Information Economics and Policy 62 (2023) 101019

Contents lists available at ScienceDirect

Information Economics and Policy

journal homepage: www.elsevier.com/locate/iep

Prioritization between asymmetric content providers☆

Jeong-Yoo Kim

Department of Economics, Kyung Hee University, 1 Hoegidong, Dongdaemunku, Seoul, 137-701, Republic of Korea

ARTICLE INFO

Article history:
Received 20 April 2022
Revised 5 January 2023
Accepted 11 January 2023
Available online 13 January 2023

JEL classification:
L11

Keywords:
Net neutrality
Congestion
Delay
Prioritization

ABSTRACT

Gautier and Somogyi (2020) showed that the monopolistic Internet service provider (ISP) can extract more surplus from consumers by giving priority to the weaker content to restore symmetry between content providers (CPs). In this study, we reexamine the issue and argue that their result depends critically on the shape of the delay cost function. We first show that under a linear delay cost, if the delay cost of contents from each CP increases with its own traffic amount, the opposite is true, that is, the ISP prefers to give priority to a strong CP, whereas it prefers to give priority to a weak CP if the delay cost of contents from an unprioritized CP decreases with its traffic amount. We confirm our insight in two specific models; the M/M/1 queuing model and the bandwidth subdivision model. We also discuss some implications of the ISP's prioritization choice for social welfare.

1. Introduction

The concept of net neutrality had a long history. In the era of the Roman Empire, all monopoly hostels, ports, and even surgeons in a community had an obligation to provide their services at a reasonable price. This is known as the principle of common carriage (common carrier) which is regarded as an essential element of net neutrality. This principle was adopted in the U.S. Communications Act of 1934 and inherited in the U.S. Telecommunications Act of 1996, although whether Internet service providers (ISPs) should be classified as Title I carrier with no common carrier obligation or Title II carrier with a common carrier obligation has been controversial. Formal net neutrality rules were established by Open Internet Order 2010[1] and propelled by Open Internet Order 2015 when the U.S. Federal Communications Commission (FCC) classified ISPs as Title II carrier (common carriers).

However, in 2017, the FCC repealed the net neutrality rules that barred ISPs from blocking or slowing Internet content or from offering paid "fast lanes." It is still controversial whether ISPs should be forced to be neutral, that is, treat all traffic the same or should be allowed to prioritize certain services over others. Nevertheless, it seems clear that certain services such as YouTube or Netflix require more bandwidth than others, that is, it seems sensible to prioritize such traffic because they would be almost unusable if streaming were subject to irregular delays.

Under the current regime, without net neutrality, it is important to decide for which priority should be granted. In a recent study, Gautier and Somogyi (2020) compared the outcomes with and without net neutrality when there is asymmetry between content providers (CPs). They showed that an ISP can extract more surplus from consumers by privileging the relatively weaker content to restore symmetry between CPs.[2] In particular, with prioritization,[3] they showed that it is preferable for an ISP to give priority to weak CPs because the resulting price is higher than the price when it gives priority to strong CPs. They used a specific function form to model the effect of delay on the utility of end users.

In this study, we argue that, in general, the priority of ISP grants crucially depends on the shape of the delay cost function. We show that if the delay cost of contents from each CP increases with its own traffic amount, the opposite is true, that is, the ISP prefers to

☆ This research was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2022S1A5A2A0304932311). We are grateful to the Editor Marc Bourreau and two anonymous referees for their extremely helpful comments.

E-mail address: jyookim@khu.ac.kr

[1] Roughly speaking, the net neutrality rules include no blocking, no throttling, and no prioritization.

[2] This issue can be regarded as a special one of intermediate market price discrimination, that is, to which downstream producer a monopolistic input supplier should give a favor. For input market price discrimination, DeGraba (1990) showed that the input supplier gives a favor to a less efficient producer in the case of linear demand, and Li (2014) argued that whether the monopolist gives a favor to the more efficient downstream firm or a less efficient downstream firm depends on the shape of the demand function.

[3] Prioritization can be either paid or unpaid. Gautier and Somogyi (2020) assume that prioritization is unpaid. We also follow this assumption in this paper.

https://doi.org/10.1016/j.infoecopol.2023.101019
0167-6245/

give priority to the strong CP, whereas it prefers to give priority to the weak CP if the delay cost of contents from the unprioritized CP decreases with its traffic amount as in Gautier and Somogyi (2020).[4]

The intuition behind this is as follows. If the ISP grants priority to the weak CP, there are two effects on the marginal end user who uses contents from a strong CP, namely, a positive transportation cost effect and a negative delay cost effect, because the traveling distance of the marginal end user is shorter but the delay cost is higher because of the absence of priority. However, the disadvantage in the delay cost (of content from the unprioritized CP) outweighs the advantage in the transportation cost for any degree of asymmetry if the delay cost of contents from the unprioritized strong CP (with higher demands) increases with the traffic amount. Consequently, the ISP will find it better to grant priority to the strong content. On the other hand, if the delay cost decreases with the traffic amount, the converse is true; the disadvantage in the delay cost is outweighed by the advantage in the transportation cost, meaning that it is better for the ISP to grant priority to weak content. We then examine our insight in two specific models; the M/M/1 queuing model of Choi and Kim (2010) and Cheng et al. (2011), and the bandwidth subdivision model of Economides and Hermalin (2012). It turns out that the M/M/1 model corresponds to the latter case, while the bandwidth subdivision model corresponds to the former. we We obtain the contrasting results that the ISP prefers to grant priority to the weak CP for any level of asymmetry in the queuing model, whereas it prefers to grant priority to the strong CP for any level of asymmetry in the bandwidth subdivision model.

We also obtain some policy implication by comparing social welfare when priority is granted to the weak CP and to the strong CP. We first show under a linear delay cost that if the delay cost of contents from each CP increases with its own traffic amount, it is socially optimal as well as privately optimal to grant priority to the strong CP, although the welfare comparison is ambiguous if the delay cost of contents from the unprioritized CP decreases with its traffic amount. We then show that in the queuing model, there is a conflict between the private incentive and the social incentive in terms of which CP is given priority. An ISP may grant priority to the weak CP in equilibrium, even though this is actually welfare inferior to the outcome when it grants priority to the strong CP. On the other hand, in the bandwidth subdivision model, it is privately and socially better for the ISP to prioritize the strong CP. However, in both models, it is socially worse than the outcome under net neutrality, regardless of who is given priority. This implies that regulation of net neutrality is needed in terms of static social welfare.[5]

The remainder of this paper is organized as follows. In Section 2, we provide an analysis of the model with a linear delay cost to examine the ISP's incentive to prioritize. In Section 3, we consider two specific models, the $M/M/1$ queuing model and the subbandwidth model to address the issue. Section 4 discusses social welfare. Finally, concluding remarks and caveats are presented in Section 5.

[4] Gautier and Somogyi (2020) make the assumption that the quality of contents from the unprioritized CP increases with its market share, which is essentially equivalent to the assumption that the delay cost of the traffic from the unprioritized CP decreases with its market share.

[5] In an earlier version, we examined the ISP's dynamic incentive to invest in network capacity along the line of Choi and Kim (2010), Krämer and Wiewiorra (2012), Bourreau et al. (2015), and Choi et al. (2018), and showed that the ISP prefers prioritize the strong CP in the queuing model and the weak CP in the subbandwidth model in terms of dynamic incentives to invest, contrary to the static preference of the ISP who prefers to prioritize the weak CP in the queuing model and the strong CP in the subbandwidth model.

2. Model

End users (consumers) are uniformly distributed over $[0,1]$. So, the mass of end users is normalized to one. CP 1 and CP 2 are located in $x_1 = a$ and $x_2 = 1$ respectively where $a \in (0,1)$ à la Gautier and Somogyi (2020). Each end user located at x requests one unit of contents from at most one content provider by traveling to x_1 or x_2 and gets some valuation v, implying that the sum of consumers who receive contents from CP 1 and CP 2 is one in which case the market is fully covered, or less than one in which case the market is partially covered. The unit transportation cost is t, so the traveling cost of the end user located at x is $t|x_i - x|$. For the time being, we restrict our attention to the case in which the market is fully covered.

End users get disutility from delay as well as utility when they receive contents. Without net neutrality, delay costs depend on which contents are prioritized, what proportion is prioritized, etc. because delay from congestion will be increasing in the amounts of traffic from each CP. We denote the delay cost of consumers who receive contents from CP i with the market share z_i by $c_i(z_i)$.[6] Due to the possibility of congestion by limited network capacity denoted by $\kappa(<1)$, we assume that $c_i'(z_i) > 0$ in case that $z_1 + z_2 = 1 > \kappa$. Under net neutrality, we assume that $c_i(.) = c(.)$ for $i = 1, 2$. We also assume that $c_i(z_i) < c_j(z_j)$ (where $j \neq i$) for any $z_i, z_j \in (0,1)$ such that $z_1 + z_2 = 1$, when a priority is given to CP i.[7]

We consider the following sequence of events. First, the ISP decides to which CP it will give priority on an unpaid basis and then sets the connection price for end users. Then, after forming expectations about the market share of each CP,[8] end users decide from which CP they request contents.[9]

Let $V_i(x)$ be the net valuation of a consumer located at x who buys contents from CP i. Then, we have

$$V_1(x) = v - c_1(z_1^e) - p - t|x - a|, \tag{1}$$

$$V_2(x) = v - c_2(z_2^e) - p - t(1 - x), \tag{2}$$

where z_i^e is the expectation about the market share of CP i and p is the common price that the monopolistic ISP charges to all consumers.

We first consider the case in which the market is fully covered. The two content providers are not symmetric in the sense that CP 1 has a higher demand than CP 2 at the common price, because $a > 0$. So, we call CP 1 and CP 2 a strong CP and a weak CP respectively à la Gautier and Somogyi (2020).

To analyze the sequential game by backward induction, we first examine the consumers' purchasing decisions. If the ISP grants priority to CP 1 (strong CP), we denote the marginal consumer, if any, who is indifferent between buying from CP 1 and CP 2 by x_s which is determined by

$$c_1(z_1^e) + t(x_s - a) = c_2(z_2^e) + t(1 - x_s). \tag{3}$$

Throughout the paper, we will assume that users' expectations are actually fulfilled. The assumption of fulfilled expectations implies

[6] Of course, the average time to transmit contents from CP i depends on z_j as well as z_i, but since $z_j = 1 - z_i$ in the case of full market coverage, $c_i(z_i) = \tilde{c}_i(z_i, 1 - z_i)$ can be interpreted as a reduced form.

[7] These assumptions on $c_i(z_i)$ hold in the M/M/1 queuing model of Choi and Kim (2010) and Cheng et al. (2011) which will be reexamined in Section 3.

[8] The assumption on the sequence in the subgame that follows the ISP's decision is not crucial. The equilibrium outcome remains unaffected even if we assume that users' expectations and decisions are made simultaneously insofar as they form correct expectations.

[9] Note that the ISP determines whom to give priority. Since we assume unpaid prioritization just as in Gautier and Somogyi (2020), CPs have no means to compete (such as a price bid for a priority) to obtain priority.

that $z_1^e = z_1 = x_s$ and $z_2^e = z_2 = 1 - x_s$. So, we obtain

$$x_s = \frac{1+a}{2} + \frac{1}{2t}[c_2(1-x_s) - c_1(x_s)]. \tag{4}$$

On the other hand, if the ISP grants priority to CP 2 (weak CP), x_w which is similarly defined as the marginal consumer indifferent between buying from CP 1 and CP 2 is determined by

$$c_1(z_1) + t(x_w - a) = c_2(z_2) + t(1 - x_w), \tag{5}$$

where $z_1 = x_w$ and $z_2 = 1 - x_w$. Accordingly, we obtain

$$x_w = \frac{1+a}{2} + \frac{1}{2t}[c_2(1-x_w) - c_1(x_w)]. \tag{6}$$

From now on, we will denote the cost functions from the prioritized CP and from the un-prioritized CP by $c_p(.)$ and $c_u(.)$ respectively. So, if CP 1 is given priority, $c_p(.) = c_1(.)$ and $c_u(.) = c_2(.)$, while $c_p(.) = c_2(.)$ and $c_u(.) = c_1(.)$ if CP 2 is given priority. In any case, by assumption, we have $c_p(z) < c_u(1-z)$. By using this notation, we can rewrite (4) and (6) that determine the interior solutions[10] $x_i \in (a, 1)$ for $i = s, w$ as

$$x_s = \frac{1+a}{2} + \frac{1}{2t}[c_u(1-x_s) - c_p(x_s)], \tag{7}$$

$$x_w = \frac{1+a}{2} + \frac{1}{2t}[c_p(1-x_w) - c_u(x_w)]. \tag{8}$$

Note that the marginal consumer is $x^* = \frac{1+a}{2}$ under net neutrality, because $c_1(x) = c_2(x)$ for all x. Equations (7) and (8) imply that $x_w < x^* < x_s$ because $c_u(1-x) > c_p(x)$ for any x under prioritization.

Proposition 1. *(i) $x_w(a) < x^*(a) < x_s(a)$ for any $a \in (0, 1)$. (ii) $x_s(a)$ and $x_w(a)$ increase in a.*

It is clear that if priority is granted to CP 1, its market share is increased ($x_s > x^*$). It is also clear that if priority is granted to the other CP, the market share of CP 1 is reduced, because its increased delay cost gives CP 1 relative disadvantage. The intuition for the second result is also clear. As a is larger, i.e., the demand for CP 1 is larger, the market share of CP 1 will be larger, regardless of who is granted priority.

Now, we consider the pricing decision of the monopoly ISP. Let p_s (or p_w respectively) be the price of the ISP when it grants priority to strong CP (or weak CP respectively). The conditions for full market coverage (i.e., $z_1 + z_2 = 1$) in two cases – which will be called individual rationality (IR) conditions – are that for $i = 1, 2$,

$$\begin{aligned} V_i(x_s) &= v - c_p(x_s) - p_s - t(x_s - a) \\ &= v - c_u(1-x_s) - p_s - t(1-x_s) \geq 0, \end{aligned} \tag{9}$$

$$\begin{aligned} V_i(x_w) &= v - c_u(x_w) - p_w - t(x_w - a) \\ &= v - c_p(1-x_w) - p_w - t(1-x_w) \geq 0. \end{aligned} \tag{10}$$

From (9) and (10), the profit-maximizing ISP will choose a price

$$p_s^* = v - c_p(x_s) - t(x_s - a) = v - c_u(1-x_s) - t(1-x_s), \tag{11}$$

if a strong CP is given priority and

$$p_w^* = v - c_p(1-x_w) - t(1-x_w) = v - c_u(x_w) - t(x_w - a) \tag{12}$$

if a weak CP is given priority. These prices make two (IR) conditions binding and maximize the profit of the ISP, given that the market is fully covered.[11] The equilibrium prices p_s^* and p_w^* depend on a, and in fact, both of them are increased with respect to an increase in a.

Proposition 2. *$p_s^*(a)$ and $p_w^*(a)$ increase in a.*

The intuition behind this proposition goes as follows. As a is larger, the transportation cost of the marginal consumer when he buys from CP 2 is lower because $x_s^*(a)$ and $x_w(a)$ increase in a, meaning that his net valuation of receiving contents (aside from the price p) is higher. This implies that the ISP can raise the price to extract all the consumer surplus of the marginal consumer. Also, this argument holds whether priority is granted to a strong CP or a weak CP.

However, the degree of a change in net valuation with respect to a change in a depends on which CP is given priority, and so would the degree of a change in the equilibrium prices. The following proposition may give some insight for the question.

Proposition 3. *Let $C^s \equiv c_p(x_s) + t(x_s - a)$ and $C^w = c_u(x_w) + t(x_w - a)$. If $c_p(\cdot)$ and $c_u(\cdot)$ are linear in their arguments with $c_u'(z) > c_p'(z) > 0$ for any $z \in (0, 1)$, (i) $C^w - C^s$ is increasing in a, and (ii) the ISP prefers to grant priority to the strong CP for any $a > 0$.*

The two costs C^s and C^w are the total costs that the marginal consumer x_s and x_w bear when priority is given to the strong CP and the weak CP respectively. This proposition implies that as a gets larger, the marginal consumer's cost gets higher when priority is given to the weak CP relative to when priority is given to the strong CP. This means that the ISP is more likely to grant priority to the strong CP as a is larger, because it can charge a higher price.[12] Proposition 3 says that it turns out that the ISP finds it better to grant priority to the strong CP for any $a > 0$ if the delay costs are linear and $c_u'(z) > c_p'(z) > 0$.

To see this, Eqs. (11) and (12), which can be rewritten as $p_s^*(a) = v - C^s(a)$ and $p_w^*(a) = v - C^w(a)$, imply that $p_s^*(a) > p_w^*(a)$ if and only if $C^w(a) > C^s(a)$. Since $C^w(0) = C^s(0)$, Proposition 3(ii) directly follows from Proposition 3(i). Contrary to Gautier and Somogyi (2020), the monopolist will find it to its advantage to grant priority to a strong CP rather than a weak CP for any $a > 0$.

To elaborate, let $c_p(x) = \alpha x + \beta$ and $c_u(y) = \alpha' y + \beta'$ where x and y are traffic volumes of the prioritized contents and the unprioritized contents respectively, $\alpha' > \alpha > 0$, and $\beta' \geq \beta \geq 0$. The assumption that $c_p(z) < c_u(1-z)$ for any $z \in (0, 1)$ can be satisfied if $\beta' - \beta \geq \alpha$. So, to be consistent with this assumption, we only consider the case that $\beta' - \beta \geq \alpha$. We have

$$\begin{aligned} C^w - C^s &= c_u(x_w) + t(x_w - a) - [c_p(x_s) + t(x_s - a)] \\ &= c_u(x_w) - c_p(x_s) + t(x_w - x_s) \\ &= (\alpha' x_w - \alpha x_s + \beta' - \beta) + t(x_w - x_s). \end{aligned} \tag{13}$$

The first term of (13) is a disadvantage in the delay cost of the marginal consumer using CP1 when priority is given to the weak CP, while the second term of (13) is an advantage in the traveling cost ($x_w - x_s < 0$). As a gets larger, we have

$$\frac{\partial (C^w - C^s)}{\partial a} = \left(\alpha' \frac{\partial x_w}{\partial a} - \alpha \frac{\partial x_s}{\partial a}\right) + t\left(\frac{\partial x_w}{\partial a} - \frac{\partial x_s}{\partial a}\right). \tag{14}$$

[10] For linear functions $c_p(x) = \alpha x + \beta$ and $c_u(y) = \alpha' y + \beta'$, a sufficient condition guaranteeing interior solutions, $x_s, x_w \in (a, 1)$, is that $\beta' - \beta = \alpha$, $a < 1$ and $a < \frac{t}{t+\alpha'-\alpha}$. For detailed derivations, see the proof of Proposition 3 in Appendix A.

[11] As an anonymous referee correctly pointed out, without the assumption of full market coverage, the ISP may find it profitable to charge a price that makes the market only partially covered, if the users' demand for Internet access is inelastic. The possibility of partial market coverage will be discussed in the end of this section.

[12] If a is too large, the market may not be fully covered because the end user located at $x = 0$ would not want to purchase from neither CP. Assumption 1 of Gautier and Somogyi excludes the possibility. We also maintain this feature by assuming that a cannot be too large.

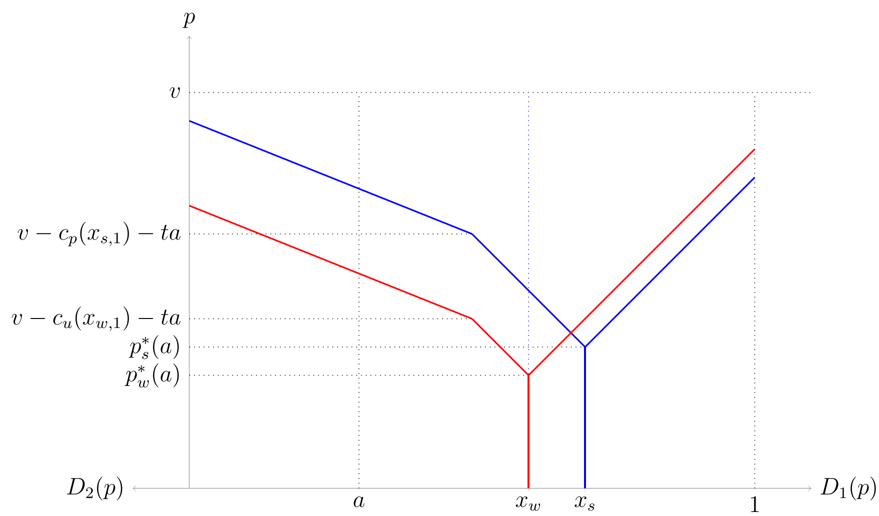

Fig. 1. Demand Functions for CP 1 and CP 2 (blue: when strong CP is prioritized) (red: when weak CP is prioritized). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Since $\frac{\partial x_w}{\partial a} = \frac{\partial x_s}{\partial a} > 0$ under the linear delay cost,[13] the sign of the first term is positive if $\alpha' > \alpha$, while the sign of the second term is zero. This means that the disadvantage in the delay cost gets larger, while the advantage in the traveling cost remains the same. This implies that the ISP finds it better to give priority to the strong CP for any $a > 0$.

Next, we consider the case in which the market is partially covered. If the ISP raises the connection price above p_s^* (or p_w^*) when it grants priority to a strong CP (or a weak CP respectively), the market is not fully covered. For $p > p_s^*$, define $x_{s,1}$ and $x_{s,2}$ by

$$V_1(x_{s,1}) = v - p - c_p(x_{s,1}) - t(x_{s,1} - a) = 0, \tag{15}$$

$$V_2(x_{s,2}) = v - p - c_p(1 - x_{s,2}) - t(1 - x_{s,2}) = 0. \tag{16}$$

Then, $x_{s,1}$ and $x_{s,2}$ determine the demand for CP 1 and CP 2 given p when a strong CP (CP1) is given a priority. Note that consumers located at $x \in (x_{s,1}, x_{s,2})$ buy contents from neither CP.

Similarly, for $p > p_w^*$, we can define $x_{w,1}$ and $x_{w,2}$ by[14]

$$V_1(x_{w,1}) = v - p - c_u(x_{w,1}) - t(x_{w,1} - a) = 0, ^{14} \tag{17}$$

$$V_2(x_{w,2}) = v - p - c_u(1 - x_{w,2}) - t(1 - x_{w,2}) = 0, \tag{18}$$

while consumers located at $x \in (x_{w,1}, x_{w,2})$ request contents from neither CP.

Figure 1 shows the demand functions for CP 1 and CP 2. Note that the market is not fully covered if $p > p_s^*$ when CP 1 is given priority and $p > p_w^*$ if CP 2 is given priority. We will find the condition for Proposition 3 to hold, i.e., a sufficient condition under which raising the price from p_s^* or p_w^* slightly will not increase the profit of the monopolistic ISP.

In the case in which the ISP slightly raises the price from p_s^* so that the market is only partially covered,[15] the total demand for Internet connection can be computed as

$$\begin{aligned} D(p) &= x_{s,1}(p) + (1 - x_{s,2}(p)) \\ &= \frac{v - p - \beta + at}{\alpha + t} + \frac{v - p - \beta'}{\alpha' + t}, \end{aligned} \tag{19}$$

when priority is given to the strong CP. The first term is the demand of consumers using CP 1 and the second term is the demand of consumers using CP 2. Then, from the profit of the ISP which is $\pi = pD(p)$, we can obtain

$$\pi'(p_s^*) = \frac{v - p_s^* - \beta + at}{\alpha + t} + \frac{v - p_s^* - \beta'}{\alpha' + t} - p_s^*\left(\frac{1}{\alpha + t} + \frac{1}{\alpha' + t}\right). \tag{20}$$

As is well known, if the ISP charges the price slightly, it has two effects; the effect of increasing revenue by the price (the first term) and the effect of reducing the demand (the second effect). From $p_s^*(a) = v - C^s(a)$ where $C^s(a)$ is found from (A.5), the condition for $\pi'(p_s^*) < 0$ is reduced to

$$v > 2\frac{(\alpha + t)[(1 + a)t + \alpha' + \beta' - \beta]}{2t + \alpha' + \alpha} + 2(\beta - at) - \frac{\frac{\beta - at}{\alpha + t} + \frac{\beta'}{\alpha' + t}}{\frac{1}{\alpha + t} + \frac{1}{\alpha' + t}}. \tag{21}$$

It is clear that (21) holds if v is large enough. Intuitively, if v is very large, p_s^* is high so the second term which is the demand-reducing effect outweighs the first term which is the revenue-increasing effect. Hence, the profit is reduced.

Similarly, if priority is given to the weak CP,

$$\begin{aligned} D(p) &= x_{w,1}(p) + (1 - x_{w,2}(p)) \\ &= \frac{v - p - \beta' + at}{\alpha' + t} + \frac{v - p - \beta}{\alpha + t}, \end{aligned} \tag{22}$$

and thus, the condition for p_w^* to be optimal is that

$$\begin{aligned} \pi'(p_w^*) = &\frac{v - p_w^* - \beta' + at}{\alpha' + t} + \frac{v - p_w^* - \beta}{\alpha + t} \\ &- p_w^*\left(\frac{1}{\alpha + t} + \frac{1}{\alpha' + t}\right) < 0, \end{aligned} \tag{23}$$

which is also satisfied if v is large enough.

Finally, it deserves to notice the crucial difference in this model and Gautier and Somogyi (2020). Gautier and Somogyi (2020) model disutility from delay as a quality degradation. Let q_i be the quality of CP i and κ is the network capacity. They

[13] This is easily shown from (A.5) and (A.6) in the proof of Proposition 3 in Appendix A.

[14] The market share for CP 1 is $x_{w,1}$ only if a is not too large, as we argued in Footnote 11.

[15] If the price is so high that $z_1 + z_2 < \kappa$, the congestion problem does not occur. To focus only on the case that the network is congested, we only consider a small deviation from p_s^* and p_w^*.

assume $q_1 = 1$ and $q_2 = \frac{\kappa - x_1}{x_2}$ when $x_1 < \kappa$ and $q_1 = \frac{\kappa}{x_1} \le 1$ and $q_2 = 0$ when $x_1 \ge \kappa$ if the ISP grants priority CP 1. If the market is covered, i.e., $x_1 + x_2 = 1$, this implies that $q_2'(x_2) > 0$ because $q_2 = \frac{\kappa-(1-x_1)}{x_2} = 1 - \frac{1-\kappa}{x_2}$ when $x_1 < \kappa$. On the other hand, we assumed that $c_1'(x_1) > 0$ and $c_2'(x_1) < 0$, or equivalently, $c_1'(x_2) < 0$ and $c_2'(x_2) > 0$ if $x_1 + x_2 = 1$. This is contrasted with the assumption of Gautier and Somogyi (2020) in the sense that the utility of CP 2's consumer is decreased as CP 2's market share increases.

If we adopt an alternative assumption that $c_2'(x_2) < 0$ as in Gautier and Somogyi (2020), i.e., we assume that $c_u(x_2) = \alpha' x_2 + \beta'$ where $\alpha' < 0 < \alpha$, the signs in (14) are reversed. Since $\alpha' < 0$, we have $\alpha' \frac{\partial x_w}{\partial a} - \alpha \frac{\partial x_s}{\partial a} < 0$ while $\frac{\partial x_w}{\partial a} - \frac{\partial x_s}{\partial a} = 0$, so that $\frac{\partial C^w}{\partial a} < \frac{\partial C^s}{\partial a}$. That is, when priority is given to the weak CP, the disadvantage in the delay cost gets smaller as a gets larger, while the advantage in the traveling cost remains the same. This implies that it is better for the ISP to give priority to the weak CP as a gets larger. This means that it is always better to give priority to the weak CP regardless of a if $\alpha' < 0$, i.e., if the delay cost of contents from the unprioritized CP decreases as its traffic amount increases (correspondingly, as the traffic from the prioritized CP decreases).

3. Specific models

In this section, we consider two specific models of prioritization depending on different assumptions on the delay cost functions.

3.1. M/M/1 queuing model

Choi and Kim (2010) and Cheng et al. (2011) use the M/M/1 queuing model to model congestion and delay in data transmission. In this section, we consider the delay cost functions that can be derived from the M/M/1 queuing model. For asymmetry between content providers, we follow the model of Gautier and Somogyi (2020).

Under net neutrality, both the interarrival time of content requests and the service time of the Internet service provider (ISP) follow exponential distributions with λ and μ. That is, the mean of the time between content requests is $1/\lambda$ and the mean of the service time is $1/\mu$. As usual, we assume that $\mu > \lambda$ to avoid the possibility that the waiting time will explode.

On the other hand, under no net neutrality allowing prioritization, the total amounts of content requests from CP 1 and CP 2 are $\tilde{\lambda}_1 = \tilde{x}\lambda$ and $\tilde{\lambda}_2 = (1-\tilde{x})\lambda$ respectively where $\tilde{x}$ is the ratio of users who request contents from CP 1, if we assume that λ is preserved after prioritization.[16] Under the preemptive priority system,[17] if CP 1 is granted priority, $\tilde{x}$ is determined by

$$c_p(\tilde{x}) + t(\tilde{x} - a) = c_u(1-\tilde{x}) + t(1-\tilde{x}), \tag{24}$$

where $c_p(\tilde{x}) = \frac{1}{\mu - \tilde{x}\lambda}$ and $c_u(1-\tilde{x}) = \frac{\mu}{\mu-\lambda}\frac{1}{\mu-\tilde{x}\lambda}$. Note that $c_u'(\tilde{y}) < 0$, where $\tilde{y} = 1 - \tilde{x}$ is the market share of the unprioritized CP. So, this delay cost corresponds to the assumption of Gautier and Somogyi (2020).

We compare two cases; the case in which the ISP grants priority to a strong CP and the case in which the ISP grants priority to a weak CP. Let x_s and x_w be $\tilde{x}$ when priority is given to a strong CP and a weak CP respectively. Then, x_s and x_w must satisfy

$$T_1^s + t(x_s - a) = T_2^s + t(1 - x_s), \tag{25}$$

$$T_1^w + t(x_w - a) = T_2^w + t(1 - x_w), \tag{26}$$

where

$$T_1^s = \frac{1}{\mu - x_s\lambda}, \tag{27}$$

$$T_2^s = \frac{\mu}{\mu - \lambda} T_1^s = \frac{\mu}{\mu-\lambda}\frac{1}{\mu - x_s\lambda}, \tag{28}$$

$$T_2^w = \frac{1}{\mu - (1-x_w)\lambda}, \tag{29}$$

$$T_1^w = \frac{\mu}{\mu-\lambda} T_2^w = \frac{\mu}{\mu-\lambda}\frac{1}{\mu-(1-x_w)\lambda}. \tag{30}$$

Here, T_i^s and T_i^w are the expected waiting time of contents from CP i $(i = 1, 2)$ when priority is given to a strong CP and a weak CP respectively. It is easy to see that $c_p(.) = T_1^s(.) = T_2^w(.)$ and $c_u(.) = T_1^w(.) = T_2^s(.)$ in this specific model.

To compare p_s^* and p_w^*, we only need to compare $C^s = T_1^s + t(x_s - a)$ and $C^w = T_1^w + t(x_w - a)$ by using the equilibrium values of x_s and x_w which can be obtained from (25) and (26).

As we saw in the previous section, granting priority to a CP entails two effects, the effect on the transportation cost and the effect on the delay cost. If the ISP grants priority to a strong CP, two conflicting effects on the marginal consumer who uses CP 1 occur. On one hand, it has a positive effect on the delay cost, and on the other hand, it has a negative effect on the transportation cost, because $x_s > x_w$ implies that the marginal consumer's traveling distance to $x = a$ is longer. However, the following proposition shows that the first effect is outweighed by the second effect in this model for any a.

Proposition 4. *$p_s^*(a) < p_w^*(a)$ for any a in the M/M/1 model.*

This proposition implies that it is always better to give priority to the weak content in the M/M/1 queuing model. To see this, compare the total cost of the marginal consumer when he uses privileged contents. If priority is given to the weak CP, the delay cost for the prioritized CP is lower because $\frac{1}{\mu-(1-x_w(a))\lambda} < \frac{1}{\mu - x_s(a)\lambda}$ due to the lower volume of traffic, although the transportation cost is higher because $t(1-x_w(a)) > t(x_s(a) - a)$ by (A.50) and Lemma 3 (in Appendix A). We can confirm a tradeoff between the delay cost and the transportation cost that was identified in Section 2.

The main driving force for Proposition 4 is that $c_u'(\tilde{y}) < 0$. Intuitively, under this feature, the utility of consumers using unprioritized contents increases as the traffic of the unprioritized contents increases as in Gautier and Somogyi (2020). We know that the traffic of unprioritized contents is $1 - x_s$ when priority is given to the strong CP, while the traffic is x_w if priority is given to the weak CP. We have $x_w(a) > 1 - x_s(a)$ for any $a > 0$. The traveling cost of the marginal consumer to the unprioritized CP is also lower when the weak CP has priority, because $x_w - a < 1 - x_s$.[18] After all, when priority is given to the weak CP, the utility of the marginal consumer

---

[16] If net neutrality is repealed, users expect the contents they request to be transmitted faster if the contents have priority, and thus they will demand more contents with priority. This implies that the request rate for prioritized contents will be generally higher than the request rate for unprioritized contents or contest under net neutrality. Thus, Kim (2022) adopts the alternative assumption that if CP 1 is granted priority and CP 2 is not, $\lambda_1 > \lambda_2$, since prioritized contents will be more frequently requested. This corresponds to the case of the variable demand in Gautier and Somogyi (2020), while the assumption that $\lambda_1 = \lambda_2 = \lambda$ corresponds to the case of the fixed demand.

[17] In the preemptive priority system, the service of a content request that is already being processed is interrupted and superseded by the prioritized content, while it is not interrupted by the arrival of the content request with priority in the non-preemptive priority system. Although Choi and Kim (2010) claim that the following waiting times are obtained under the assumption of the non-preemptive priority system, they are, in fact, the ones that can be obtained under the preemptive priority system. See Kim (2022).

[18] We need to prove this for completing the proof of Proposition 4, but the Appendix A provides an alternative proof of Proposition 4.

of requesting unprioritized contents is higher and accordingly the price that makes the (IR) condition binding (i.e., that makes the marginal consumer's net valuation zero) is higher.

3.2. Bandwidth subdivision model of economides and hermalin

Economides and Hermalin (2012) model prioritized service and delay in a different way.[19] Let k be a bandwidth that can be interpreted as the capacity of the ISP's content transmission. Also, let k_p and k_u be the sub-bandwidths allocated to contents from the prioritized CP and the unprioritized CP respectively, where $k_p + k_u = k$ and $k_p > k_u$. So, unlike the M/M/1 queuing model in which the ISP processes prioritized contents before unprioritized contents, prioritized contents and unprioritized contents are assumed to use separate portions of the bandwidth (fast lane and slow lane, respectively).

We consider the following congestion model under unpaid prioritization à la Economides and Hermalin (2012). First, given k_p and k_u,[20] the ISP decides whom to grant priority by allocating k_p and then chooses the connection fee for end users p. Second, end users choose from which CP they request contents.

We assume that the valuation v for one unit of contents that each end user requests is discounted by the average transmission time. Let the adjustment factor be $\rho(\tau_i) = \rho\tau_i$,[21] where $\rho > 0$, $\tau = \frac{x_i}{k_i}$ is the average transmission time of contents from CP i and x_i is the total amount of contents from CP i.[22] Then, the discounted valuation from the service of CP i is $\frac{v}{\rho\tau_i}$. For simplicity, we assume that $k = 1$ which is equal to the size of end users (the total amount of traffic), implying that $k_p > \frac{1}{2} > k_u$.

By abusing notation, let $x_s(a)$ and $x_w(a)$ be the location of the borderline end user who is indifferent between two CPs when the ISP grants priority to the strong CP and the weak CP respectively. Assuming that v is large enough to cover the whole market, we can find $x_s(a)$ and $x_w(a)$ by the following two Eqs. (31) and (32) respectively;

$$\frac{v}{\rho}\frac{k_p}{x} - t(x-a) = \frac{v}{\rho}\frac{k_u}{1-x} - t(1-x), \tag{31}$$

$$\frac{v}{\rho}\frac{k_u}{x} - t(x-a) = \frac{v}{\rho}\frac{k_p}{1-x} - t(1-x). \tag{32}$$

Let $v_u(y) = \frac{v}{\rho}\frac{k_u}{y}$. Then, $v_u'(y) = -\frac{v}{\rho}\frac{k_u}{y^2} < 0$. So, this is essentially equivalent to the assumption that $c_u'(y) > 0$, unlike the case of Gautier and Somogyi (2020).

Let $p_s^*(a)$ and $p_w^*(a)$ be the prices that maximize the profit of the ISP when it grants priority to the strong CP and to the weak CP respectively. Then, the conditions for full market coverage imply that

$$\frac{v}{\rho}\frac{k_p}{x_s} - p_s - t(x_s-a) = \frac{v}{\rho}\frac{k_u}{1-x_s} - p_s - t(1-x_s) = 0, \tag{33}$$

$$\frac{v}{\rho}\frac{k_u}{x_w} - p_w - t(x_w-a) = \frac{v}{\rho}\frac{k_p}{1-x_w} - p_w - t(1-x_w) = 0, \tag{34}$$

so that the profit-maximizing prices will be

$$p_s^*(a) = \frac{v}{\rho}\frac{k_p}{x_s} - t(x_s-a) = \frac{v}{\rho}\frac{k_u}{1-x_s} - t(1-x_s), \tag{35}$$

if a strong CP is given priority and

$$p_w^*(a) = \frac{v}{\rho}\frac{k_u}{x_w} - t(x_w-a) = \frac{v}{\rho}\frac{k_p}{1-x_w} - t(1-x_w) \tag{36}$$

if a weak CP is given priority.

The following proposition shows that the ISP prefers to prioritize the strong CP in this congestion model, contrary to the M/M/1 queuing model.

Proposition 5. *$p_s^*(a) > p_w^*(a)$ for any $a > 0$ and for any k_p, k_u such that $k_p > k_u$ in the congestion model of Economides and Hermalin (2012).*

It is interesting that the result obtained in the queuing model is reversed in this model. To make comparison easy, compare the total costs of the marginal consumer only when he uses unprioritized contents. The discounted valuation is higher when priority is given to the strong CP, because $\frac{k_u}{1-x_s} > \frac{k_u}{x_w}$, but the transportation cost is also higher, because $t(1-x_s) > t(x_w-a)$. Again, we can see the tradeoff between the delay cost and the transportation cost.

To see this more clearly, let V^s and V^w be the net valuations that the marginal consumer (x_s and x_w) enjoys from using unprioritized contents when priority is given to the strong CP and the weal CP respectively. Then, we have

$$V^s = \frac{v}{\rho}\frac{k_u}{1-x_s} - t(1-x_s),$$

$$V^w = \frac{v}{\rho}\frac{k_u}{x_w} - t(x_w-a).$$

When priority is given to the strong CP, the distance in the traveling cost is clearly higher because $1-x_s > x_w - a$. If the discounted utility were lower when the traffic amount is lower ($1-x_s < x_w$) as in the $M/M/1$ queuing model, it would be unambiguously better for the ISP to grant priority to the weak CP. However, the discounted utility from the lower traffic is higher because $v_u'(y) < 0$ in this subbandwidth model. So, it is possibly better for the ISP to grant priority to the strong CP. Then, why does the ISP prefer to grant priority to the strong CP for any $a > 0$? In other words, what is the main driving force for obtaining the completely opposite results in the queuing model and in the subbandwidth model? The essential difference between the two models is that an increase in the delay cost of prioritized contents due to an increase in the traffic also increases the delay cost of unprioritized contents in the queuing model, whereas an increase in discounting for prioritized contents decreases discounting for unprioritized contents. If x_p is the traffic from the prioritization CP, delay costs of the prioritized contents and unprioritized contents ($\frac{1}{\mu - x_p\lambda}$) and ($\frac{\mu}{\mu-\lambda}\frac{1}{\mu - x_p\lambda}$) both increase in x_p in the queuing model, whereas discounting for prioritized contents, $\frac{x_p}{1/2+\Delta}$, increases in x_p but discounting for unprioritized contents, $\frac{1-\tilde{x}}{1/2-\Delta}$, decreases in x_p.

Under the M/M/1 queuing model, the traffic amounts of the two CPs are not very different, because the delay costs of prioritized contents and unprioritized contents are affected in the same direction as we described above. However, under the subbandwidth model, a difference in the traffic amounts make the difference in delay costs between prioritized contents and unprioritized contents larger. So, the traffic difference of the two CPS tends to be large. Accordingly, the delay cost effect outweighs the traveling cost effect under the subbandwidth model, implying that the ISP prefers to prioritize the strong CP.

Finally, we consider the ISP's choice of Δ, i.e., how to allocate bandwidth to two CPs. Since the only revenue source of the ISP is the network access fee from users under unpaid prioritization and the network access fee is always higher when priority is given to the strong CP as shown in Proposition 5, we only need to examine

[19] Hermalin and Katz (2007) discuss the issue of net neutrality in terms of providing different qualities rather than providing different priorities.

[20] We will briefly discuss the issue of how to allocate the bandwidth k to the prioritized CP and the unprioritized CP, i.e., how to determine k_p and k_u at the end of this section.

[21] The adjustment factor $\rho(\tau_i)$ is defined slightly differently from Economides and Hermalin (2012) who assume that $\rho(\tau_i)$ is decreasing in τ_i.

[22] Since we assume that each end user requests one unit of contents, the total amount of contents from CP i is the same as the size of end users who request from CP i.

a change of $p_s(\Delta)$ with respect to a change in Δ. Note that the access fee $p_s(\Delta)$ is determined by the borderline user when priority is given to the strong CP. The next proposition shows that ISP's profit decreases as Δ increases.

Proposition 6. *The ISP prefers less difference in the bandwidths between prioritized CP and unprioritized CP.*

If the ISP increases Δ, i.e., amplifies asymmetry in the capacities for the prioritized CP and the unprioritized CP, it has two effects when the marginal consumer uses prioritized CP. It increases the transportation cost and also increases discounting of the value for prioritized contents. However, an increase in Δ itself reduces delay for prioritized contents. So, the overall effect may be ambiguous. However, this proposition implies that the two indirect negative effects outweigh the direct positive effect.

This proposition also suggests that the ISP prefers net neutrality to prioritization in this model, because $p_s(0;a) = p_w(0;a) > p_s(\Delta;a)$ for any $\Delta > 0$ and for any $a > 0$.

4. Social welfare

In this section, we discuss the implications of our results regarding the ISP's prioritization choice on social welfare.

4.1. Model with linear delay cost

We compare social welfare when the ISP grants priority to the strong CP and to the weak CP. If we assume that the market is fully covered and all end users get the same utility from the Internet service, we only need to compare the total costs consisting of the transportation cost (TC) and the delay cost which can be measured by the waiting time (T).

Let SC^s and SC^w be the total social cost when priority is given to the strong CP and to the weak CP respectively. In the model with a linear delay cost in Section 2, they are computed as follows:

$$\begin{aligned}SC^s &= \int_0^a (c_p(x_s) + t(a-x))dx + \int_a^{x_s} (c_p(x_s) + t(x-a))dx \\ &\quad + \int_{x_s}^1 (c_u(1-x_s) + t(1-x))dx, \\ SC^w &= \int_0^a (c_u(x_w) + t(a-x))dx + \int_a^{x_w} (c_u(x_w) + t(x-a))dx \\ &\quad + \int_{x_w}^1 (c_p(1-x_w) + t(1-x))dx.\end{aligned}$$

Thus, the difference is computed as

$$\begin{aligned}SC^w - SC^s &= \int_0^{x_w} K_1 dx - \int_{x_s}^1 K_2 dx \\ &\quad + \int_{x_w}^{x_s} (c_p(1-x_w) + t(1-x) - c_p(x_s) - t(x-a))dx, \\ &= \int_0^{x_w} (\alpha' x_w + \beta' - \alpha x_s - \beta)dx \\ &\quad + \int_{x_s}^1 (\alpha(1-x_w) + \beta - \alpha'(1-x_s) - \beta')dx \\ &\quad + \int_{x_w}^{x_s} (\alpha(1-x_w) + \beta + t(1-x) - \alpha x_s - \beta - t(x-a))dx,\end{aligned} \tag{37}$$

where $K_1 = c_u(x_w) - c_p(x_s) > 0$ and $K_2 = c_u(1-x_s) - c_p(1-x_w) > 0$. Note that $K_1 > K_2$ because

$$\begin{aligned}K_1 - K_2 &= \alpha' x_w + \beta' - \alpha x_s - \beta - [\alpha'(1-x_s) + \beta' - \alpha(1-x_w) - \beta'] \\ &= \alpha'(x_s + x_w - 1) - \alpha(x_s + x_w - 1) \\ &= (\alpha' - \alpha)(x_s + x_w - 1) > 0,\end{aligned}$$

where $\alpha' > \alpha > 0$ and $x_s + x_w > 1$ for any $a > 0$. Since $x_w > 1 - x_s$, it is clear that $\int_0^{x_w} K_1 dx > \int_{x_s}^1 K_2 dx$. In Fig. 4, P is the area between $c_u(x_w) + t|a-x|$ and $c_p(x_s) + t|a-x|$ from $x = 0$ to $x = x_w$, and S is the area between $c_u(1-x_s) + t(1-x)$ and $c_p(1-x_w) + t(1-x)$ from $x = x_s$ to $x = 1$. They correspond with the first term and the second term of (37) respectively. The area $Q - R$, which is the difference between two triangles, corresponds with the third term of (37). It is also clear from Fig. 4 that it is positive, because $C_w(a) > C_s(a)$. This implies that social welfare is higher when priority is given to the strong CP under a linear delay cost.

Proposition 7. *Social welfare is higher when priority is given to the strong CP than when it is given to the weak CP if $c_p(x) = \alpha x + \beta$ and $c_u(y) = \alpha' y + \beta'$ with $\alpha' > 0$.*

The formal proof is omitted, because it is clear from the above discussion and Fig. 4.

However, if $\alpha' < 0$, it implies that $K_1 < K_2$ because $(\alpha' - \alpha)(x_s + x_w - 1) < 0$. In this case, although $C^w < C^s$ implies that the third term of (37) is negative, it is ambiguous to compare $SW^w(a)$ and $SW^s(a)$ because it is not guaranteed that $\int_0^{x_w} K_1 dx < \int_{x_s}^1 K_2 dx$ due to $x_w > 1 - x_s$.

4.2. Queuing model

In an $M/M/1$ queing model, we have

$$\begin{aligned}SC^s &= \int_0^a (T_1^s + t(a-x))dx + \int_a^{x_s} (T_1^s + t(x-a))dx \\ &\quad + \int_{x_s}^1 (T_2^s + t(1-x))dx \\ &= T^s + TC^s, \\ SC^w &= \int_0^a (T_1^w + t(a-x))dx + \int_a^{x_w} (T_1^w + t(x-a))dx \\ &\quad + \int_{x_w}^1 (T_2^w + t(1-x))dx \\ &= T^w + TC^w,\end{aligned}$$

where

$$\begin{aligned}T^s &= \int_0^{x_s} T_1^s dx + \int_{x_s}^1 T_2^s dx = \frac{x_s}{\mu - x_s\lambda} + \frac{\mu}{\mu - \lambda}\frac{1-x_s}{\mu - x_s\lambda}, \\ T^w &= \int_0^{x_w} T_1^w dx + \int_{x_w}^1 T_2^w dx \\ &= \frac{x_w}{\mu - (1-x_w)\lambda} + \frac{\mu}{\mu - \lambda}\frac{x_w}{\mu - (1-x_w)\lambda},\end{aligned}$$

$$\begin{aligned}TC^s &= t\left[\int_0^a (a-x)dx + \int_a^{x_s} (x-a)dx + \int_{x_s}^1 (1-x)dx\right], \\ &= t\left[x_s^2 - (1+a)x_s + a^2 + \frac{1}{2}\right],\end{aligned} \tag{38}$$

$$\begin{aligned}TC^w &= t\left[\int_0^a (a-x)dx + \int_a^{x_w} (x-a)dx + \int_{x_w}^1 (1-x)dx\right], \\ &= t\left[x_w^2 - (1+a)x_w + a^2 + \frac{1}{2}\right].\end{aligned} \tag{39}$$

The following proposition implies that social welfare is higher when priority is granted to the strong CP.

Proposition 8. *The total social cost is lower when priority is granted to the strong CP, i.e., $SC^s(a) < SC^w(a)$, in the queuing model.*

We know that the total delay costs are the same regardless of which CP is given priority, i.e., $T^s = T^w$ due to the invariance result of Choi and Kim (2010). Thus, this proposition says that the total transportation cost is lower when priority is given to the strong CP. This is again because $x_s(a) + x_w(a) < 1 + a$, i.e., $x_s - \frac{1+a}{2} < \frac{1+a}{2} - x_w$ for any $a > 0$, implying that the marginal consumer is closer to $\frac{a+1}{2}$ which minimizes the total transportation costs when priority is given to the strong CP (CP 1) than when priority is given to the weak CP (CP 2). Moreover, the equilibrium outcome under prioritization yields lower welfare than the equilibrium outcome under net neutrality, because the traveling cost is minimized under net neutrality. This justifies the regulation of net neutrality.

4.3. Bandwidth subdivision model

In the subbandwidth model of Economides and Hermalin, social welfare can be defined by the discounted utility of users minus their transportation costs. Let SW^s, SW^w and U^s, U^w be social welfare and the discounted utility of users when the ISP gives priority to the strong CP and to the weak CP, respectively. Then, we have

$$\begin{aligned} SW^s &= U^s - TC^s \\ &= \int_0^a \left(\frac{v}{\alpha} \frac{k_p}{x_s} - t(a-x) \right) dx + \int_a^{x_s} \left(\frac{v}{\alpha} \frac{k_p}{x_s} - t(x-a) \right) dx \\ &\quad + \int_{x_s}^1 \left(\frac{v}{\alpha} \frac{k_u}{1-x_s} - t(1-x) \right) dx, \\ SW^w &= U^w - TC^w \\ &= \int_0^a \left(\frac{v}{\alpha} \frac{k_u}{x_w} - t(a-x) \right) dx + \int_a^{x_w} \left(\frac{v}{\alpha} \frac{k_u}{x_w} - t(x-a) \right) dx \\ &\quad + \int_{x_w}^1 \left(\frac{v}{\alpha} \frac{k_p}{1-x_w} - t(1-x) \right) dx. \end{aligned}$$

As proved in the appendix, straightforward algebra shows that $U^s = U^w$ (neutrality result), while $TC^s < TC^w$, implying that $SW^s > SW^w$. This is again simply because the borderline user is closer to $\frac{a+1}{2}$, meaning that the total transportation cost is lower when priority is given to the strong CP if $a > 0$.

Proposition 9. *For any $a > 0$, social welfare is higher when priority is granted to the strong CP, i.e., $SW^s(a) > SW^w(a)$, in the congestion model of Economides and Hermalin (2012).*

The neutrality result that $U^s = U^w$ is mainly due to the assumption that the adjustment factor is linear in delay $\tau_i = \frac{x_i}{k_i}$. For example, if the adjustment factor is quadratic, i.e., $\alpha(\tau_i) = \tau_i^2$, it may be socially better to grant priority to the weak CP, because otherwise the utility of a large amount of contents from the strong CP is discounted too much.[23]

It is also easy to see that social welfare under prioritization is lower than social welfare under net neutrality because the sum of the transportation costs are minimized under net neutrality, whereas the total utility of users is the same under the two regimes if the adjustment factor is linear. Again, this implies that net neutrality is socially desirable in this model of Economides and Hermalin (2012). This result is contrasted with Economides and Hermalin (2012) that derive conditions under which net neutrality is welfare superior to prioritization. In our model, users have different preferences for the two CPs. Prioritization gives some advantage to one of the two CPs thereby distorting the choice of users between the CPs which otherwise would be split equally among users. This negative effect of prioritization does not occur in Economides and Hermalin (2012), because they do not assume such different preferences of users towards CPs.[24]

5. Conclusion and caveats

In this paper, we showed that the ISP may give priority to the weaker content or to the stronger content, depending on the delay costs. The general insight is that which content is granted a priority has two conflicting effects, the traveling cost effect favoring the weak CP and the delay cost effect favoring the strong CP, and that the first effect outweighs the second effect if the delay cost of the unprioritized contents increases as the traffic increases.

Throughout the paper, we assumed that there are only two CPs and all users are informed of which CP is prioritized or not when they make decisions. However, it may be too optimistic to assume that all users know whether each individual CP is prioritized or not. In this incomplete information case, the ISP may strategically disclose the priority information of all CPs or may engage in strategic obfuscation by not revealing the information strategically. It may be an interesting issue to compare social welfare in the two cases and to examine the implications of the policy of the mandatory disclosure of priority information.

Of course, we admit that our simple model is still restrictive in the sense that the decision of whom to give priority is, in reality, determined by various other factors, in particular, heterogeneity in contents in terms of the disutility from delay, the sensitivity of the content requests to the possibility of delay, etc. Nonetheless, we believe that our insight will be at least worth to the ISP who considers prioritization when net neutrality is repealed.

Declaration of Competing Interest

The authors declare that they have no conflict of interest.

Data availability

Data will be made available on request.

Proof of Proposition 1. Eq. (25) can be rewritten as

$$\phi^{LS}(x_s) = \phi^{RS}(x_s), \tag{A.1}$$

where $\phi^{LS}(x) \equiv c_p(x) + t(x-a)$ and $\phi^{RS}(x) \equiv c_u(1-x) + t(1-x)$. Total differentiation of (A.1) leads to

$$\phi_1^{LS} dx_s + \phi_2^{LS} da = \phi_1^{RS} dx_s, \tag{A.2}$$

where $\phi_1^{LS} \equiv \frac{\partial \phi^{LS}}{\partial x}$, $\phi_2^{LS} \equiv \frac{\partial \phi^{LS}}{\partial a}$ and $\phi_1^{RS} \equiv \frac{\partial \phi^{RS}}{\partial x}$. From (A.2), we obtain $\frac{dx_s(a)}{da} = \frac{\phi_2^{LS}}{\phi_1^{RS} - \phi_1^{LS}} > 0$ because $\phi_1^{RS} = -c_u' - t < 0$, $\phi_1^{LS} = c_p' + t > 0$ and $\phi_2^{LS} = -t < 0$.

Similarly, Eq. (26) can be written as $\phi^{LW}(x_w) = \phi^{RW}(x_w)$ where $\phi^{RW}(x) \equiv c_u(x) + t(x-a)$ and $\phi^{LW}(x) \equiv c_p(1-x) + t(1-x)$. By total differentiation, we get $\frac{dx_w(a)}{da} = \frac{\phi_2^{RW}}{\phi_1^{LW} - \phi_1^{RW}} > 0$ because $\phi_1^{LW} < 0$, $\phi_1^{RW} > 0$ and $\phi_2^{RW} < 0$. (Note that the symbols "L" and "R" are switched for the weak content because we find it more convenient for proving Proposition 4.) □

[23] To see this, assuming that $v = 1$, the weighted sum of utility can be computed as $U(x) = \frac{(\Delta+1/2)^2}{x} + \frac{(\Delta-1/2)^2}{1-x}$ where x is the market share of the prioritized CP. Then, we have $U'(x) = \frac{\Gamma}{x^2(1-x)^2}$ where $\Gamma = (\Delta + 1/2 - x)(2x\Delta - \Delta - 1/2) < 0$ if Δ is not very large, because we assume that $\Delta + 1/2 > x$ for the prioritized CP. This implies that consumer utility is higher when priority is given to the weak CP, because $x_s(a) > 1 - x_w(a)$ for $a > 0$.

[24] Although we assume that users have inelastic demands for contents, Economides and Hermalin (2012) allow variable demands for contents with respect to the transmission time and argue that net neutrality is welfare superior to any prioritization if the discounted utility function is concave by using Jensen's inequality. Kim (2022) also considered the case of variable demands and obtained a similar result as Economides and Hermalin (2012) in a product differentiation model. See Proposition 3 of Kim (2022).

Proof of Proposition 2. From (11) and (12), we have

$$\frac{dp_s^*(a)}{da} = c_u' \frac{dx_s(a)}{da} + t\frac{dx_s(a)}{da} > 0,$$

$$\frac{dp_w^*(a)}{da} = c_p' \frac{dx_w(a)}{da} + t\frac{dx_w(a)}{da} > 0,$$

due to $c_p', c_u' > 0$ and Proposition 1. □

Proof of Proposition 3. Let $c_p(x) = \alpha x + \beta$ and $c_u(y) = \alpha' y + \beta'$ where $\alpha' > \alpha > 0$ and $\beta' \geq \beta \geq 0$. We assume that $\beta' - \beta \geq \alpha$ to be consistent with the assumption that $c_p(z) < c_u(1-z)$ for any $z \in (0,1)$.

(i) From (7) and (8), we have

$$x_s = \frac{(1+a)t + \alpha' + \beta' - \beta}{2t + \alpha' + \alpha}, \tag{A.3}$$

$$x_w = \frac{(1+a)t + \alpha + \beta - \beta'}{2t + \alpha' + \alpha}. \tag{A.4}$$

Note that $\frac{\partial x_s}{\partial a} = \frac{\partial x_w}{\partial a}$. Also, if $\beta' - \beta \geq \alpha$, we can easily check that $x_s, x_w < 1$ if $a < 1 - \frac{\beta'-\beta-\alpha}{t}$, and $x_s, x_w > a$ if $a < \frac{t+\alpha+\beta-\beta'}{t+\alpha'+\alpha} (< 1)$.

Accordingly, we have

$$\begin{aligned} C^s(a) &= c_p(x_s) + t(x_s - a) \\ &= \frac{(\alpha + t)[(1+a)t + \alpha' + \beta' - \beta]}{2t + \alpha' + \alpha} + \beta - ta, \end{aligned} \tag{A.5}$$

$$\begin{aligned} C^w(a) &= c_u(x_w) + t(x_w - a) \\ &= \frac{(\alpha' + t)[(1+a)t + \alpha + \beta - \beta']}{2t + \alpha' + \alpha} + \beta' - ta. \end{aligned} \tag{A.6}$$

Note that $C^s(0) = C^w(0)$. Also, (A.5) and (A.6) lead to

$$C^w(a) - C^s(a) = A(1+a)t + B,$$

where

$$A = \frac{(\alpha' - \alpha)}{2t + \alpha' + \alpha}, \tag{A.7}$$

$$B = \frac{(\alpha' + t)(\alpha + \beta - \beta') - (\alpha + t)(\alpha' + \beta' - \beta)}{2t + \alpha' + \alpha}. \tag{A.8}$$

Since $\alpha' > \alpha$ it is clear that $A > 0$. Therefore, $C^w(a) - C^s(a)$ is increasing in a.

(ii) This follows directly from (i), because $C^w(0) = C^s(0)$. □

Proof of Proposition 4. Let $z = 1 - x$. Eqs. (25) and (26) are rewritten as

$$\frac{1}{\mu - x_s\lambda} + t(x_s - a) = \frac{\mu}{\mu - \lambda}\frac{1}{\mu - x_s\lambda} + t(1 - x_s), \tag{A.9}$$

$$\frac{1}{\mu - z_w\lambda} + tz_w = \frac{\mu}{\mu - \lambda}\frac{1}{\mu - z_w\lambda} + t(1 - a - z_w), \tag{A.10}$$

where $z_w = 1 - x_w$. Eqs. (A.9) and (A.10) can be rearranged into

$$\frac{\theta}{\mu - x_s\lambda} = t(2x_s - a - 1), \tag{A.11}$$

$$\frac{\theta}{\mu - z_w\lambda} = t(2z_w + a - 1), \tag{A.12}$$

where $\theta = \frac{\lambda}{\mu-\lambda} > 0$. If $a = 0$, it follows from (A.11) and (A.12) that $x_s = z_w = 1 - x_w$, i.e., x_s and x_w must be symmetric around $\frac{1}{2}$, assuming the uniqueness of the solution. (The uniqueness will be proved in Lemma 1.)

Let $f(x) = \frac{1}{\mu-\lambda x}$ and $g(x) = \sigma f(x)$ where $\sigma = \frac{\mu}{\mu-\lambda} > 1$. Then, we have

$$f'(x) = \frac{\lambda}{(\mu - \lambda x)^2} > 0, \tag{A.13}$$

$$f''(x) = \frac{2\lambda^2}{(\mu - \lambda x)^3} > 0. \tag{A.14}$$

Let the solutions for (A.9) and (A.10) given a be $x_s(a)$, $z_w(a)$ and the corresponding x_w be $x_w(a)$. Figure 2 illustrates $x_s(a)$ and $x_w(a)$ which are determined from (A.9) and (A.10) when $a > 0$.

Let us denote the left hand side (LHS) and the right hand side (RHS) of (A.9) by $\phi^{LS}(x)$ and $\phi^{RS}(x)$ respectively, and denote LHS and RHS of (A.10) by $\phi^{LW}(z)$ and $\phi^{RW}(z)$ respectively.

With a change in a, the solutions for $x_s(a)$ and $z_w(a)$ must move along $\phi^{RS}(x)$ and $\phi^{LW}(z)$ respectively which are defined by

$$\phi^{LW}(z) = f(z) + ty, \tag{A.15}$$

$$\phi^{RS}(x) = g(x) + t(1 - x). \tag{A.16}$$

To elaborate on the definitions of C^s and C^w, we have

$$C^s(a) = \phi^{RS}(x_s(a)) = \phi^{LS}(x_s(a)), \tag{A.17}$$

$$C^w(a) = \phi^{RW}(x_w(a)) = \phi^{LW}(x_w(a)). \tag{A.18}$$

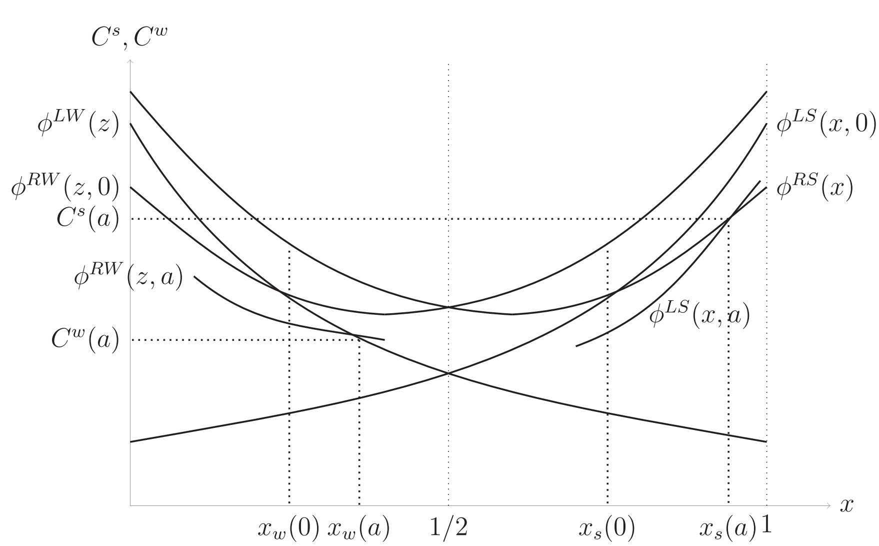

Fig. 2. Comparison between C^s and C^w ($a > 0$).

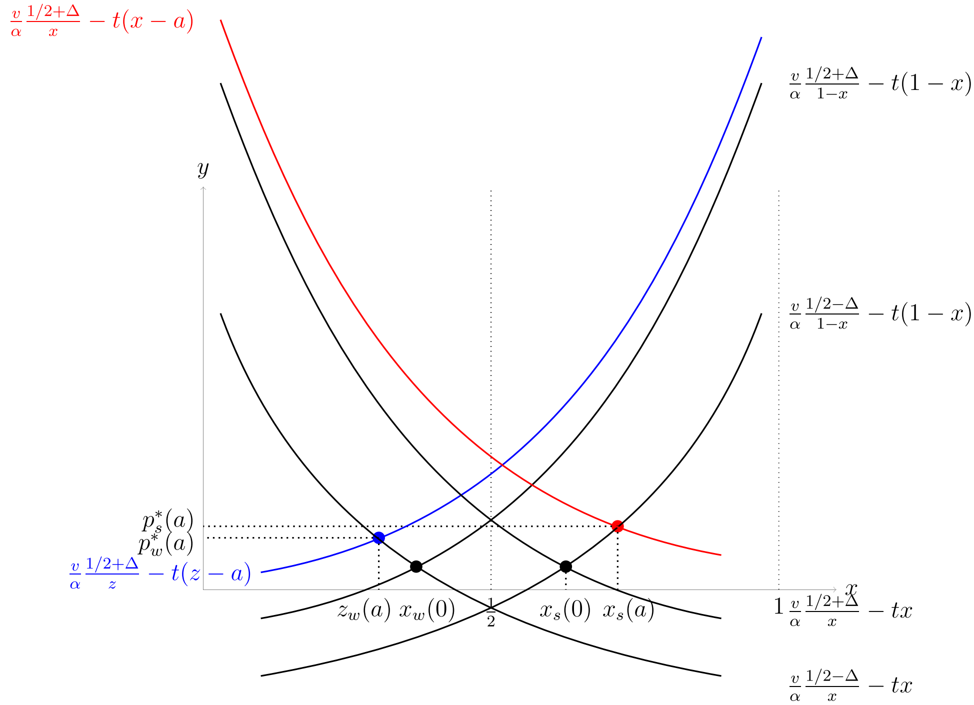

Fig. 3. Comparison between $p_s(a)$ and $p_w(a)$ $(a > 0)$.

Then, we know that $C^s(0) = C^w(0)$, i.e., $\phi^{RW}(x_w(0)) = \phi^{LS}(x_s(0))$ by symmetry between $x_s(0)$ and $x_w(0)$. We also know that $\frac{dx_s(0)}{da} = -\frac{dz_w(0)}{da} = \frac{dx_w(0)}{da}$ from (A.11) and (A.12). Also, by taking derivatives of (A.15) and (A.16), we obtain

$$\frac{\partial \phi^{LW}(z)}{\partial x} = \frac{\partial \phi^{LW}(z)}{\partial z}\frac{\partial z}{\partial x} = -(f'(z) + t) = -f'(z) - t, \tag{A.19}$$

$$\frac{\partial \phi^{RS}(x)}{\partial x} = g'(x) - t = \sigma f'(x) - t. \tag{A.20}$$

Now, let us compare $C^s(a)$ and $C^w(a)$. Since $C^s(0) = C^w(0)$ and $\frac{dx_s(0)}{da} = \frac{dx_w(0)}{da}$, we only need to compare changes in $g(x_s(a))$ and $f(z_s(a))$ in a, because the effects of a change in a on the second terms of $\phi^{LW}(z)$ and $\phi^{RS}(x)$ are the same. Since $g'(x) > 0$ and $f'(z) > 0$ (i.e., $\frac{\partial f(1-x)}{\partial x} = \frac{\partial f(1-x)}{\partial z}\frac{\partial z}{\partial x} = -f'(z) < 0$), it follows that $C^s(a) > C^w(a)$ for any $a > 0$. (More specifically, $g(x_s(a)) = \sigma f(x_s(a)) > \sigma f(x_s(0)) > f(z_w(0)) > f(z_w(a))$, since $f'(x) > 0$, $x_s(0) = z_w(0)$, $x_s(a) > x_s(0)$ and $z_w(0) > z_w(a)$ for any $a > 0$.) This implies that $p_s^*(a) < p_w^*(a)$ for any $a > 0$. □

Lemma 1. *$x_s(0)$ and $x_w(0)$ are unique if any.*

Proof of Lemma 1. The uniqueness of $x_s(0)$ is clear from the observation that $x_s(0)$ is a solution for the quadratic equation $(2x - 1)(\mu - x\lambda) = \frac{\theta}{t}$ where $\frac{\mu}{\lambda} > 1$. Thus, due to symmetry, $x_w(0)$ is also unique. □

Proof of Proposition 5. Define $y_s(a)$ and $y_w(a)$ by

$$y_s(a) \equiv \frac{v}{\alpha}\frac{k_p}{x_s(a)} - t(x_s(a) - a) = \frac{v}{\alpha}\frac{k_u}{1 - x_s(a)} - t(1 - x_s(a)), \tag{A.21}$$

$$y_w(a) \equiv \frac{v}{\alpha}\frac{k_u}{x_w(a)} - t(x_w(a) - a) = \frac{v}{\alpha}\frac{k_p}{1 - x_w(a)} - t(1 - x_w(a)). \tag{A.22}$$

Let $k_p = \frac{1}{2} + \Delta$ and $k_u = \frac{1}{2} - \Delta$ where $\Delta > 0$. Consider the benchmark case that $a = 0$. Then, it is clear that $y_w(0) = y_s(0)$ because $x_w(0)$ and $x_s(0)$ are symmetric around $\frac{1}{2}$. (See Fig. 3) If $a > 0$, $x_s(a)$ and $y_s(a)$ are determined by (A.21), i.e.,

$$y_s(a) \equiv \frac{v}{\alpha}\frac{1/2 + \Delta}{x_s(a)} - t(x_s(a) - a) = \frac{v}{\alpha}\frac{\frac{1}{2} - \Delta}{1 - x_s(a)} - t(1 - x_s(a)). \tag{A.23}$$

On the other hand, to compare $y_w(a)$ with $y_s(a)$, let $z = 1 - x$ and rewrite (A.22) in terms of z. Then, (A.22) can be rewritten as

$$\frac{v}{\alpha}\frac{1/2 - \Delta}{1 - z_w(a)} - t(1 - z_w(a) - a) = \frac{v}{\alpha}\frac{\frac{1}{2} + \Delta}{z_w(a)} - tz_w(a). \tag{A.24}$$

Let the solution of (A.24) be $z_w(a)$. Note that $x_s(0) = z_w(0)$ so that $y_s(0) = y_w(0)$. Now, suppose that a is increased marginally from $a = 0$. To compare $\frac{dy_w(a)}{da}$ and $\frac{dy_s(a)}{da}$, define

$$\Phi^{LS}(x, a) = \frac{v}{\alpha}\frac{1/2 + \Delta}{x} - t(x - a), \tag{A.25}$$

$$\Phi^{RS}(x, a) = \frac{v}{\alpha}\frac{1/2 - \Delta}{1 - x} - t(1 - x), \tag{A.26}$$

$$\Phi^{LW}(z, a) = \frac{v}{\alpha}\frac{1/2 - \Delta}{1 - z} - t(1 - z - a), \tag{A.27}$$

$$\Phi^{RW}(z, a) = \frac{v}{\alpha}\frac{1/2 + \Delta}{z} - tz. \tag{A.28}$$

Then, $x_s(a)$ and $z_w(a)$ satisfy

$$F^S(x, a) = \Phi^{LS}(x, a) - \Phi^{RS}(x, a) = 0, \tag{A.29}$$

$$F^W(z, a) = \Phi^{RW}(z, a) - \Phi^{LW}(z, a) = 0, \tag{A.30}$$

respectively. To compare $\frac{\partial x_s(a)}{\partial a}$ and $\frac{\partial z_w(a)}{\partial a}$, we differentiate (A.29) and (A.30) totally to get

$$F_x^S dx_s + t da = 0, \tag{A.31}$$

$$F_z^W dz_w + t da = 0, \tag{A.32}$$

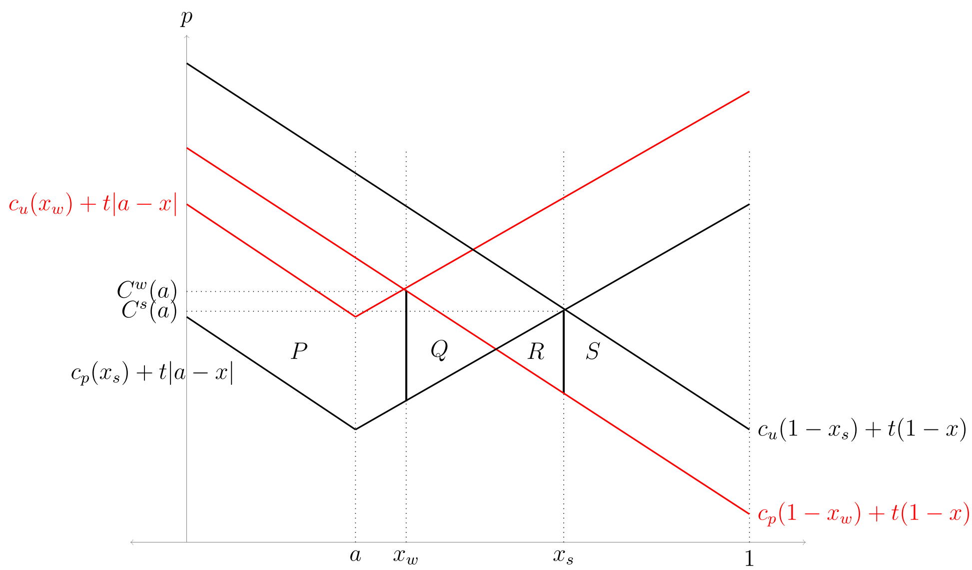

Fig. 4. Welfare Comparison under Linear Delay Cost.

where $F_x^S(x,a) = \Phi_x^{LS}(a) - \Phi_x^{RS}(a)$ and $F_z^W(z,a) = \Phi_z^{RW}(a) - \Phi_z^{LW}(a)$. Therefore, it follows that

$$\frac{dx_s}{da} = -\frac{t}{F_x^S}, \tag{A.33}$$

$$\frac{dz_w}{da} = -\frac{t}{F_z^W}, \tag{A.34}$$

where

$$F_x^S(x,a) = -\left[\frac{1/2+\Delta}{x^2} + \frac{1/2-\Delta}{(1-x)^2} + 2t\right], \tag{A.35}$$

$$F_z^W(z,a) = -\left[\frac{1/2+\Delta}{z^2} + \frac{1/2-\Delta}{(1-z)^2} + 2t\right]. \tag{A.36}$$

Therefore, we obtain

$$\frac{\partial x_s(a)}{\partial a} = -\frac{t}{\frac{1/2+\Delta}{x^2} + \frac{1/2-\Delta}{(1-x)^2} + 2t}, \tag{A.37}$$

$$\frac{\partial x_w(a)}{\partial a} = -\frac{\partial z_w(a)}{\partial a} = \frac{t}{\frac{1/2+\Delta}{(1-x)^2} + \frac{1/2-\Delta}{x^2} + 2t}. \tag{A.38}$$

To show that $\frac{\partial x_s(a)}{\partial a} > \frac{\partial x_w(a)}{\partial a}$, it needs to show that

$$\frac{1/2+\Delta}{x_s^2} + \frac{1/2-\Delta}{(1-x_s)^2} < \frac{1/2+\Delta}{(1-x_w)^2} + \frac{1/2-\Delta}{x_w^2},$$

or equivalently,

$$(1/2+\Delta)\left[\frac{1}{x_s^2} - \frac{1}{(1-x_w)^2}\right] + \left(\frac{1}{2} - \Delta\right)\left[\frac{1}{(1-x_s)^2} - \frac{1}{x_w^2}\right] < 0.$$

Since $x_s(a) > 1 - x_w(a)$, we have $\frac{1}{x_s^2} - \frac{1}{(1-x_w)^2} < 0$ and $\frac{1}{(1-x_s)^2} - \frac{1}{x_w^2} > 0$. Therefore, we only need to show that

$$\nabla \equiv \left[\frac{1}{(1-x_s)^2} - \frac{1}{x_w^2}\right] - \left[\frac{1}{(1-x_w)^2} - \frac{1}{x_s^2}\right] > 0. \tag{A.39}$$

Due to Lemma 2, we have

$$\nabla = \left[\frac{1}{(1-x_s)^2} + \frac{1}{x_s^2}\right] - \left[\frac{1}{x_w^2} + \frac{1}{(1-x_w)^2}\right] > 0, \tag{A.40}$$

since $x_s + x_w > 1$ implies that $\frac{1}{2} - x_w < x_s - \frac{1}{2}$. Therefore, it follows that $\frac{\partial x_s(a)}{\partial a} > \frac{\partial x_w(a)}{\partial a}$ for any $a > 0$. Since $x_s(a)$ and $z_w(a)$ are determined on the curves $\Phi^{RS}(x,0)$ and $\Phi^{LW}(z,0)$ which are symmetric around $\frac{1}{2}$, the proof is completed. (See Fig. 3) □

Lemma 2. *Define* $h(x) = \frac{1}{x^2} + \frac{1}{(1-x)^2}$ *for* $x \in (0,1)$. *Then,* $h(x)$ *is symmetric around* $x = \frac{1}{2}$ *and* $h'(x) \gtreqless 0$ *iff* $x \gtreqless \frac{1}{2}$.

Proof of Proposition 6. We differentiate (A.29) with respect to Δ and get

$$F_x^S dx_s + F_\Delta d\Delta = 0, \tag{A.41}$$

where $F_\Delta^S(x,a) = \frac{1/2+\Delta}{x^2} + \frac{1/2-\Delta}{(1-x)^2} + 2t$ and $F_\Delta^S = \frac{1}{x_s} + \frac{1}{1-x_s}$. Thus, we obtain

$$\frac{dx_s(\Delta)}{d\Delta} = \frac{\frac{1}{x} + \frac{1}{1-x}}{\frac{1/2+\Delta}{x^2} + \frac{1/2-\Delta}{(1-x)^2} + 2t} > 0. \tag{A.42}$$

Now, to find $\frac{dp_s(\Delta)}{d\Delta}$, we have

$$\begin{aligned}\frac{\Phi^{LS}}{d\Delta} &= -\Phi_x^{LS}\frac{dx_s(\Delta)}{d\Delta} + \Phi_\Delta^{LS} \\ &= -\left[\frac{1/2+\Delta}{x^2} + t\right]\frac{\frac{1}{x} + \frac{1}{1-x}}{\frac{1/2+\Delta}{x^2} + \frac{1/2-\Delta}{(1-x)^2} + 2t} + \frac{1}{x} < 0.\end{aligned} \tag{A.43}$$

To show that $\frac{\Phi^{LS}}{d\Delta} < 0$, we can see from (A.43) that we only need to show that

$$x_s > \frac{\frac{1/2+\Delta}{x_s^2} + \frac{1/2-\Delta}{(1-x_s)^2} + 2t}{\left(\frac{1}{x_s} + \frac{1}{1-x_s}\right)\left(\frac{1/2+\Delta}{x_s^2} + t\right)}. \tag{A.44}$$

Let $\chi \equiv x\left(\frac{1}{x} + \frac{1}{1-x}\right)\left(\frac{1/2+\Delta}{x_s^2} + t\right) - \left[\frac{1/2+\Delta}{x^2} + \frac{1/2-\Delta}{(1-x)^2} + 2t\right]$. We have

$$\chi = \frac{x_s}{1-x_s}\left(\frac{1/2+\Delta}{x_s^2} + t\right) - t > 0, \tag{A.45}$$

since $x_s > \frac{1}{2}$ so that $\frac{x_s}{1-x_s} > 1$. This implies that $\frac{dp_s(\Delta)}{d\Delta} = \frac{d\Phi^{LS}}{d\Delta} < 0$. □

Proof of Proposition 8. To compare $SC^s(a)$ and $SC^w(a)$, we first compare the delay costs $T^s(a)$ and $T^w(a)$. Letting $z_w = 1 - x_w$, we have

$$T^s(a) = \frac{x_s(a)}{\mu - x_s(a)\lambda} + \frac{\mu}{\mu - \lambda}\frac{1 - x_s(a)}{\mu - x_s(a)\lambda},$$

$$T^w(a) = \frac{z_w(a)}{\mu - z_w(a)\lambda} + \frac{\mu}{\mu-\lambda}\frac{1-z_w(a)}{\mu - z_w(a)\lambda}.$$

Define $H(x) = xf(x) + \sigma(1-x)f(x) = [x + \sigma(1-x)]f(x)$ where $f(x) = \frac{1}{\mu - x\lambda}$ and $\sigma = \frac{\mu}{\mu-\lambda} > 1$. We have

$$\begin{aligned} H'(x) &= (1-\sigma)f(x) + [x+\sigma(1-x)]f'(x) \\ &= (1-\sigma)\frac{1}{\mu - x\lambda} + [x+\sigma(1-x)]\frac{\lambda}{(\mu-x\lambda)^2} \\ &= \frac{(1-\sigma)\mu + \sigma\lambda}{(\mu - x\lambda)^2} \\ &= 0, \end{aligned}$$

which implies that $T^s(a) = T^w(a)$ for any a. This is mainly due to Lemma 3(iii) of Choi and Kim (2010).

Now, we compare the transportation costs $TC^s(a)$ and $TC^w(a)$. Let $\tilde{x}$ be the marginal consumer who is indifferent between CP 1 and CP 2. From (38) and (39), we can write the total transportation cost as

$$\begin{aligned} TC(\tilde{x}) &= t\left[\int_0^a (a-x)dx + \int_a^{\tilde{x}} (x-a)dx + \int_{\tilde{x}}^1 (1-x)dx\right], \\ &= t\left[\tilde{x}^2 - (1+a)\tilde{x} + a^2 + \frac{1}{2}\right]. \end{aligned}$$

This is minimized at $\tilde{x} = \frac{a+1}{2}$ which can be attained under net neutrality.

To show that $SC^s(a) < SC^w(a)$, it suffices to show that $x_s(a) - \bar{x}(a) < \bar{x}(a) - x_w(a)$ where $\bar{x} = \frac{a+1}{2}$, due to symmetry of $TC(\tilde{x})$ around $\tilde{x} = \frac{a+1}{2}$.

From (A.11) and (A.12), we have

$$\left[\frac{\lambda^2}{(\mu-\lambda)(\mu - x_s\lambda)^2} + 2t\right]dx_s = tda, \tag{A.46}$$

$$\left[\frac{\lambda^2}{(\mu-\lambda)(\mu - z_w\lambda)^2} + 2t\right]dz_w = -tda. \tag{A.47}$$

Therefore, we obtain

$$\frac{x_s}{da} = \frac{1}{2+X} \in (0, \frac{1}{2}), \tag{A.48}$$

$$\frac{x_w}{da} = -\frac{z_w}{da} = \frac{1}{2+Y} \in (0, \frac{1}{2}), \tag{A.49}$$

where

$$X = \frac{\lambda^2}{t(\mu-\lambda)(\mu - x_s\lambda)^2} > 0,$$

$$Y = \frac{\lambda^2}{t(\mu-\lambda)(\mu - z_w\lambda)^2} > 0.$$

Thus, we have

$$\frac{dx_s}{da} + \frac{dx_w}{da} < 1, \tag{A.50}$$

implying that

$$\frac{dx_s}{da} - \frac{d\bar{x}}{da} < \frac{d\bar{x}}{da} - \frac{dx_w}{da}, \tag{A.51}$$

where $\frac{d\bar{x}}{da} = \frac{1}{2}$. This implies that $x_s(a) + x_w(a) < 1 + a$, i.e., $x_s(a) - \frac{1+a}{2} < \frac{1+a}{2} - x_w(a)$ by Lemma 3. By symmetry of $TC(\tilde{x})$ around $\tilde{x} = \frac{1+a}{2}$, it directly follows that $TC^s(a) < TC^w(a)$, implying that $SC^s(a) < SC^w(a)$. □

Lemma 3. *(i)* $x_s(0) + x_w(0) = 1$ *and (ii)* $\frac{x_s(a)}{da} + \frac{x_w(a)}{da} < 1 \iff$ *(iii)* $x_s(a) + x_w(a) < 1 + a$.

Proof of Lemma 3. ($\Longrightarrow$) From (i), we have $x_s(0) + x_w(0) = 1$ when $a = 0$. How, if a is increased by Δa, (ii) implies that

$$\begin{aligned} & \Delta x_s(a) + \Delta x_w(a) < \Delta a \\ \iff & (x_s(a) - x_s(0)) + (x_w(a) - x_w(0)) < a \\ \iff & x_s(a) + x_w(a) < x_s(0) + x_w(0) + a = 1 + a. \end{aligned}$$

($\Longleftarrow$) This is trivial by differentiating (iii). □

Proof of Proposition 9. We have

$$\begin{aligned} U^s &= \frac{v}{\alpha}\left[\int_0^{x_s} \frac{1/2+\Delta}{x_s}dx + \int_{x_s}^1 \frac{1/2-\Delta}{1-x_s}dx\right] \\ &= \frac{v}{\alpha}, \\ U^w &= \frac{v}{\alpha}\left[\int_0^{x_w} \frac{1/2+\Delta}{x_w}dx + \int_{x_w}^1 \frac{1/2-\Delta}{1-x_w}dx\right] \\ &= \frac{v}{\alpha}. \end{aligned}$$

Thus, it remains to show that $TC^s(a) < TC^w(a)$. From (A.52) and (A.53), we have

$$\frac{dx_s}{da} = \frac{1}{(1/t)\left[\frac{1/2+\Delta}{x_s^2} + \frac{1/2-\Delta}{(1-x_s)^2}\right] + 2} \in (0, \frac{1}{2}), \tag{A.52}$$

$$\frac{dx_w}{da} = -\frac{dz_w}{da} = \frac{1}{(1/t)\left[\frac{1/2+\Delta}{x_w^2} + \frac{1/2-\Delta}{(1-x_w)^2}\right] + 2} \in (0, \frac{1}{2}). \tag{A.53}$$

Therefore, it follows that

$$\begin{aligned} \frac{dx_s}{da} + \frac{dx_w}{da} = & \frac{1}{(1/t)\left[\frac{1/2+\Delta}{x_s^2} + \frac{1/2-\Delta}{(1-x_s)^2}\right] + 2} \\ & + \frac{1}{(1/t)\left[\frac{1/2+\Delta}{x_w^2} + \frac{1/2-\Delta}{(1-x_w)^2}\right] + 2} < 1, \end{aligned} \tag{A.54}$$

which implies that $\frac{dx_s}{da} - \frac{1}{2} < \frac{1}{2} - \frac{dx_w}{da}$, i.e., $TC^s(a) < TC^w(a)$ for any $a > 0$. This completes the proof. □

References

Bourreau, M., Kourandi, F., Valletti, T., 2015. Net neutrality with competing internet platforms. J. Ind. Econ. 63, 30–73.

Cheng, H.K., Bandyopadhyay, S., Guo, H., 2011. The debate on net neutrality: a policy perspective. Inf. Syst. Res. 22, 60–82.

Choi, J.P., Jeon, D.-S., Kim, B.C., 2018. Net neutrality, network capacity, and innovation at the edges. J. Ind. Econ. 66, 172–204.

Choi, J.P., Kim, B.C., 2010. Net neutrality and investment incentives. Rand J. Econ. 41, 446–471.

DeGraba, P., 1990. Input market price discrimination and the choice of technology. Am. Econ. Rev. 80, 1246–1253.

Economides, N., Hermalin, B., 2012. The economics of network neutrality. Rand J. Econ. 43, 602–629.

Gautier, A., Somogyi, R., 2020. Prioritization vs zero-rating: discrimination on the internet. Int. J. Ind Organiz 73, 1–25.

Hermalin, B., Katz, M., 2007. The economics of product-line restrictions with an application to the network neutrality debate. Inf. Econ. Policy 19, 215–248.

Kim, J.Y., 2022. On the invariance result of net neutrality. Rev. Netw. Econ. 20, 139–157.

Krämer, J., Wiewiorra, L., 2012. Network neutrality and congestion sensitive content providers: implications for content variety, broadband investment, and regulation. Inf. Syst. Res. 23, 1303–1321.

Li, Y., 2014. A note on third degree price discrimination in intermediate good markets. J. Ind. Econ. 62. 554–554