), \text{ where } n_i = \frac{1}{2} + \frac{f_j - f_i}{2t} + \lambda u_i.$$

In any interior equilibrium of the third stage competition for consumers, each ISP i chooses f_i to maximize its total profit from both CPs and consumers, given f_j. With the first order condition, symmetry, and the relationship $\lambda = \frac{h}{t}$, we can derive the symmetric equilibrium subscription price:

$$f^*(\widehat{\pi}^{CP}; h) = \frac{t + 2hU(\alpha)}{1+4h} - \frac{(1+2h)\widehat{\pi}^{CP}}{1+4h}, \tag{11}$$

where $U(\alpha) = \underline{u} + (1-\alpha) \sum_{k=H,L} \nu_k^* \theta_k u(q_k)$ is the consumer gross utility. Then each ISP's equilibrium profit is given by

$$\Pi_i^*(\alpha; \widehat{\mathbf{c}}) = \left(\frac{1}{2} + \lambda \left(U(\alpha) - f^*\right)\right) \cdot (f^* + \widehat{\pi}^{CP}),$$

where f^* is from (11). With some algebra, we find that

Proposition 3 *Consider any symmetric equilibrium.*

(i) Each ISP earns the profit

$$\Pi_i^*(\alpha; \widehat{\mathbf{c}}) = \left[\frac{1}{2} + \lambda \frac{(1+2h) \cdot \widehat{\Pi}(\alpha) - t}{1+4h}\right] \left[\frac{t + 2h \cdot \widehat{\Pi}(\alpha)}{1+4h}\right] \tag{12}$$

where $\widehat{\Pi}(\alpha) = U(\alpha) + \widehat{\pi}^{CP}$ *and* $\lambda = \frac{h}{t}$.

(ii) For $h = t\lambda = 0$, each ISP's profit is always equal to the standard Hotelling profit of $t/2$.

(iii) For $h = t\lambda > 0$, maximizing joint market equilibrium profit of the ISPs is equivalent to maximizing $\widehat{\Pi}(\alpha) = U(\alpha) + \widehat{\pi}^{CP} = \Pi^M(\alpha)\big|_{p=\widehat{c}q}$. In other words, the ISPs maximizing joint profit behave as a monopoly ISP facing homogeneous consumers with inelastic subscription.

The result of Proposition 3-(iii) is worth elaborating on. Note first that for $h = t\lambda > 0$, the equilibrium profit (12) is an increasing function of $\widehat{\Pi}(\alpha)$. This implies that what the competing ISPs maximize when jointly choosing quality levels and access charges is equivalent to what a monopoly ISP would maximize when it faces homogenous consumers with inelastic subscription, which we analyzed as a benchmark in section 3.2: $U(\alpha) + \widehat{\pi}^{CP}$ is exactly what the monopoly ISP of the benchmark maximizes. Recall that $\widehat{\Pi}(\alpha) \leq \underline{u} + u^{FB} - c^{FB}$, where the equality holds only when the ISPs capture the entire CPs' surplus with the first-degree price discrimination. This implies that there are potentially two sources of distortions in the objective of the ISPs compared to social welfare: *the ISPs neglect the rent of the CPs and endogenous subscription of* *consumers*. Later, we will study how these two distortions play differently in non-neutral and neutral networks.

What drives our result is that for any given $\left(U(\alpha), \widehat{\pi}^{CP}\right)$, the competition on consumer side leads to an equilibrium subscription fee of the form

$$f^* = \beta U(\alpha) - (1-\beta)\widehat{\pi}^{CP} + \frac{t}{1+4h}, \tag{13}$$

where $\beta = \frac{2h}{1+4h} \in (0, 1/2)$ for $h = t\lambda > 0$ and $\beta = 0$ for $h = t\lambda = 0$. When consumer subscription is inelastic ($\lambda = 0$), the equilibrium consumer subscription fee is given by $f^* = t - \widehat{\pi}^{CP}$ and hence each ISP obtains the standard Hotelling profit $t/2$. In this case, any profit from the content side is completely dissipated away in the competition for the consumer side. Imposing net neutrality has no impact on each ISP's profit. This is reminiscent of the *profit neutrality* result in the literature on the competition of telecommunications networks (Laffont, Rey, Tirole, 1998a) where the access charge has no impact on the networks' profits in a Hotelling model (hence with inelastic consumer subscription). Our result is stronger in the sense that the profit depends neither on the level of access charge nor on the number of product lines ISPs are allowed to offer.

When consumer subscription is elastic with $h = t\lambda > 0$, the profit neutrality result no longer holds; the pass-through rate of $\widehat{\pi}^{CP}$ into f is not complete and only partial. As $\widehat{\pi}^{CP}$ increases, the consumer subscription price decreases but *less* than the change in $\widehat{\pi}^{CP}$ for any $h > 0$. This can be seen from

$$-1 < \frac{\partial f^*}{\partial \widehat{\pi}^{CP}} = -(1-\beta) < -\frac{1}{2}. \tag{14}$$

Inequality (14), in turn, implies that each ISP's total profit per consumer increases with $\widehat{\pi}^{CP}$:

$$0 < \frac{\partial (f^* + \widehat{\pi}^{CP})}{\partial \widehat{\pi}^{CP}} = \beta < \frac{1}{2}.$$

More generally, an ISP's profit is equal to the number of consumers multiplied by the profit per consumer. The number of consumers is given by $\frac{1}{2} + \lambda(1-\beta)\widehat{\Pi}(\alpha)$ plus a constant, which linearly increases with $\widehat{\Pi}(\alpha)$. Furthermore, from (13), profit per consumer $f^* + \widehat{\pi}^{CP}$ is given by $\beta\widehat{\Pi}(\alpha)$ plus a constant and hence linearly increases with $\widehat{\Pi}(\alpha)$. Therefore, the ISPs will choose quality levels and access charges to maximize $\widehat{\Pi}(\alpha)$ and replicate the monopolistic solution derived in section 3.2.

6 ISPs' Choice of Quality and Access Charges

In this section, we analyze the ISPs' choice of quality levels and corresponding access charges. We showed that the competing ISPs maximize the same objective as a monopoly ISP facing homogenous consumers and that off-net cost pricing must hold in any equilibrium. One potential issue is that not all off-net costs can be supported as equilibrium prices for CPs since an ISP might have an incentive to deviate from off-net cost pricing in stage 2. However, we show that for the access charge optimally agreed on by the ISPs, this issue does not arise and off-net cost pricing is both necessary and sufficient for an equilibrium in the CP market.

We proceed in two steps. First, we consider a constrained benchmark case in which no ISP is allowed to deviate from the off-net cost pricing in stage 2. Therefore the ISPs behave the same way as the monopoly ISP facing homogenous consumers would behave in section 3.2. This is because there is one-to-one correspondence between the retail price of content delivery and the choice of access charge from the off-net cost pricing, $p(q_k) = \widehat{c}_k q = (c + a_k - c_T)q_k$. In other words, any second-best quality-price combinations that would be chosen by the monopolistic ISP can be replicated by agreeing to the same quality levels and appropriate choice of access fees. Essentially, both the monopolistic ISP and competing ISPs have the same objective function and the same instruments.

Second, we consider the original case in which any ISP is allowed to deviate from off-net cost pricing in stage 2 and prove that there is no profitable deviation when the ISPs agree on the qualities and access charges that would implement the monopoly benchmark outcome.

6.1 Non-Neutral Network

We now consider ISPs' choice of quality levels and access charges in the non-neutral network for $h = t\lambda > 0$. According to Proposition 3-(iii), the ISPs collectively choose the quality levels and the corresponding access charges to maximize $\widehat{\Pi}(\alpha)$. Furthermore, according to Proposition 2, maximizing $\widehat{\Pi}(\alpha)$ in the non-neutral network requires to serve both types of CPs (i.e., no exclusion). Given the quality pairs (q_H, q_L) which the ISPs agreed to offer to each type of CPs, no ISP is allowed to deviate from the off-net cost pricing in the constrained benchmark case. Given the off-net cost pricing constraint, the ISPs can indirectly choose the equilibrium price of each quality by appropriately choosing the access charge.

From (5) and off-net cost pricing, it is immediate that the access charges will be chosen as follows to replicate the monopolistic solution:

$$a_H^* = \alpha\left(\theta_H u(q_H^*) - \Delta\theta u(q_L^*)\right)/q_H^* - c_O \qquad \text{and} \qquad a_L^* = \alpha\theta_L u(q_L^*)/q_L^* - c_O. \tag{15}$$

Proposition 4 *Consider a non-neutral network under Assumption 1:*

(i) Suppose that no ISP is allowed to deviate from the off-net cost pricing.

(a) The ISPs offer quality levels (q_H^*, q_L^*) such that $q_H^* = q_H^{FB}$ for any $\alpha \in [0,1]$ and $q_L^*(\alpha)$ is determined by (6): $q_L^*(0) = q_L^{FB}$ and $q_L^*(\alpha)$ strictly decreases with α.

(b) The ISPs choose access charges (a_H^*, a_L^*) given by (15). This leads to the following retail prices for CPs: $p_H^* = \alpha\theta_H u(q_H^*) - \alpha\Delta\theta u(q_L^*(\alpha))$ and $p_L^* = \alpha\theta_L u(q_L^*(\alpha))$.

(ii) When the ISPs agreed on (q_H^*, q_L^*) and (a_H^*, a_L^*), there is no profitable deviation from the off-net cost pricing in stage 2 and the ISPs implement the outcome that would be chosen by a monopoly ISP facing homogeneous consumers.

The proof of Proposition 4-(ii) is provided in the Appendix.

6.2 Neutral network

As in the non-neutral network, the monopolistic ISP solution can be replicated by an appropriate choice of the access charge if they are not allowed to deviate from the off-net cost pricing. More specifically, the ISPs will serve only high type CPs for $\alpha > \alpha^N$, and they will cooperatively choose the delivery quality level of $\widetilde{q}^*(\alpha) = q_H^{FB}$ and the access charge of $\widetilde{a}^*(\alpha) = \alpha\theta_H u(q_H^{FB})/q_H^{FB} - c_O$ to replicate the monopolistic solution. For $\alpha < \alpha^N$, the ISPs choose to serve both types of CPs with $\widetilde{q}^*(\alpha) = \widetilde{q}(\alpha)$ and the corresponding access charge of $\widetilde{a}^*(\alpha) = \alpha\theta_L u(\widetilde{q}(\alpha))/\widetilde{q}(\alpha) - c_O$.

Proposition 5 *Consider a neutral network under Assumption 2.*

(i) Suppose that no ISP is allowed to deviate from the off-net cost pricing. Then there exists a unique threshold level of α, denoted by $\alpha^N \in (0,1)$.

(a) The ISPs offer $\widetilde{q}^*(\alpha) = q_H^{FB}$ for $\alpha > \alpha^N$ and $\widetilde{q}^*(\alpha) = \widetilde{q}(\alpha)$ otherwise. $\widetilde{q}(\alpha)$ is determined by (8) and it is higher than q_L^{FB} and strictly decreases with α. The cut-off value α^N is determined by (10). The ISPs serve only high type CPs for $\alpha > \alpha^N$ and serve both types otherwise.

(b) *The ISPs choose an access charge* $\widetilde{a}^*(\alpha) = \alpha\theta_H u(q_H^{FB})/q_H^{FB} - c_O$ *for* $\alpha > \alpha^N$ *and* $\widetilde{a}^*(\alpha) = \alpha\theta_L u(\widetilde{q}(\alpha))/\widetilde{q}(\alpha) - c_O$ *otherwise. This generates a retail price for CPs such that* $\widetilde{p}^*(\alpha) = \alpha\theta_H u(q_H^{FB})$ *for* $\alpha > \alpha^N$ *and* $\widetilde{p}^*(\alpha) = \alpha\theta_L u(\widetilde{q}(\alpha))$ *otherwise.*

(ii) *When the ISPs agreed on* $\widetilde{q}^*(\alpha)$ *and* $\widetilde{a}^*(\alpha)$*, there is no profitable deviation from the off-net cost pricing in stage 2 and the ISPs implement the outcome that would be chosen by a monopoly ISP facing homogeneous consumers.*

The proof of Proposition 5-(ii) is provided in the Appendix.

The parameter α represents the surplus division between CPs and end users when they interact through ISPs and indicates which side to focus for the ISP to extract rents. As α increases, CPs capture more surplus and the extraction of rents from the CP side becomes more important. As a result, ISPs distort the quality for the low type CPs further down to reduce the rent of the high type CPs under the non-neutral network. Under the neutral network, ISPs exclude the low type CPs when α is high enough (i.e., $\alpha > \alpha^N$) to reduce the rents of the high type CPs under assumption 2. This finding may have important policy implications. For instance, α may capture how much CPs can extract consumer surplus through micropayments. From this perspective, the concern about potential exclusion of CPs in a neutral network can be heightened if the business model of CPs shifts from advertising-based one with free access to the one with micropayments that directly charge consumers for content.

A simple application of the implicit function theorem to equation (10) that defines the critical value α^N yields the following comparative statics results.

Corollary 1 $\frac{\partial\alpha^N}{\partial c} = -\frac{\widetilde{q}^*}{\nu\Delta\theta u(\widetilde{q}^*)} < 0$ *and* $\frac{\partial\alpha^N}{\partial\nu} = \frac{(1-\alpha^N)\Delta\theta u(\widetilde{q}^*)-(\theta_H u(q_H^{FB})-cq_H^{FB})}{\nu\Delta\theta u(\widetilde{q}^*)} = \frac{c\widetilde{q}^*-\theta_L u(\widetilde{q}^*)}{\nu^2\Delta\theta u(\widetilde{q}^*)} < 0$

As Corollary 1 shows, the exclusion strategy is more likely to occur when the marginal cost of delivery increases and the proportion of high-type CP increases, with all other things being equal. This result has some implication for mobile Internet networks that are constrained by the scarcity of bandwidth imposed by physical laws and thus have a higher delivery cost (i.e., higher c) compared to fixed Internet networks with fiber optic cables; the non-neutral network is likely to increase the allocative efficiency and may provide justifications for differential treatments of mobile networks.

6.3 Comparison: quality choices and each group's payoff

Figure 2 shows the optimal quality schedules for both network regimes. In a non-neutral network, there is no distortion in the quality for high type CPs and a downward distortion in the quality for low type CPs. As α decreases, this distortion becomes smaller and becomes zero when $\alpha = 0$ (i.e., $q^*(\alpha = 0) = q_L^{FB}$). In a neutral network, the ISPs serve only high type CPs for $\alpha > \alpha^N$ and choose $\widetilde{q}^*(\alpha) = q_H^{FB}$; for $\alpha \leq \alpha^N$, they choose a quality $\widetilde{q}^*(\alpha)(> q_L^{FB})$ and serve both types where $\widetilde{q}^*(\alpha)$ decreases with α.

The ISPs realize a strictly higher profit in a non-neutral network than in a neutral network by the revealed preference argument; in a non-neutral network the ISPs could always choose an equal quality for both delivery services if this would give a higher profit. Since the equilibrium number of subscribed consumers increases with consumer surplus (gross of transportation cost) and this consumer surplus increases with the ISPs' profits,[16] consumer surplus is higher in the non-neutral network than in the neutral network.

Low type CPs always receive zero rents for any α regardless of net neutrality regulation. Comparison of high type CPs' surplus depends on whether α is higher or lower than α^N. If $\alpha \leq \alpha^N$, the relationship of $\widetilde{q}^*(\alpha) > q_L^{FB} \geq q_L^*(\alpha)$ implies that high type CPs obtain a higher payoff in the neutral network than in the non-neutral network. If $\alpha > \alpha^N$, the reverse holds since high type CPs obtain no rent in the neutral network while they obtain a strictly positive rent in the non-neutral network.

6.4 Bill and Keep

Here we further discuss access charges chosen by ISPs. In particular, we study conditions under which ISPs choose "bill and keep" (i.e., zero access charge), a widely used access charge settlement mechanism in a peering arrangement between ISPs. In this regard, we point out that implementing negative access charges is impractical because of the ISPs' incentives to game the system by sending artificially generated traffics to a rival ISP to generate positive revenues. Thus, we expect the ISPs to adopt bill and keep if their preferred access charges are negative.

Let us first study the access charges in the first-best world. The prices described in Proposition 1 are the unique ones that implement the first-best outcome under the budget constraint: each ISP and each CP realizes zero profit and the marginal consumer is indifferent between subscribing

[16] Recall the result in section 5 that the number of consumers is given by $\frac{1}{2} + \lambda(1-\beta)\widehat{\Pi}(\alpha)$ plus a constant, which linearly increases with $\widehat{\Pi}(\alpha)$.

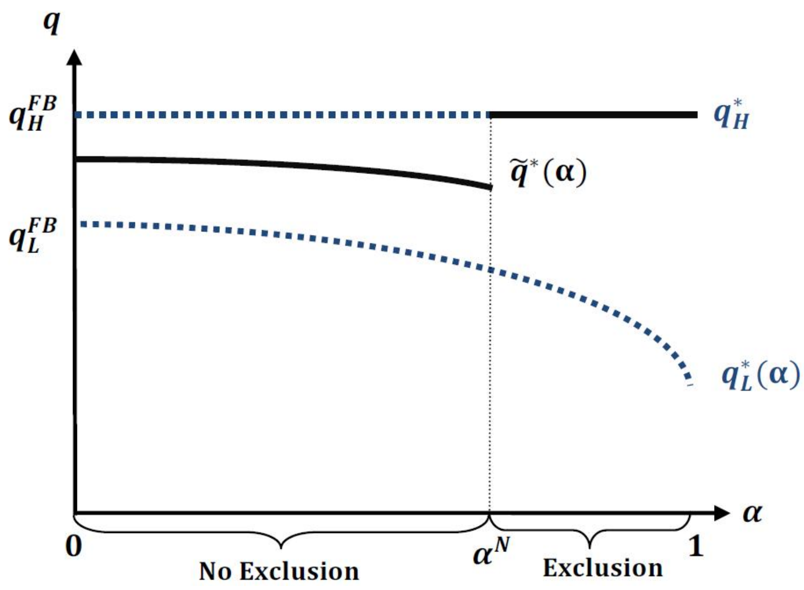

Figure 2: The optimal quality schedules

and not subscribing. In particular, the price charged for each CP of type θ to use service of quality $q^{FB}(\theta)$ is $\alpha\theta u(q^{FB}(\theta))$. Suppose now that a social planner chooses access charges to implement these prices for CPs through off-net cost pricing. Then, the socially optimal access charge is given by

$$a^{FB}(\theta) = \alpha\theta u(q^{FB}(\theta))/q^{FB}(\theta) - c_O. \tag{16}$$

Thus, we have

$$\text{Bill \& Keep (i.e., } a^{FB}(\theta) \leq 0) \qquad \text{iff} \qquad \alpha\theta u(q^{FB}(\theta))/q^{FB}(\theta) - c_O \leq 0.$$

It shows that bill and keep is optimal if α is low enough. When α is small enough, most surplus from interaction between a consumer and a CP is captured by the consumer. Therefore, the social planner finds it optimal to subsidize content side by charging access charges below the termination cost. Propositions 4 and 5 show that this property is qualitatively preserved when access charges are chosen by ISPs that maximize their joint profits, regardless of whether networks are neutral or not. ISPs have an incentive to subsidize content side while making profits from consumer side when α is small.

Specifically, the comparison between (15) and (16) yields the following result: In the non-neutral network, we have

$$a_H^* < a^{FB}(\theta_H) \quad \text{and} \quad a_L^* > a^{FB}(\theta_L).$$

The market equilibrium access charge for the high type content, a_H^*, is lower than the first-best level because of the need to provide rents to high type CPs to satisfy the IC constraint. By contrast, the market access charge for the low type content, a_L^*, is higher than the first-best level due to the downward distortion in the quality for low type CPs. This implies that bill and keep would be more likely to be chosen for high type CPs while it is less so for low type CPs in the market equilibrium compared to the first-best world.

In the neutral network, conditional on no exclusion, $\widetilde{q}(\alpha) > q_L^*(\alpha)$ implies

$$\widetilde{a}^* < a_L^*.$$

In addition, if $\Delta\theta\left[u(q_H^*) - u(q_L^*(\alpha))\right]/q_H^* > \theta_L\left[u(\widetilde{q}(\alpha))/\widetilde{q}(\alpha) - u(q_H^*)/q_H^*\right]$ holds, we obtain

$$\widetilde{a}^* < a_H^*.$$

In the absence of exclusion, the neutral network thus would choose bill and keep more often than the non-neutral network.

7 Social Welfare: Neutral vs. Non-Neutral Network

In this section, we compare social welfare under neutral and non-neutral networks to assess the merit of net neutrality regulations. Recall that a neutral network cannot outperform a non-neutral network for the two extreme cases of full or zero extraction of surplus by CPs vis-à-vis consumers (see Proposition 2). For $\alpha = 1$, the neutral network without second-degree price discrimination results in the exclusion of low type CPs, while the non-neutral network does not entail such exclusion.[17] For $\alpha = 0$, the neutral network provides a suboptimal quality for both types of CPs, while the non-neutral network provides the first-best quality for each type of CP. We investigate whether this result is robust to intermediate cases of $\alpha \in (0, 1)$ and find that the social welfare ranking between the two regimes would be reversed for intermediate values of α.

As we provided the intuition for this result in the introduction, the single quality restriction

[17] By assumptions 1 and 2, we have limited our attention to the parameter space where the non-neutral network serves both types while the neutral network entails exclusion of low type CPs.

can be welfare-enhancing because of its smaller quality distortion in a neutral network compared to a non-neutral network, despite offering a suboptimal quality for the high type CPs. Note that the ISPs have two sources of revenues, the one from the CP side and the one from the consumer side. When the ISPs choose the quality for low type CPs, they face a trade-off between extracting the rent of the high type CPs and extracting the consumer surplus. More precisely, a downward distortion in the quality for low type CPs has the benefit of extracting more rent from high type CPs and the cost of extracting less consumer surplus. This implies that when the ISPs focus on making revenue from consumer side rather than from CP side, there will be less distortion in quality. Because the ISPs have more instruments to extract CPs' surplus in a non-neutral network than in a neutral network, it can happen that they focus on extracting CPs' surplus under non-neutral networks while they focus on extracting consumer surplus under neutral networks. In our model, this arises when α becomes smaller than α^N. For $\alpha \in \left(0, \alpha^N\right)$, the ISPs always distort downward the quality for low type CPs under non-neutral networks whereas under neutral networks, they serve both types of CPs with a quality higher than q_L^{FB}. For this reason, neutral networks may provide a higher social welfare than non-neutral networks.[18]

Given this insight, now we provide more rigorous mathematical derivation for our result. The social welfare in the non-neutral network with optimal quality choices can be written as

$$W^* = N^* \cdot \omega^* - T(N^*) \tag{17}$$

where ω^* the net social surplus per consumer at the optimal quality choices with $q_H^* = q_H^{FB}$ and $q_L^* = q_L^*(\alpha)$, that is,

$$\omega^* = \underline{u} + \sum_{k=H,L} \nu_k[\theta_k u(q_k^*) - cq_k^*] = \underline{u} + \nu(\theta_H u(q_H^{FB}) - cq_H^{FB}) + (1-\nu)(\theta_L u(q_L^*(\alpha)) - cq_L^*(\alpha)).$$

Recall that the number of consumers is given by $N^* = 2\left[\frac{1}{2} + \lambda \frac{(1+2h)\cdot\widehat{\Pi}(\alpha) - t}{1+4h}\right]$ where $\widehat{\Pi}(\alpha) = \omega^* - \nu\alpha\Delta\theta u(q_L^*(\alpha))$. We take the first-order derivative of the social welfare with respect to α as

$$\frac{dW^*}{d\alpha} = N^* \frac{\partial \omega^*}{\partial \alpha} + (\omega^* - T'(N^*)) \frac{\partial N^*}{\partial \alpha}. \tag{18}$$

[18] If we relax Assumption 2 and consider the case where the neutral networks entail no exclusion and the quality distortion effect is high enough in the non-neutral networks, social welfare may be higher in the neutral networks when α is close or equal to 1.

The first term in (18) is negative as $\frac{\partial \omega^*}{\partial \alpha} < 0$: this is because the quality distortion increases in α, i.e., $\frac{dq_L^*}{d\alpha} < 0$. Its second term is also negative as $\omega^* - T'(N^*) > 0$ and $\frac{\partial N^*}{\partial \alpha} < 0$. Notice that $\omega^* - T'(N^*)$ is positive as long as N^* is smaller than the first-best level, which holds for any $\alpha \in [0, 1]$.[19] We have seen in the monopoly benchmark that the monopoly profit decreases as α increases, which implies that N^* decreases with α. Using the first-order optimal quality condition for the low-type CPs, $(1 - \nu)(\theta_L u'(q_L^{SB}) - c) - \nu\alpha\Delta\theta u'(q_L^{SB}) = 0$ (see equation (6)), we obtain an expression as follows:

$$\frac{dW^*}{d\alpha} = N^* \cdot \nu\alpha\Delta\theta u'(q_L^*)\frac{dq_L^*}{d\alpha} - (\omega^* - T'(N^*))2\lambda\frac{(1+2h)}{1+4h}\nu\Delta\theta u(q_L^*) < 0. \tag{19}$$

Similarly, we can define social welfare under the neutral network and for the same reason as in the non-neutral network, we find that the social welfare in the neutral network also decreases in α for any $\alpha < \alpha^N$:

$$\frac{d\widetilde{W}^*}{d\alpha} = \widetilde{N}^* \cdot \alpha\nu\Delta\theta u'(\widetilde{q}^*)\frac{d\widetilde{q}^*}{d\alpha} - \left(\widetilde{\omega}^* - \widetilde{T}^{*\prime}\right)2\lambda\frac{(1+2h)}{1+4h}\nu\Delta\theta u(\widetilde{q}^*) < 0. \tag{20}$$

Recall that the quality adjustment to the change in α can be derived as (7) for the non-neutral network and (9) for the neutral one.

To facilitate the comparison further and gain more intuition, let us consider a utility function with Arrow-Pratt constant absolute risk aversion (CARA), e.g., $u(q) = A - \frac{B}{r}\exp(-rq)$ where r measures the degree of risk aversion with positive constants A and B. Then, we can obtain a clear comparison of

$$\left|\frac{dq_L^*(\alpha)}{d\alpha}\right| > \left|\frac{d\widetilde{q}^*}{d\alpha}\right| \quad \text{for } \forall\nu \in (0, 1) \tag{21}$$

from $-\frac{u'(\widetilde{q}^*)}{u''(\widetilde{q}^*)} = -\frac{u'(q_L^*)}{u''(q_L^*)} = r$. This implies that the ISPs' quality degradation gradient for the low-type in the non-neutral network is steeper than the one for the uniform quality in the neutral network as α increases. In addition, we find $u'(q_L^*) > u'(\widetilde{q}^*)$ from $q_L^* < \widetilde{q}^*$ for any utility function with $u'' < 0$. Since the non-neutral network provides consumer surplus (gross of transportation cost) at least as high as that under the neutral network, we have $N^* \geq \widetilde{N}^*$ where the equality holds for $\lambda = 0$. Hence, we find that the social welfare decreases more quickly as α increases in the non-neutral network compared to the neutral network, i.e., $\left|\frac{dW^*}{d\alpha}\right| > \left|\frac{d\widetilde{W}^*}{d\alpha}\right|$ if the market

[19]Even if $\alpha = 0$ and hence the ISPs choose the first-best qualities, N^* is smaller than the first-best level since the first-best outcome can be implemented only with zero profit of the ISPs.

expansion is highly limited ($\lambda \approx 0$).

Given this understanding, let us finally compare the level of social welfare under two different network regimes. Recalling the definition of α_N in (10), the per consumer net surplus in the neutral network at $\alpha = \alpha^N$ can be expressed as

$$\begin{aligned}
\widetilde{\omega}^*|_{\alpha=\alpha^N} &= \underline{u} + (\theta_L + (1-\alpha^N)\nu\Delta\theta)u(\widetilde{q}^*) - c\widetilde{q}^* + \alpha^N\nu\Delta\theta u(\widetilde{q}^*) \\
&= \underline{u} + \nu \cdot (\theta_H u(q_H^{FB}) - cq_H^{FB}) + \alpha^N\nu\Delta\theta u(\widetilde{q}^*)
\end{aligned}$$

This simplifies the comparison between $\widetilde{\omega}^*$ and ω^* evaluated at $\alpha = \alpha^N$ as the comparison between $\alpha^N\nu\Delta\theta u(\widetilde{q}^*)$ and $(1-\nu)(\theta_L u(q_L^{SB}) - cq_L^{SB})$:

$$\widetilde{\omega}^*|_{\alpha=\alpha^N} > \omega^*|_{\alpha=\alpha^N} \Leftrightarrow \alpha^N\nu\Delta\theta u(\widetilde{q}^*) > (1-\nu)(\theta_L u(q_L^*) - cq_L^*), \tag{22}$$

Under Assumptions 1-2, per consumer social welfare can be higher in a neutral network relative to a non-neutral network as long as (22) is satisfied. Note that this does not ensure that a neutral network always dominates a non-neutral network at α^N because of $N^* \geq \widetilde{N}^*$. However, for a sufficiently small λ, the difference in number of consumers subscribed is of second-order relative to the difference in per consumer welfare. Hence, we can state that

Proposition 6 *Consider the CARA utility function for which Assumptions 1 and 2, and (22) are satisfied. For a sufficiently small λ, there exists always intermediate level of α, which is weakly smaller than α^N but is greater than zero, under which the total social surplus also is higher in the neutral network than in the non-neutral network.*

Figure 3 illustrates a plausible case in which a neutral network may yield a higher total social welfare than a non-neutral network. Let us wrap up our discussion on social welfare with numerical simulations that verify our findings. Consider a parameter space of $\theta_H = 40$, $\theta_L = 30$, $\underline{u} = 5$, $\nu = 0.74$, $t = 1$, and $h = 1/4$. If we consider a CARA utility function such as $u(q) = 1 - \frac{1}{2}e^{-2q}$, it satisfies all the assumptions that we make and the neutral network yields the higher social surplus than the non-neutral network over the range of $\alpha \in [0.4334, 0.9901]$. When we consider a quadratic utility function $u(q) = \sqrt{q}$, then for the same parameters we find the neutral network provide the higher social welfare compared to the non-neutral network for $\alpha \in [0.5254, 0.8711]$.

Through the sequence of intuition, analysis, and numerical simulations, we find that the merit

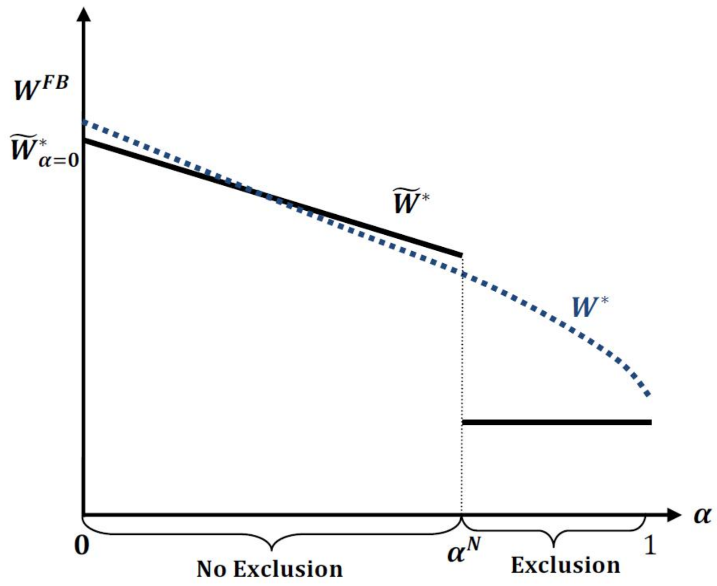

Figure 3: Welfare under neutral and non-neutral networks

of net neutrality regulation may depend crucially on content providers' business models. Though we frame our discussion in the context of the recent debate on network neutrality in the Internet, our analysis can be more generally interpreted. In particular, our findings suggest that welfare implications of second-degree price discrimination in two-sided platform markets can crucially depend on the relative allocation of the total surplus between the two sides.

8 Concluding Remarks

In this paper we have analyzed competition between interconnected networks when content is heterogeneous in its sensitivity to delivery quality. The heterogeneity of content calls for the multi-tiered Internet to reflect the need for differential treatment of packets depending on its sensitivity. With interconnected networks, however, the assurance of a certain level of delivery quality requires cooperation among networks. To address this issue, we have developed a framework of two-sided markets in which ISPs compete with each other to serve as platforms that connect CPs and end consumers. We have considered two regimes under which packets can be delivered: a neutral regime in which all packets are required to be delivered with the same quality (speed) and a non-neutral regime under which ISPs are allowed to offer multi-tiered services with different delivery quality levels. We derived conditions under which social welfare can be higher in a neutral network. The conditions highlight the importance of CPs' business models in the evaluation of net neutrality regulation.

Looking forward, this paper is a first step towards incorporating heterogeneous content in the analysis of interconnection issues. There are many worthwhile extensions that call for further analysis. One limitation of our analysis is its static nature. We have not analyzed dynamic investment incentives facing ISPs and CPs by assuming away capacity constraints for ISPs and by considering a fixed mass of active CPs. The effects of net neutrality regulation on ISPs' capacity expansions and CPs' entry decisions are important issues.[20]

It would be an important research agenda to develop a model that can capture differences between mobile networks and fixed networks. Mobile networks are becoming an increasingly important channel for Internet content delivery. Fixed and mobile Internet networks are inherently different in many dimensions, most importantly the scarcity of bandwidth for mobile networks imposed by physical laws. These differences are recognized by the recent FCC rule on net neutrality. The new FCC rule, announced on December 21, 2010, reaffirmed its commitment to the basic principle of net neutrality by prohibiting ISPs from "unreasonable discrimination" of web sites or applications, but wireless telecommunications were exempted from such anti-discrimination rules. Our model may lend a new justification for asymmetric regulation between fixed and mobile networks. For instance, imagine a situation in which the mobile networks are more constrained in their capacity and expansion possibilities. The network operators thus may prefer to serve only the high type CPs under neutral network, instead of providing somewhat jittery content delivery by serving uniform speed to heterogeneous CPs. If CPs in the mobile networks adopt business models of more content-usage based charge system that enables them to extract more surplus from consumers than the ad-financed system, our model suggests that net neutrality regulation may be beneficial for fixed networks but not for mobile networks.

Finally, we assumed a homogeneous and exogenous business model by assuming the same level of surplus extraction (parameterized by α) for CPs. The analysis can be extended to heterogeneous business models that would be endogenously derived.

[20] Choi and Kim (2010) addresses the dynamic investment issue, but with a monopolistic ISP. Njoroge, Ozdaglar, Stier-Moses, and Weintraub (2010) study investment incentives with multiple ISPs, but neither interconnection between ISPs nor the role of CPs' business model in net neutrality regulation is considered.

References

[1] Armstrong, M. “Network Interconnection in Telecommunications.” *Economic Journal*, Vol. 108 (448), 1998: pp 545-564.

[2] Armstrong, M. “Competition in Two-Sided Markets,” *Rand Journal of Economics*, Vol. 37(3), Autumn 2006, pp. 668-691.

[3] Armstrong, M. and J. Wright. “Mobile Call Termination.” *Economic Journal*, Vol. 119(538), 2009, F270-F307.

[4] Bourreau, M., F. Kourandi, and T. Valletti, “Net Neutrality with Competing Internet Platforms.” mimeo, 2012.

[5] Casadesus-Masanell, R. and G. Llanes, “Investment Incentives in Proprietary and Open-Source Two-Sidede Platforms.” HBS Working Paper 12-114, 2012.

[6] Choi, J.P. and B-C. Kim. “Net Neutrality and Investment Incentives.” *Rand Journal of Economics*, Vol. 41(3), 2010, pp. 446-471.

[7] Dessein, W. “Network Competition in Nonlinear Pricing.” *Rand Journal of Economics*, Vol. 34(4), 2003, pp. 593-611.

[8] Economides, N. and B. Hermalin. “The Economics of Network Neutrality.” *Rand Journal of Economics*, forthcoming.

[9] Economides, N. and J. Tåg, “Net Neutrality on the Internet: A Two-sided Market Analysis” *Information Economics and Policy*, Vol. 24, 2012, pp.91-104.

[10] Hagiu, A. and R.S. Lee, “Exclusivity and Control,” *Journal of Economics and Management Strategy*, 2011, pp. 679-701.

[11] Hermalin, B. and M. Katz, “The Economics of Product-Line Restrictions With an Application to the Network Neutrality Debate.” *Information Economics and Policy*, Vol.19(2), 2007, pp.215-248.

[12] Hurkens, S. and D.S. Jeon. “Promoting Network Competition by Regulating Termination Charges.” *International Journal of Industrial Organization*, Vol.30, 2012, pp. 541-552.

[13] Jeon, D.S., J.J. Laffont and J. Tirole. “On the ‘Receiver-Pays’ Principle.” *Rand Journal of Economics*, Vol.35(1), 2004, pp. 85-110.

[14] Jullien, B. and W. Sand-Zantman “Congestion Pricing and Net Neutrality.” TSE Working paper series 12-327, 2012.

[15] Laffont, J.J., S. Marcus, P. Rey and J. Tirole. “Internet Interconnection and the Off-Net-Cost Pricing Principle.” *Rand Journal of Economics*, Vol. 34, 2003, pp. 370-390.

[16] Laffont, J.J, S. Marcus, P. Rey and J. Tirole, “Internet Peering.” *American Economic Review*, Vol. 91(2), 2001, pp. 287-291.

[17] Laffont, J.J., P. Rey and J. Tirole. "Network Competition I: Overview and Nondiscriminatory Pricing." *Rand Journal of Economics*, Vol. 29(1), 1998a, pp. 1-37.

[18] Laffont, J.J., P. Rey and J. Tirole. "Network Competition II: Price Discrimination." *Rand Journal of Economics*, Vol. 29(1), 1998b, pp. 38-56.

[19] Lee, R.S. and Tim Wu. "Subsidizing Creativity through Network Design: Zero-Pricing and Net Neutrality. " *Journal of Economics Perspective*, Vol.23(3), 2009, pp. 61-76.

[20] Njoroge, P., A.E. Ozdaglar, N.E. Stier-Moses, , G.Y. Weintraub. "Investment in Two Sided Markets and the Net Neutrality Debate." Columbia Business School DRO (Decision, Risk and Operations) Working Paper No. 2010-05.

[21] Schuett, F. "Network Neutrality: A Survey of the Economic Literature." *Review of Network Economics*: Vol. 9(2), 2010, Article 1.

Appendix

Proof of Lemma 1

Let n_i be the number of consumers subscribing to ISP i in stage 3. First, under a neutral network, consider an equilibrium price for CPs $p(q)$ given that the ISPs previously agreed on (q, a). Suppose first that at $p(q)$ both types buy connections from ISP i. Let σ_i be the ISP i's market share in the content side of the market. Then, in stage 3, each ISP maximizes the following problem when they compete on the consumer side of the market.

$$\begin{aligned} & \max_{f_i}\ \Pi_i = n_i\sigma_i(p(q) - cq) + n_j\sigma_i(p(q) - \widehat{c}q) + n_i\sigma_j(\widehat{c} - c)q + n_i f_i \\ = \ & (n_i + n_j)\sigma_i(p(q) - \widehat{c}q) + n_i(\widehat{c} - c)q + n_i f_i \end{aligned}$$

Let $p(q)$ be in equilibrium p^* and $p_i = p_i(q)$. ISP i's profit from stage 2 on can be represented as

$$\Pi_i = \pi_i(p_i, f_i, f_j; p_j = p^*).$$

From the first order condition, we have

$$\frac{d\Pi_i}{dp_i} = \frac{\partial \Pi_i}{\partial p_i} + \frac{\partial \Pi_i}{\partial f_j}\frac{\partial f_j}{\partial p_i} + \frac{\partial \Pi_i}{\partial f_i}\frac{\partial f_i}{\partial p_i} = 0.$$

From the envelope theorem, this becomes

$$\frac{d\Pi_i}{dp_i} = \frac{\partial \Pi_i}{\partial p_i} + \frac{\partial \Pi_i}{\partial f_j}\frac{\partial f_j}{\partial p_i} = 0,$$

Now let us consider an approximation of the frictionless CP market: CPs are uniformly distributed over the Hotelling line and the two ISPs are on the extreme points of the line and we study the first order condition $\frac{d\Pi_i}{dp_i} = 0$ when we make the transportation cost of the Hotelling line denoted by τ converge to zero. Then, we have

$$\sigma_i = \frac{1}{2} - \frac{p_i - p^*}{2\tau}.$$

Hence, ISP i's profit is given by

$$\Pi_i = (n_i + n_j)(\frac{1}{2} - \frac{p_i - p^*}{2\tau})(p_i - \widehat{c}q) + n_i(\widehat{c} - c)q + n_i f_i.$$

Then, we can derive from the first-order condition with respect to p_i

$$\left.\frac{\partial \Pi_i}{\partial p_i}\right|_{p_i = p^*} = (n_i + n_j)\left[\frac{1}{2} - \frac{p^* - \widehat{c}q}{2\tau}\right]$$

and from the two first order conditions with respect to f_i and f_j

$$\left.\frac{\partial f_j}{\partial p_i}\right|_{p_i = p^*} = -\frac{A + \lambda\left[\frac{p^* - \widehat{c}q}{2\tau}\right]}{B}$$

where

$$A = \frac{\frac{\lambda}{4t}}{\frac{1}{2t} + 2\lambda}, B = \frac{3}{2t} + 2\lambda$$

Hence,

$$\left.\frac{d\Pi_i}{dp_i}\right|_{p_i=p^*} = (n_i + n_j)\left[\frac{1}{2} - \frac{p^* - \widehat{c}q}{2\tau}\right] - \frac{A + \lambda\left[\frac{p^* - \widehat{c}q}{2\tau}\right]}{B}\left\{-\frac{\lambda}{2}(p^* - \widehat{c}q) + \frac{1}{2t}\left[(\widehat{c} - c)q + f_i\right]\right\} = 0$$

Multiplying the equation by 2τ gives

$$(n_i + n_j)\left[\tau - (p^* - \widehat{c}q)\right] - \frac{2\tau A + \lambda\left[p^* - \widehat{c}q\right]}{B}\left\{-\frac{\lambda}{2}(p^* - \widehat{c}q) + \frac{1}{2t}\left[(\widehat{c} - c)q + f_i\right]\right\} = 0.$$

As τ converges to zero, satisfying the equation requires the off-net cost pricing $p^* = \widehat{c}q$.

Proof of Proposition 4 (ii)

We show that there is no profitable deviation from the off-net cost pricing when the monopolistic benchmark solution $(q_H^*, q_L^*(\alpha))$ and the associated access charges (a_H^*, a_L^*) are agreed upon. Let $(\overline{q}, \underline{q})$ represent the qualities allocated to high and low types, respectively, in any deviation. Note that ISP i is indifferent between winning CPs of a given type and losing them. Therefore, we need to consider only two deviation possibilities: ISP i can deviate to induce both types to buy $q_L^*(\alpha)$ or to buy q_H^*.

Consider first the deviation of ISP i to induce both types to consume the quality $q_L^*(\alpha)$ intended for the low type CPs in the proposed equilibrium, i.e., $(\overline{q}, \underline{q}) = (q_L^*(\alpha), q_L^*(\alpha))$. Since the IC constraint for the high type is binding and high type CPs are indifferent between the two qualities in the monopolistic solution, the best way for ISP i to achieve this deviation to set a price at an epsilon discount of the off-net cost pricing for $q_L^*(\alpha)$: the price it charges after the deviation is essentially $p_i(q_L^*(\alpha)) = \alpha\theta_L u(q_L^*(\alpha))$. The CP side profit (per consumer) from this deviation is given by $\pi_{dev}^{CP} = p_i(q_L^*(\alpha)) - cq_L^*(\alpha) = \alpha\theta_L u(q_L^*(\alpha)) - cq_L^*(\alpha)$. Note that a_L^* the per unit access charge for the low quality delivery two ISPs agreed on in stage 1 is given by $a_L^* = \alpha\theta_L u(q_L^*(\alpha))/q_L^*(\alpha) - c_O$, with the off-net cost pricing of $p_i(q_L^*(\alpha)) = \alpha\theta_L u(q_L^*(\alpha)) = (a_L^* + c - c_T)q_L^*(\alpha)$. This implies that the stage 3 competition after the deviation leads to a symmetric equilibrium in which $(\overline{q}, \underline{q}) = (q_L^*(\alpha), q_L^*(\alpha))$ and $\pi_i^{CP} = \pi_j^{CP} = \pi^{CP} = (a_L^* - c_T)q_L^*(\alpha)$. This cannot give a higher profit than the upper bound that each ISP can obtain without deviation; otherwise, we have a contradiction because the upper bound is not achieved by the ISPs in the first place.

Consider now the deviation of ISP i to induce both types to consume high quality, i.e., $(\overline{q}, \underline{q}) = (q_H^*, q_H^*)$. This requires ISP i to charge $p_i(q_H^*) = \alpha\theta_L u(q_H^*)$ to induce low type CPs to purchase high quality delivery. Let (N, s_i, s_j) represent the total number of consumers subscribed and each ISP's consumer market share in stage 3. Then, ISP i's profit from the CP side is

$$\begin{aligned} & N\left[\alpha\theta_L u(q_H^*) - s_i c q_H^* - (1 - s_i)(c + a_H^* - c_T)q_H^*\right] \\ = \; & N\left[\alpha\theta_L u(q_H^*) - (c + a_H^* - c_T)q_H^* + s_i(a_H^* - c_T)q_H^*\right] \\ = \; & N\left[-\alpha\Delta\theta(u(q_H^*) - u(q_L^*(\alpha))) + s_i(a_H^* - c_T)q_H^*\right], \end{aligned}$$

where we use
$$a_H^* = \alpha \left[\theta_H u(q_H^*) - \Delta\theta u(q_L^*(\alpha))\right] / q_H^* - c_O.$$
and ISP j's profit from the CP side is
$$Ns_j(a_H^* - c_T)q_H^*.$$

Our proof strategy is to show a general result (Lemma 3) that when an ISP attracts all CPs with the same quality of delivery, the ISPs total profit (from the CP and the consumer side) is decreasing with the access charge associated with that quality, then return to the above specific set-up $(\overline{q}, \underline{q}) = (q_H^*, q_H^*)$. Before proving this, let us describe the setting under which Lemma 3 is obtained.

Specifically, fix $(\overline{q}, \underline{q}) = (q, q)$ and suppose that initially $(\overline{q}, \underline{q}) = (q, q)$ is implemented with the off-net cost pricing such that it generates $\pi_i^{CP} = \pi_j^{CP} = \pi^{CP}$: Then, we get
$$\begin{aligned} (a - c_T)q &= \pi^{CP}; \\ p(q) &= (c + a - c_T)q = \alpha\theta_L u(q). \end{aligned}$$

Consider now an asymmetric situation with a new access charge $a' = a + \Delta a$ with $\Delta a > 0$ in which ISP i is assumed to win all CPs with the same retail price
$$p(q) = \alpha\theta_L u(q).$$

Then, ISP i's profit from the CP side is
$$N\left[\alpha\theta_L u(q) - s_i cq - (1 - s_i)(c + a' - c_T)q\right]$$
whereas ISP j's profit from the CP side is
$$Ns_j(a' - c_T)q = Ns_j\left[(a - c_T)q + \Delta aq\right].$$

Note that
$$(c + a' - \Delta a - c_T)q = \alpha\theta_L u(q).$$

Hence, ISP i's profit from the CP side is
$$N\left[-\Delta aq + s_i(a' - c_T)q\right] = N\left[-(1 - s_i)\Delta aq + s_i(a - c_T)q\right].$$

Note that a' will affect (N, s_i, s_j), which is determined in stage 3. Given this, here is the lemma:

Lemma 3 *ISP i's total profit (from the content side and the consumer side) is higher when $\Delta a = 0$ than when $\Delta a > 0$.*

Proof. Let $\Pi_i(\Delta a)$ denote the total profit for ISP i when the access charge is given by $a' = a + \Delta a$. Then, we have
$$\Pi_i(\Delta a) = n_i(f_i, f_j)(f_i + \pi^{CP}) - n_j(f_i, f_j)\Delta aq$$

By using the envelope theorem, we have

$$\frac{d\Pi_i(\Delta a)}{d\Delta a} = -n_j + \left[\frac{\partial n_i}{\partial f_j}(f_i + \pi^{CP}) - \frac{\partial n_j}{\partial f_j}\Delta a q\right]\frac{df_j}{d\Delta a}$$

The expression in the square bracket is positive since $\frac{\partial n_i}{\partial f_j} > 0$ and $\frac{\partial n_j}{\partial f_j} < 0$. By totally differentiating the first order conditions for f_i and f_j, we can easily derive a comparative static result that $\frac{df_j}{d\Delta a} < 0$. The intuition is that an increase in access charge is equivalent to a subsidy by ISP i that captures the whole CP market to ISP j for each consumer ISP j attracts. ISP i competes more aggressively to attract consumers to reduce the subsidy and ISP j also competes more aggressively to attract consumers to increase the subsidy. As a result, competition in the consumer market is intensified. Taken together, we have $\frac{d\Pi_i(\Delta a)}{d\Delta a} < 0$, which shows that the winning ISP's overall profit decreases with the access charge. ■

As we planned, now use the above lemma to show that a deviation to induce the quality choice of q_H^* by both types of CPs is not profitable. To see this, consider $(\overline{q}, \underline{q}) = (q_H^*, q_H^*)$, $\Delta a q_H^* = \alpha\Delta\theta(u(q_H^*) - u(q_L^*(\alpha)))$ and $a' = a_H^*$. Hence, we have

$$\begin{aligned}
(a - c_T)\, q_H^* &= (a_H^* - c_T - \Delta a)\, q_H^* \\
&= \alpha\left[\theta_H u(q_H^*) - \Delta\theta u(q_L^*(\alpha))\right] - c q_H^* - \alpha\Delta\theta\left[u(q_H^*) - u(q_L^*(\alpha))\right] \\
&= \alpha\theta_L u(q_H^*) - c q_H^*.
\end{aligned}$$

With the access charge a given by $(a - c_T)\, q_H^* = \alpha\theta_L u(q_H^*) - c q_H^*$, the off-net cost pricing leads to $(c + a - c_T)\, q_H^* = \alpha\theta_L u(q_H^*)$. Then, from Lemma 3, we proved that the total profit of ISP i upon deviation is smaller than the profit it obtains in a symmetric equilibrium with $(\overline{q}, \underline{q}) = (q_H^*, q_H^*)$ and a satisfying $(c + a - c_T)\, q_H^* = \alpha\theta_L u(q_H^*)$. Furthermore, the profit in this symmetric equilibrium is what the ISPs could achieve through the off-net cost pricing and should give each ISP a profit smaller than the upper bound. This ends the proof.

In sum, therefore, there is no profitable deviation from the upper bound of the joint profits characterized in Proposition 4.

Proof of Proposition 5 (ii)

Here we show that the upper bound of the joint profits in the neutral network can be achieved when any ISP is allowed to deviate from the off-net cost pricing.

First, suppose that no type is excluded in the upper bound $(q_H, q_L) = (\widetilde{q}, \widetilde{q})$. Then, it is clear that there is no profitable deviation because increasing price for CPs by ISP i attracts no CPs and hence does not affect π_i^{CP} and π_j^{CP}, and decreasing the price only reduces π_i^{CP}.

Second, consider the case in which low type is excluded $(q_H, q_L) = (q_H^{FB}, 0)$. More precisely, suppose that the two ISPs agreed on providing quality q_H^{FB} at access charge $a^* = \alpha\theta_H u(q_H^{FB})/q_H^{FB} - c_O$. Then, off-net cost pricing leads to $p^*(q_H^{FB}) = \alpha\theta_H u(q_H^{FB})$ and each ISP i realizes a profit of $\nu s_i(a^* - c_T) q_H^{FB}$.

The previous argument can be applied to show that there is no profitable deviation conditional on that only high type is served. Hence, it is enough to consider ISP i's deviation to serve both types such that $(q_H, q_L) = (q_H^{FB}, q_H^{FB})$; then it will choose $p_i(q_H^{FB}) = \alpha\theta_L u(q_H^{FB})$ and obtain a profit of

$$\begin{aligned} & N\left[\alpha\theta_L u(q_H^{FB}) - s_i c q_H^{FB} - (1-s_i)\left(c + a^* - c_T\right) q_H^{FB}\right] \\ = \; & N\left[-\alpha\Delta\theta u(q_H^{FB}) + s_i\left(a^* - c_T\right) q_H^{FB}\right] \end{aligned}$$

ISP j's profit is

$$N\left[s_j\left(a^* - c_T\right) q_H^{FB}\right].$$

Hence, we can apply the previous lemma 3. Consider $(\overline{q}, \underline{q}) = (q_H^{FB}, q_H^{FB})$, $\Delta a q_H^{FB} = \alpha\Delta\theta u(q_H^{FB})$ and $a' = a^*$. Hence, we have

$$\begin{aligned} \left(a - c_T\right) q_H^* &= \left(a^* - c_T - \Delta a\right) q_H^{FB} \\ &= \alpha\theta_H u(q_H^{FB}) - c_O q_H^{FB} - \alpha\Delta\theta u(q_H^{FB}) \\ &= \alpha\theta_L u(q_H^{FB}) - c_O q_H^{FB}. \end{aligned}$$

With the access charge a given above, the off-net cost pricing leads to

$$\left(c + a - c_T\right) q_H^{FB} = \alpha\theta_L u(q_H^{FB}).$$

From Lemma 3, the total profit of ISP i upon deviation is smaller than the profit it obtains in a symmetric equilibrium with $(\overline{q}, \underline{q}) = (q_H^{FB}, q_H^{FB})$ and a that satisfies $\left(c + a - c_T\right) q_H^{FB} = \alpha\theta_L u(q_H^{FB})$. Furthermore, the profit in this symmetric equilibrium is what the ISPs could achieve through off-net cost pricing and should give each ISP a profit smaller than the upper bound.

Hence, the upper bound of the joint profits characterized in Proposition 5 can be always achieved by neutral networks.