RAND Journal of Economics
Vol. 43, No. 4, Winter 2012
pp. 602–629

The economics of network neutrality

Nicholas Economides*

and

Benjamin E. Hermalin**

Under the current regime for Internet access, "network neutrality," parties are billed only by the Internet service provider (ISP) through which they connect to the Internet; pricing is not contingent on the content being transmitted. Recently, ISPs have proposed that content and applications providers pay them additional fees for accessing the ISPs' residential clients, as well as fees to prioritize certain content. We analyze the private and social implications of such fees when the network is congested and more traffic implies greater delays. We derive conditions under which network neutrality would be welfare superior to any feasible scheme for prioritizing service.

1. Introduction

■ At issue when a network or platform facilitates the transactions of other parties are the prices it charges these parties for this facilitation. Since its commercialization in the mid-1990s, a party (website or household) connecting to the Internet pays only its direct provider of access. Other carriers on the network, even if they are subsequently involved in transporting on behalf of that party, do not collect payment from that party.

Over time, the market for connecting commercial websites has become relatively competitive. At the same time, however, there is significant monopoly power in residential broadband Internet access, that is, the connections households make to their Internet service provider (ISP). Moreover, even when ISPs might compete for residential customers, the fact that these customers almost always connect through a single ISP (single home) means that an individual household's ISP has a monopoly on others' access to it.

To date, consumers' ISPs have not directly charged websites for the content that passes over ISPs' networks into consumers' homes, so-called last-mile service. In summer 2005, the Federal Communications Commission (FCC) reclassified Internet services in a way that allowed

*New York University, University of California, Berkeley, and NET Institute; economides@stern.nyu.edu.
**University of California, Berkeley; hermalin@haas.berkeley.edu.

The authors thank seminar participants at Telecommunications Policy Research Conference 2010, the Stern School of Business, the University of California, Berkeley, the 38th Research Conference on Communication, Information and Internet Policy, TILEC Conference on the Law and Economics of Media and Telecommunication 2011, the 25th Summer IO Conference, and the University of Southern California. The authors also thank two anonymous referees and the editor, James Hosek, for many valuable suggestions. The financial support of the Newhouse Foundation and of the Thomas and Alison Schneider Distinguished Professorship in Finance is gratefully acknowledged.

Copyright © 2012, RAND.

for the possibility of ISPs charging content and application providers for last-mile service.[1] In fall 2005, AT&T proposed that a new fee be paid directly to it by application and content providers whose information packets were carried by AT&T to residential customers, irrespective of where those application and content providers connected to the Internet. In particular, even application and content providers that did not connect to the Internet through AT&T would be subject to being charged by AT&T. Soon other telecom and cable TV companies proposed doing the same. In short, residential-access ISPs proposed introducing fees to application and content providers (which are now zero). Additionally, they proposed dividing their residential-access connection pipe into "fast-access" and "slow-access" lanes and charge application and content providers on the basis of the speed of access they chose. The status quo (zero fees and no discrimination) has been dubbed *network neutrality*, so imposing fees to the "other side" of the network or introducing price discrimination are considered departures from network neutrality.

The issue of network neutrality is controversial and complex. In October 2009, the FCC proposed rules that would impose network neutrality by law. Broadly, the proponents of neutrality are consumer groups, the Obama administration,[2] and companies such as Google, Skype, Amazon eBay, and Microsoft, whereas the opponents are Cisco Systems and telecom and cable companies.

There are many advocacy papers written on the subject, but significantly less economic research. In the latter category, Economides and Tåg (2012), in a model of differentiated consumers and content providers and monopoly or duopoly ISPs, compare a world of zero fees by ISPs to one with positive fees. They find that, for most parameter values, total surplus is higher at zero fees. With respect to the issue of price discrimination via differential quality of service, Hermalin and Katz (2007) find that restricting an ISP to a single quality level has the following effects: (i) application and content providers that would otherwise have purchased slow access are excluded from the market; (ii) providers "in the middle" of the market utilize higher and more efficient speeds than otherwise; and (iii) providers at the top utilize lower and less efficient speeds than otherwise. Total surplus may rise or fall, although their analysis suggests that prohibiting discrimination is likely to harm welfare. Choi and Kim (2010) analyze prioritization using a queuing model.[3] They consider two application or content providers that are competing for eyeballs. In equilibrium, prioritization is bought by the more efficient provider, leading to consumers switching to this provider. This improves productive efficiency but also induces utility losses from consumers who do not buy their most preferred service. In their model, the net effect of departing from neutrality on welfare is, thus, ambiguous.

Our model differs from the earlier literature in a number of ways. First, we explicitly assume congestion. This means there could be a purely allocative reason to depart from network neutrality beyond any reasons stemming from the exercise of monopoly power. Second, we assume that ISPs are not producing their own content and applications for which they may seek priority vis-à-vis the content and applications of independent producers. This assumption abstracts away from one incentive ISPs could have to depart from network neutrality.[4] Additionally, unlike Choi

[1] Until that point, telecommunication facility based Internet transmissions had been subject to common-carrier regulation that compelled nondiscrimination. Other Internet transmissions, those not telecommunication facility based, were not subject to common-carrier regulation. Thus, DSL service was considered a common-carrier service, and so subject to nondiscrimination provisions. Cable modem service, in contrast, was not considered common-carrier service, and therefore was not subject to such provisions. In 2005, the FCC reclassified Internet transmissions from "telecommunications services" to "information services." See *Nat'l Cable & Telecomm. Assn. v. Brand X Internet Services*, 125 S. Ct. 2688 (2005). This implied that there were no longer "nondiscrimination" restrictions on Internet service pricing.

[2] See web.archive.org/web/20100304063856/http://www.barackobama.com/pdf/issues/technology/Fact_Sheet_Innovation_and_Technology.pdf. See also Federal Communications Commission Report and Order 10-201, released December 23, 2010.

[3] Like Choi and Kim (2010), Cheng, Bandyopadhyay, and Guo (2011), and Krämer and Wiewiorra (2010) also employ an M/M/1 queuing model to study network neutrality.

[4] In particular, there is a fear that ISPs that are also content providers, such as Comcast, may engage in vertical foreclosure. The recent battle between Comcast and Level 3 Communications (see, e.g., "Comcast Fee Ignites Fight over Videos on Internet" in the *New York Times*, November 30, 2010) illustrates these concerns.

© RAND 2012.

and Kim and others, we do not take the amount of traffic (e.g., number of packets) sent by a given content provider as fixed. Rather, the amount of content purchased by each household from any given content provider can vary. This is a critical extension insofar as it means expansion in bandwidth does not necessarily increase speed because larger bandwidth will attract more traffic. In this sense, our model allows for the effect, commonly observed with physical highways, that adding lanes does not always significantly reduce commute times.[5] In turn, this has important implications for the sort of second-degree price discrimination via differential quality considered by Choi and Kim, Hermalin and Katz, and others.

Another key distinction between our work and some previous analyses is that we seek to model the provision of quality (transmission speeds) from first principles, rather than in a "black box" manner (e.g., as in Hermalin and Katz). This proves critical for the following reason. One question is whether, absent price-discrimination motives, there would be a welfare gain from offering multiple service tiers. In Hermalin and Katz, because different tiers are simply assumed to be differentially costly to provide and there is, by implicit assumption, no tradeoff in the quality assigned one content provider with respect to the quality that can be assigned another, one arrives at the answer that the provision of differential quality levels is welfare superior to a single quality level. Hence, in their model, the only reason *not* to have multiple tiers is that it will lead to greater distortions via pricing than would neutrality (a common quality level). In contrast, here we show that, for a fixed amount of bandwidth, a case can be made that welfare is greater under neutrality than under multiple tiers.

The next section of the article presents our model. Because we wish to allow for varying amounts of content consumption, the queuing approach of Choi and Kim and others is not well suited to our purposes. Instead, we model last-mile service as a "pipe" of given bandwidth. Tiering (multiple speeds) is captured by allowing the ISP to allocate portions of the bandwidth to the traffic from different content providers. This is equivalent to the ISP giving priority to the content from different content providers via guarantees about average transmission speeds. Unlike some models, which limit the content providers to making money solely from advertising, here the content providers can earn income by selling content directly to households, from advertising, or from some combination thereof. To avoid complications in the modeelling and welfare analysis, we assume the content providers are not in direct competition with each other; rather, each is a monopoly in its product space. Households place different values on the content of different content providers; critically, they are more sensitive to delay when receiving content from some providers than from others.

The principal section of the article is Section 3, where we analyze the welfare implications of departing from network neutrality. In that section, we show that a sufficient statistic for welfare is the total amount of content traded in equilibrium (Proposition 1). This result is reminiscent of results concerning the welfare implications of third-degree price discrimination, although there is a significant difference insofar as discrimination means differential times rather than differential prices. One implication of Proposition 1 is that welfare can never be enhanced by excluding or blocking some content providers (a practice of which some ISPs have been accused). A second implication of Proposition 1 is that, for tiering to be welfare maximizing, it is necessary that delay times be shorter for content with greater elasticities of demand with respect to transmission time than for content with smaller elasticities. This result is reminiscent of Ramsey pricing, although the mechanics are different because we are considering different times rather than different prices. If the aforementioned elasticities do not vary across the different types of content, then an absence of tiering (i.e, a uniform level of service) is welfare maximizing.

A key point of the article is that content that is more time sensitive, in a sense made precise in Section 2, need not be the content with the greater elasticity of demand with respect to transmission time. This is relevant because the price-discrimination schemes based on differential transmission speeds that residential ISPs have proposed using in their pricing to the content providers must

[5] See, for instance, Duranton and Turner (2011) for evidence concerning physical highways.

© RAND 2012.

provide faster service to more time-sensitive content.[6] If it turns out that such content is not also the content with the greater elasticities with respect to transmission time, then such price discrimination will result in service prioritization (speeds) that are the reverse of what would be welfare maximizing (see Proposition 5 in Section 4). Network neutrality would then be welfare superior to any *implementable* price-discrimination (prioritization) scheme of this sort. Moreover, we show in the same section that mandating neutrality may be necessary: absent regulation, a profit-maximizing ISP can have private incentives to implement prioritization (tiering) even when neutrality is welfare superior.

Driving much of our results is the endogeneity of household consumption decisions. Just as on a physical highway, where cars abandon slow-moving lanes for fast-moving lanes, giving priority to some content over others, will cause households to consume more of the former and less of the latter *ceteris paribus*. It is this ability of households to adjust their consumption that explains, in part, Proposition 1: the fact that households choose how to allocate the greater content that one division of bandwidth provides over another means households will tend to prefer the former division to the latter. An additional force behind Proposition 1 is that content providers' profits are increasing in total content sold.

An issue in the network neutrality debate is whether relaxing the neutrality requirement would provide ISPs greater incentives to increase bandwidth. We consider that issue in Section 5. Unlike earlier work (e.g., Choi and Kim, 2010), which found that allowing an ISP to discriminate had ambiguous effects on its incentives to increase bandwidth, we find that an ability to discriminate unambiguously results in the ISP installing greater bandwidth. This effect is welfare enhancing. Whether it is strong enough to counterbalance the static inefficiency of discrimination is ambiguous: we derive results that suggest that when household utility is a significantly greater component of welfare than content providers' profits, then network neutrality can still be the welfare-superior policy even accounting for the ISP's bandwidth-building incentives.

Like any model, some assumptions are necessary to make the analysis tractable. We explore relaxing these assumptions in Section 6. We show there that our primary results can continue to hold if we allow the ISP to unit-price content; if households are heterogeneous in their preferences or if different households consume different content entirely; if content providers vary in terms of their marginal costs and advertising rates; and when we consider a variety of different assumptions for the household utility function. In particular, via an example, we show that our main finding—the possibility that network neutrality is welfare maximizing among the *feasible* (implementable) tiering schemes (Proposition 5)—can still hold given a very different demand structure.

2. Model

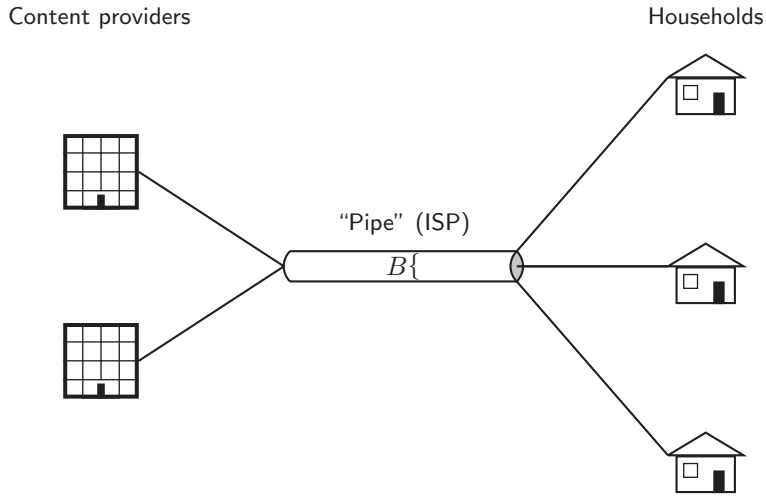
□ **Structure and technology.** Figure 1 shows the basic structure and technology we have in mind. Households want to engage with content (or application) providers. These providers' content must pass through a "pipe" controlled by a monopoly, the ISP, to reach households. The pipe has a bandwidth, B. This should be interpreted as there being the capacity for B "units" of content (e.g., packets) to go from the content providers to the households per unit of time. We assume that the ISP can dedicate portions of the bandwidth or otherwise guarantee priority to different content providers or groups of providers; that is, it can divide the bandwidth into

[6] In theory, forms of price discrimination besides second-degree price discrimination via prioritization exist. In particular, it has been suggested to us that ISPs could engage in perfect discrimination vis-à-vis content providers because ISPs can identify what content comes from which content providers; that is, an ISP should charge an access fee to each content provider exactly equal to that provider's profit from transacting with the ISP's residential customers (e.g., Verizon should charge Amazon an access fee equal to 100% of all Amazon's profits earned transacting with Verizon's residential customers). We discuss the possibility of such perfect discrimination in greater depth in Section 4; we note now, however, that there are reasons to doubt such discrimination is truly feasible or permissible. Moreover, to the best of our knowledge, no ISP has ever sought permission to engage in such firm-specific pricing.

© RAND 2012.

FIGURE 1

SCHEMATIC REPRESENTATION OF TECHNOLOGY



$B_1, \ldots, B_J$ subbandwidths, where $\sum_{j=1}^{J} B_j = B$ (or do the equivalent thereof via the granting of priority).[7]

We assume a continuum of content providers of measure one, indexed by $\theta \in [\underline{\theta}, \bar{\theta}) \subseteq \mathbb{R}_+$. The distribution of θ is $F : [\underline{\theta}, \bar{\theta}) \to [0, 1]$. Assume the derivative $F'(\cdot)$ exists and is positive for all $\theta \in (\underline{\theta}, \bar{\theta})$. As a slight abuse of notation, we will sometimes write $F(\Theta)$ to denote the proportion of application-provider types that are in set Θ.

Let $X(\theta)$ denote the units of content sent by content provider θ. If $\Theta \subset [\underline{\theta}, \bar{\theta})$ is a measurable subset of content providers with dedicated bandwidth B_Θ, then

$$t(\Theta) \equiv \frac{\int_\Theta X(\theta) dF(\theta)}{B_\Theta}$$

is the time necessary to send all of the content of those content providers in Θ. Observe $t(\Theta)$ is a measure of the congestion faced by content providers in Θ, and we will treat it as such in what follows. We use $\tau(\theta)$ to denote the delivery time of content provider θ's content. If θ is in Θ, then $\tau(\theta) = t(\Theta)$.

□ **Consumers and content providers.** Assume there is a continuum of households (consumers) of measure one. Each household potentially engages in trade with each content provider. We assume a household's utility is quasilinear and additively separable over the content from different providers. A household's *marginal* utility from the xth unit of content from content provider θ is taken to be

$$m\left(\frac{x}{\alpha(\tau(\theta), \theta)}\right), \tag{1}$$

where the adjustment factor, $\alpha(\tau(\theta), \theta)$, reflects the congestion in transmission, $\tau(\theta)$, some indication of the value the household assigns that content, and how much the household cares

[7] In particular, this formulation is equivalent to one in which an ISP guarantees average transmission speeds. To see this, a content provider's *average* speed equals content ÷ total time = content ÷ (content ÷ allocated bandwidth) = allocated bandwidth.

© RAND 2012.

about delay or congestion vis-à-vis that content.[8] A household's overall utility is

$$\mathcal{U} = y + \int_{\underline{\theta}}^{\bar{\theta}} \left(\int_0^{x(\theta)} m\left(\frac{x}{\alpha(\tau(\theta), \theta)} \right) dx \right) dF(\theta), \tag{2}$$

where y is the numéraire good and $x(\theta)$ is its consumption of the θth provider's good (number of packets bought).

Specification (2) embeds a number of assumptions, which we discuss in depth in Section 6. As we demonstrate there, the assumption of additive separability across different content, a common assumption in the literature, can be justified as a reasonable approximation to more general utility functions. Another embedded assumption is that households are homogeneous in their preferences. This assumption is made primarily for convenience; as we show in Section 6, our principal results—Propositions 1–5—continue to hold even when households have heterogeneous preferences (including wishing to consume different types of content).

We assume marginal utility, $m(\cdot)$, is twice differentiable and decreasing. We assume, further, that marginal utility is never "too convex"; specifically, we assume

$$zm''(z) + m'(z) < 0 \tag{3}$$

for all $z \in_+$. We assume that households prefer faster content delivery to slower content delivery *ceteris paribus*; that is, $\tau > \tau'$ implies $\alpha(\tau, \theta) < \alpha(\tau', \theta)$ for all θ. We further assume that households view content from higher-θ content providers to be more time sensitive than from lower-θ content providers; that is, for all $\theta > \theta'$ and all $\tau > \tau'$,

$$\alpha(\tau, \theta) - \alpha(\tau', \theta) < \alpha(\tau, \theta') - \alpha(\tau', \theta') \tag{4}$$

(in other words, α exhibits decreasing differences). At this juncture, we make no assumptions about how α varies with content type, θ, holding time constant.

We assume that consumption of the content providers' goods plus any hookup fee paid the ISP never consumes a household's entire income. This and the assumption of additively separable and quasilinear utility mean that each household acquires the amount of the θth content provider's product that equates marginal utility to marginal cost (i.e., price, p); consequently, household demand is

$$x(p, \theta) = \alpha(\tau(\theta), \theta) m^{-1}(p) \equiv \alpha(\tau(\theta), \theta)\omega(p). \tag{5}$$

(Note the implicit definition of $\omega : \mathbb{R}_+ \to \mathbb{R}_+$.) Observe that demand for content falls with congestion. Expression (5) plays a key role in the analysis that follows, so it is worth pausing to interpret it. This expression embodies the highly plausible assumption that demand for any particular type of content increases if that content is delivered more quickly, where the rate of increase in demand depends on what the content is (i.e., the effect on the demands for content θ' and θ'' from the same change in delivery time may vary). An issue, however, is how this demand effect should be modelled. Here, it is, essentially, a counterclockwise rotation of demand. Alternatively, it could, say, be modelled as a parallel shift out in demand. Our assumption of rotational (multiplicative) change makes the ensuing analysis highly tractable, albeit at the loss of some generality. In Section 6 we show, via a particular example, that our main result, Proposition 5, is still applicable when the effect of faster delivery is a parallel shift out in demand rather than a rotation.

A content provider's profit is

$$(q + p - c)x(p, \theta) - s\,,$$

[8] We assume this form for marginal utility, rather than the more general $m(x, \tau(\theta), \theta)$, because it is particularly tractable. In Section 6, we show, via an example, that assuming (1) is not necessary for establishing that network neutrality can be superior to any feasible discrimination scheme.

© RAND 2012.

where q is the advertising rate, c is the marginal cost of content production and transmission, and s is a payment to the ISP. Observe that we are assuming that the advertising rate and content cost are common across content providers.[9]

Assumption (3) implies $\omega(\cdot)$ is log concave. To rule out infinite consumption, assume that $\lim_{p\downarrow c-q} \omega(p) < \infty$.[10] These assumptions are thus sufficient for a content provider's pricing problem,

$$\max_p (q + p - c)x(p, \theta) - s\,, \tag{6}$$

to have a unique and finite solution. Observe that solving (6) is equivalent to solving

$$\max_p (q + p - c)\omega(p);$$

hence, the solution to (6) is independent of θ. Call that solution p^*.[11] For future reference, define

$$\pi = (q + p^* - c)\omega(p^*);$$

hence, content provider θ's maximized profit is $\alpha(\tau(\theta), \theta)\pi - s$. We refer to π as the "equilibrium gross profit factor."

Household (consumer) surplus from trade with content provider θ is

$$\int_{p^*}^{\infty} x(p, \theta)dp = \alpha(\tau(\theta), \theta) \int_{p^*}^{\infty} \omega(p)dp \equiv \alpha(\tau(\theta), \theta)\sigma\,.$$

(Note the implicit definition of σ—the "equilibrium consumer surplus factor.")

Total welfare, the sum of content providers' profits and households' consumer surplus, is

$$W = \int_{\underline{\theta}}^{\bar{\theta}} (\pi + \sigma)\alpha(\tau(\theta), \theta)dF(\theta). \tag{7}$$

3. Welfare

■ Let $\Theta_1, \ldots, \Theta_N$ denote a partition of $[\underline{\theta}, \bar{\theta})$ such that each Θ_n is measurable with respect to F. Let B_n denote the bandwidth allocated to Θ_n ($\sum_{n=1}^{N} B_n = B$). As a shorthand, let $t_n = t(\Theta_n)$. An identity that must hold in equilibrium is

$$t_n = \frac{1}{B_n} \int_{\Theta_n} x(p^*, \theta)dF(\theta) = \frac{1}{B_n}\omega(p^*) \int_{\Theta_n} \alpha(t_n, \theta)dF(\theta). \tag{8}$$

Observe that $\sum_{n=1}^{N} B_n t_n$ equals the total amount of content, X, sent in equilibrium given this division of the bandwidth.

Consider two partitions, $\Theta_1, \ldots, \Theta_N$ and $\widetilde{\Theta}_1, \ldots, \widetilde{\Theta}_M$, and corresponding bandwidth allocations $B_1, \ldots, B_N$ and $\tilde{B}_1, \ldots, \tilde{B}_M$, respectively. Denote $t(\Theta_n)$ by t_n and $t(\widetilde{\Theta}_m)$ by $\tilde{t}_m$.

[9] Such an assumption is a reasonable first-order approximation because there is a competitive market for advertising and the contents' marginal costs are often the same, zero. Like any approximation, however, it is imperfect and there are exceptions (e.g., video or audio resellers who must pay significant royalties per item). We explore this further in Section 6, showing that the article's primary results hold even if $q - c$ varies across content providers provided the variation is independent of the household preference parameter θ.

[10] An alternative assumption would be to require $\lim_{p\downarrow 0} \omega(p) < \infty$ and to assume $p \geq 0$; the latter condition matches reality insofar as practical issues generally preclude websites, even those that make money through advertising, from paying their visitors.

[11] That each content provider's price is independent of its type is both a positive and negative from a modelling perspective: positive insofar as it allows us to abstract away from the effects that changes in tiering might cause in content providers' pricing; and negative insofar as a common price is unrealistic. In Section 6, we show that a variation of the model with different provider prices can be entertained without changing our results.

© RAND 2012.

Equilibrium welfare under these two bandwidth allocations is thus

$$W = \sum_{n=1}^{N}(\pi+\sigma)\int_{\Theta_n}\alpha(t_n,\theta)dF(\theta) = \frac{\pi+\sigma}{\omega(p^*)}\sum_{n=1}^{N}B_n t_n = \frac{\pi+\sigma}{\omega(p^*)}X \tag{9}$$

and

$$\widetilde{W} = \sum_{m=1}^{M}(\pi+\sigma)\int_{\widetilde{\Theta}_m}\alpha(\tilde{t}_m,\theta)dF(\theta) = \frac{\pi+\sigma}{\omega(p^*)}\sum_{m=1}^{M}\tilde{B}_m\tilde{t}_m = \frac{\pi+\sigma}{\omega(p^*)}\widetilde{X}. \tag{10}$$

Comparing (9) to (10), the following result is immediate:

Proposition 1. Given two alternative divisions of the total bandwidth, one is welfare superior to the other if and only if it results in more content being carried in equilibrium than the other.

An obvious, but important, corollary is the following.

Corollary 1. Network neutrality is welfare superior (at least weakly) to any division of the bandwidth if and only if no division of the bandwidth leads to more content being sent in equilibrium.

Proposition 1 and its corollary are reminiscent of the result that a necessary condition for third-degree price discrimination to increase welfare vis-à-vis uniform pricing is that price discrimination increase the total amount sold vis-à-vis the amount sold under uniform pricing (see, e.g., Varian, 1985, 1989). Our results differ from the results concerning third-degree price discrimination because here purchasers face different transmission times rather than, as under third-degree price discrimination, different prices. There is a certain analogy between transmission times and prices insofar as utility is decreasing in both. However, unlike in the analysis of third-degree price discrimination, there is no general fundamental identity analogous, say, to Roy's identity, which would link transmission speed and welfare. Consequently, our analysis cannot be as general, in terms of utility functions, as Varian's.

It might at first seem odd that total content sent is a sufficient statistic for deciding among different bandwidth allocations and that *how* the bandwidth is allocated is not an explicit part of Proposition 1. To better understand this result, observe that households adjust their consumption to achieve the same marginal utility across the content of the different content providers in equilibrium. Hence, for any two types of content, θ and θ', we have in equilibrium

$$\omega(p^*) = \frac{x(p^*,\theta)}{\alpha(\tau(\theta),\theta)} = \frac{x(p^*,\theta')}{\alpha(\tau(\theta'),\theta')}. \tag{11}$$

Observe that

$$\int_0^{x(\theta)} m\left(\frac{x}{\alpha(\tau,\theta)}\right)dx = u\left(\frac{x(\theta)}{\alpha(\tau,\theta)}\right)\alpha(\tau,\theta),$$

where $u(z) = \int_0^z m(x)dx$. Given (11), it follows that welfare is

$$W = \int_{\underline{\theta}}^{\bar{\theta}}\left(u(\omega(p^*))\frac{x(\theta)}{\omega(p^*)} + (q-c)x(\theta)\right)dF(\theta) \propto \int_{\underline{\theta}}^{\bar{\theta}}x(\theta)dF(\theta) = X;$$

that is, welfare is in fixed proportion to the total amount of content sent.

Yet another interpretation is the following. Households are free to adjust their consumption. From the perspective of the households, more total content is similar to more total income in a conventional consumer-choice model: more income is better than less because a consumer is free to allocate that income as she sees fit. Similarly, more content will tend to make households better off because they can allocate it as they see fit. Additionally, content providers' gross profits

© RAND 2012.

increase in the amount they sell by a constant factor. It is not surprising then that welfare increases in the amount of content carried.

An ongoing policy debate is whether residential ISPs should be permitted to block certain traffic. For example, it has been alleged that the American ISP Comcast blocked BitTorrent.[12] As a rationale for actions that might hinder BitTorrent, Comcast cited congestion alleviation.[13] It is, therefore, worth considering whether welfare can be improved by blocking some content while providing remaining content a common level of service.

Proposition 2. Suppose the ISP provides a common class of service but excludes a positive measure of content providers. In the resulting equilibrium, welfare is less than it would be were no content providers excluded.

Proof. Let the excluded content belong to Θ_{ex} and the nonexcluded content belong to Θ_{in}, $\Theta_{ex} \cup \Theta_{in} = [\underline{\theta}, \bar{\theta})$. By assumption, $F(\Theta_{ex}) > 0$.

Suppose, contrary to the proposition, that welfare given the exclusion of Θ_{ex} was not less than it would be under network neutrality (i.e., no exclusion). It follows from Proposition 1 that if $\hat{X}$ is total content carried in equilibrium given exclusion and X is total content carried in equilibrium under neutrality, then

$$\hat{X} \geq X. \tag{12}$$

Let $\hat{t}$ and t^* be the equilibrium transmission times under exclusion and neutrality, respectively. Given that bandwidth is fixed, it follows that

$$\hat{t} = \frac{\hat{X}}{B} \geq \frac{X}{B} = t^*.$$

Hence, $\alpha(\hat{t}, \theta) \leq \alpha(t^*, \theta)$ for all θ. This yields

$$\hat{X} = \omega(p^*) \int_{\Theta_{in}} \alpha(\hat{t}, \theta) dF(\theta) \leq \omega(p^*) \int_{\Theta_{in}} \alpha(t^*, \theta) dF(\theta) < \omega(p^*) \int_{\underline{\theta}}^{\bar{\theta}} \alpha(t^*, \theta) dF(\theta) = X. \tag{13}$$

But (13) contradicts (12). The result follows *reductio ad absurdum*. *Q.E.D.*

Intuitively, the only way for welfare to increase is if more content is traded. But trading more content necessarily increases a *common* transmission time, which reduces demand. Hence, it is impossible to increase welfare by excluding some content: the excluded content is no longer traded and, as just shown, the volume of the nonexcluded content traded in equilibrium cannot exceed the total volume traded under neutrality.[14]

The next issue is deriving conditions that allow us to determine whether dividing the bandwidth (tiering service) increases or decreases total content sent. To that end, assume $\alpha(\cdot, \cdot)$ is twice differentiable in both arguments.[15] Observe that the bandwidth used by content θ is $x(\theta)/\tau(\theta) \times dF(\theta)$. The overall bandwidth constraint can, thus, be written as

$$\int_{\underline{\theta}}^{\bar{\theta}} \frac{x(\theta)}{\tau(\theta)} dF(\theta) = B. \tag{14}$$

[12] See "Comcast Blocks Some Internet Traffic," an Associated Press article by Peter Svensson, October 19, 2007. Accessed on the MSNBC website.

[13] Source: see footnote 12.

[14] The reader may question this conclusion on the grounds that we have assumed homogeneous households, worrying that the use of, say, BitTorrent represents an externality imposed by certain types of households on other types of households. As we show in Section , this concern is misplaced: Proposition holds even if we extend the analysis to account for heterogeneous households (see, in particular, Proposition 15).

[15] At the end of this section, we present an alternative analysis that does not rely on calculus. See Proposition 4.

© RAND 2012.

When households can adjust consumption, welfare is given by (7) above and, because $x(\theta)$ then equals $\alpha(\tau(\theta), \theta)\omega(p^*)$, the bandwidth constraint (14) can be written as

$$\omega(p^*)\int_{\underline{\theta}}^{\bar{\theta}} \frac{\alpha(\tau(\theta), \theta)}{\tau(\theta)} dF(\theta) = B\,. \tag{14'}$$

Let λ denote the Lagrange multiplier on the constraint. The Lagrangean associated with constrained welfare maximization is

$$\int_{\underline{\theta}}^{\bar{\theta}} \alpha(\tau(\theta), \theta)\left(\pi + \sigma - \lambda\omega(p^*)\frac{1}{\tau(\theta)}\right) dF(\theta). \tag{15}$$

Observe one maximizes (15) pointwise; that is, to maximize (15) one maximizes, for each θ,

$$\alpha(\tau(\theta), \theta)\underbrace{\left(\pi + \sigma - \lambda\omega(p^*)\frac{1}{\tau(\theta)}\right)}_{Z}$$

with respect to $\tau(\theta)$. The term labelled Z must be positive because otherwise the expression would be maximized by allocating infinite time, but allocating infinite time guarantees Z is positive; *reductio ad absurdum*, $Z > 0$. Maximization programs are invariant with respect to increasing transformations, so $\tau(\theta)$ must maximize

$$\log(\alpha(\tau(\theta), \theta)) + \log\left(\pi + \sigma - \lambda\omega(p^*)\frac{1}{\tau(\theta)}\right). \tag{16}$$

Assuming (16) has an interior maximum, it follows from well-known comparative-statics results that the $\tau(\theta)$ that maximizes (16) is increasing in θ if the cross-partial derivative of (16) with respect to θ and $\tau(\theta)$ is positive; it is decreasing in θ if that cross-partial derivative is negative; and it is constant if that cross-partial is everywhere zero. The sign of the cross-partial derivative of (16) is the same as that of the cross-partial derivative of

$$\log(\alpha(\tau(\theta), \theta)). \tag{17}$$

The sign of the cross-partial derivative of (17) is a statement about how the elasticity of demand for content with respect to transmission time varies across content. That elasticity is given by

$$\varepsilon(\tau, \theta) \equiv -\frac{\partial \log(\alpha(\tau, \theta)\omega(p^*))}{\partial \log(\tau)} = -\frac{\partial \log(\alpha(\tau, \theta))}{\partial \log(\tau)} = -\tau\frac{\partial \log(\alpha(\tau, \theta))}{\partial \tau} > 0\,.$$

This establishes that the crucial determinant of how content should be prioritized (if at all) depends on how transmission-time elastic demand is for different content. Formally, we have the following.

Proposition 3. Suppose the elasticity of demand for content with respect to transmission time is monotone in content type holding time constant (i.e., $\partial\varepsilon(\tau, \theta)/\partial\theta$ has a common sign for all τ and θ). Then if, for all times, the elasticity of demand for content θ with respect to transmission time exceeds that of content θ' (i.e., $\varepsilon(\tau, \theta) > \varepsilon(\tau, \theta')$ all τ), a welfare maximizing allocation of bandwidth across content providers is such that the equilibrium transmission time for θ content is shorter than for θ' content (i.e., $\tau(\theta) < \tau(\theta')$). If the elasticity of demand with respect to transmission time is invariant with respect to content type, then network neutrality is welfare maximizing.

As intuition, suppose, initially, that content θ and θ' had the same delivery times. Suppose, too, that demand for content θ is more transmission-time elastic than for content θ'; that is, the equilibrium consumption of content θ is more affected by a change in delivery time than is the equilibrium consumption of content θ'. Then shifting bandwidth to θ from θ'—that is, reducing the delivery time for θ but increasing it for θ'—would increase total content traded

© RAND 2012.

because the increase in trade of θ would more than offset the reduction in the trade of θ'. Such a shift would be welfare enhancing because total content traded is a sufficient statistic for welfare (Proposition 1). The result, Proposition 3, follows. In many ways, this result is reminiscent of optimal Ramsey pricing (except, here, different users face different times, not prices).

It is important to recognize that the assumption that higher-θ content is more time sensitive (i.e., that $\frac{\partial^2 \alpha}{\partial \theta \partial \tau} < 0$) does *not* imply demand for that content is more elastic with respect to time. For example, consider

$$\alpha(\tau, \theta) = \theta \exp\left(-\frac{\tau}{\theta}\right) \Rightarrow \frac{\partial^2 \alpha}{\partial \theta \partial \tau} = -\frac{\tau \exp(-\tau/\theta)}{\theta^2} < 0, \text{ but } \frac{\partial}{\partial \theta}\left(-\frac{\partial \log(\alpha)}{\partial \tau}\tau\right) = \frac{-\tau}{\theta^2} < 0.$$

In other words, *it is possible that welfare can be maximized by retarding more time-sensitive content in favor of less time-sensitive content.*

This last possibility may, at first, seem perverse. It is, however, a possibility because households adjust their consumption. Hence, as noted, what matters is the time elasticity of demand. Given that need not coincide with time sensitivity, as defined by (4), there is no reason why welfare maximization requires speeding the delivery of more time-sensitive content at the expense of less time-sensitive content.

An alternative analysis, which does not rely on calculus, is feasible when the identity (8) can be solved for t_n.[16] To that end, suppose that α can be decomposed as

$$\alpha(\tau, \theta) = \gamma(\tau)v(\theta), \tag{18}$$

where, consistent with earlier assumptions, $\gamma(\cdot) > 0$, $v(\cdot) > 0$, $\gamma(\cdot)$ decreasing, and $v(\cdot)$ increasing. Assume $v(\cdot)$ is measurable and $\gamma(\cdot)$ is continuous on $\mathbb{R}_+$. Define $g(t) = t/\gamma(t)$. Because $\gamma(\cdot)$ is decreasing, $g(\cdot)$ is a strictly increasing function. It follows that $g(\cdot)$ has an inverse, $g^{-1}(\cdot)$, and that this inverse is also an increasing function. Given that $g(t)$ is formed by multiplying t by an increasing positive function, it is not unreasonable to assume further that $g(\cdot)$ is convex.[17] If $g(\cdot)$ is convex, it follows that $g^{-1}(\cdot)$ is concave. The solution to (8) is

$$t_n = g^{-1}\left(\frac{\omega(p^*)}{B_n}\int_{\Theta_n} v(\theta)dF(\theta)\right). \tag{19}$$

Given neutrality (no bandwidth division), the equilibrium time to send all content in equilibrium is

$$t^* = g^{-1}\left(\frac{\omega(p^*)}{B}\int_{\underline{\theta}}^{\bar{\theta}} v(\theta)dF(\theta)\right). \tag{20}$$

We can now establish the following.

Proposition 4. Suppose the adjustment function, $\alpha(\tau, \theta)$, is multiplicatively separable in delivery time and type (i.e., is given by (18)). Suppose too that the function defined by $t \mapsto t/\gamma(t)$ is a convex function. Then, network neutrality is welfare superior to any division of the bandwidth and strictly welfare superior to any division that either excludes some measurable segment of content providers or results in different groups of content providers having their content providers having their content being sent with different transmission times in equilibrium.

The proof, which uses Jensen's inequality to show that total content under neutrality exceeds total content under any division, can be found in the Appendix. Note that if $v(\cdot)$ and $\gamma(\cdot)$ are differentiable, then (18) implies the elasticity of demand with respect to time is constant across content types; that is, Proposition 4 is consistent with Proposition 3.

[16] Observe the leftmost term in (8) is linearly increasing in t_n, whereas the rightmost term is decreasing in t_n; hence, if (8) has a solution in t_n, it is unique.

[17] Although $x\psi(x)$ is not necessarily convex on $\mathbb{R}_+$ for all increasing functions $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ (e.g., $x(1 + (x-1)^3)$ is not convex), $x\psi(x)$ will be convex for a large set of $\psi(\cdot)$ including $\psi(x) = x^{1/y}$, $y > 1$, and $\psi(x) = \log(x+1)$. Of course, $x\psi(x)$ is convex if $\psi(\cdot)$ is convex.

© RAND 2012.

4. Residential ISPs charging content providers

■ In this section, we focus on the incentives of the residential ISP to charge content providers for carrying their content the "last mile" to households (the ISP's subscribers). In particular, we wish to compare the ISP's private incentives with what would be welfare maximizing.

A possible pricing regime would be one in which the ISP sets a common fee, s, to content providers and provides those who paid it a common level of service. If a positive measure of content providers chose not to pay, then welfare would necessarily be reduced by Proposition 2 To summarize, we have the following.

Corollary 2. Suppose the ISP provides a common class of service but charges any content provider that wishes access to its subscribers (households) an access fee $s > 0$. If, in the resulting equilibrium, a set of content providers of positive measure chooses not to purchase access, then welfare would be increased by prohibiting the ISP from charging such an access fee.

On one hand, Corollary 2 might not seem surprising: after all, it is well-known that monopoly pricing that excludes buyers willing to pay more than a good's marginal cost is welfare reducing. On the other hand, although the ISP's monetary cost is zero, there is a *social* marginal cost due to congestion. As is well-known, pricing to exclude low-value uses can be a way to raise welfare in the face of congestion. In this light, Corollary 2 is not obvious. The result holds because households already internalize congestion in making their purchase decisions and, hence, it is impossible to create a benefit vis-à-vis congestion alleviation by excluding content providers.

In the network neutrality debate, an issue has been whether more sophisticated pricing by residential ISPs to the content providers should be permitted. Central to this debate has been second-degree price discrimination via prioritized service; that is, differential prices for differential classes (tiers) of service.[18] We now turn to such pricing.

An immediate consequence of our definition of time sensitivity, expression (4), is the following result.

Lemma 1. Suppose there are two tiers of service with corresponding equilibrium transmission times t and t', $t > t'$. Suppose the ISP charges content providers s and s', respectively, for the different tiers. Consider two content-provider types θ and θ', $\theta > \theta'$. Then if the low-type θ' prefers the faster tier t' at price s' (i.e., $\pi\alpha(t', \theta') - s' \geq \pi\alpha(t, \theta') - s$), the higher-type θ strictly prefers the faster tier t' at price s' (i.e., $\pi\alpha(t', \theta) - s' > \pi\alpha(t, \theta) - s$).

A reasonable assumption is that there is no demand for content that will never be delivered (equivalently, that has been allocated no bandwidth). We thus maintain the following assumption henceforth.

Assumption 1. For all content types θ, $\lim_{\tau\to\infty} \alpha(\tau, \theta) = 0$.

Given this assumption, a content provider that chooses not to purchase access to an ISP's subscribers can be considered to have "purchased" infinite delay at a price of zero. This insight and Lemma 1 yield the following result.

Corollary 3. If a lower type of content provider is willing to pay for access to an ISP's subscribers (i.e., buy a tier of service with finite transmission time), then a higher type strictly prefers to pay for access.

In other words, if pricing by the ISP causes some content providers to exit, it will be the lowest-type content providers that exit.

[18] At the end of this section, we briefly discuss other forms of price discrimination.

© RAND 2012.

A well-known consequence of Lemma 1 and Corollary 3 is that any feasible (i.e., incentive-compatible) price-discrimination scheme that the ISP can employ vis-à-vis the APs must have the feature that, in equilibrium, quality (i.e., speed) is everywhere nondecreasing in type; hence, $\tau(\cdot)$ is necessarily a nonincreasing function.[19] Such an allocation of transmission times cannot be better—*but could be worse*—than network neutrality if Proposition 4 applies or if the time elasticity of demand is nonincreasing in content type (Proposition 3). When conditions satisfy Proposition 4, this conclusion follows because network neutrality is welfare maximizing. When the time elasticity of demand is decreasing with content type, this conclusion follows because a feasible scheme (in which $\tau(\cdot)$ is nonincreasing) goes in the opposite direction relative to the welfare maximizing scheme (in which $\tau(\cdot)$ is increasing); that is, network neutrality is closer to the welfare maximizing scheme than would be a feasible discrimination scheme. To summarize, we have the following.

Proposition 5. Let the adjustment function α be twice differentiable in its arguments. Suppose the elasticity of demand with respect to transmission time is nonincreasing in content type, holding time constant (i.e., $\partial\varepsilon(\tau,\theta)/\partial\theta \leq 0$ for all τ and θ). Then, any feasible price-discrimination (tiered-service) scheme that the ISP can implement vis-à-vis the APs the content providers will be welfare inferior to network neutrality.

Alternatively, let the adjustment function α satisfy the assumptions of Proposition 4. Then any feasible price-discrimination scheme that the ISP can implement vis-à-vis the content providers will be welfare inferior to network neutrality.

It is, however, important to note that if the time elasticity of demand is increasing in content type (i.e., $\partial\varepsilon(\tau,\theta)/\partial\theta > 0$), then there is an incentive-compatible prioritization scheme that is welfare superior to network neutrality. On the other hand, as a general rule, the welfare-maximizing scheme will not maximize the ISP's profit; hence, the answer to the question "Will the ISP adopt a price-discrimination scheme that is welfare superior to network neutrality when the time elasticity of demand is increasing in content type?" is it depends: one cannot guarantee that discrimination will be welfare superior because it is possible that the ISP could "overdo it."

Knowledge that a pricing scheme could be welfare inferior to neutrality is meaningful only if the ISP would pursue such pricing if not otherwise restrained. Given that one cannot be sure that a monopolist will price in a less-than-welfare maximizing manner even in the simplest of monopoly settings,[20] it cannot be that *always* restraining the ISP is welfare improving. In what follows, therefore, we focus on a few cases that demonstrate the possibility that the ISP would impose non-neutral schemes in situations in which neutrality would be welfare superior, that is, cases in which restraining the ISP would be welfare improving.

In considering the ISP's pricing to the content providers, we need also to be mindful that this is a two-sided market and, hence, be cognizant of the effect of the ISP's pricing to content providers on the prices it can charge households. The assumption that households are homogeneous means that the ISP can fully capture a household's equilibrium net benefit from trade with content providers. Hence, the hookup fee, η, will equal

$$\sigma \int_{\underline{\theta}}^{\bar{\theta}} \alpha(\tau(\theta),\theta) dF(\theta). \tag{21}$$

[19] See, for example, Tirole (1988, chapter 3) for why quality must be nondecreasing in type under incentive-compatible second-degree price discrimination.

[20] Consider a monopoly with a constant marginal cost normalized to zero, which faces demand

$$D(p) = \begin{cases} \max\{0, 2-p\}, & \text{if } p > 1 \\ 1, & \text{if } p \leq 1 \end{cases}.$$

(as would arise if a given consumer wanted at most one unit and the distribution of consumers' willingnesses to pay was uniform on [1, 2]). It is readily seen that the profit-maximizing price is $p = 1$, a price at which welfare is a maximum.

© RAND 2012.

Recalling the expression for total welfare (7), it follows from (21) that

$$\eta = \frac{\sigma}{\pi + \sigma} W, \tag{22}$$

where W is equilibrium welfare. The hookup fee is thus a constant proportion of equilibrium total welfare. The ISP will thus internalize, to an extent, overall welfare.

An immediate consequence of this analysis of pricing to households is the following.

Proposition 6. Suppose the ISP's pricing scheme to content providers is not welfare maximizing; then, the imposition of restrictions on the ISP vis-à-vis its pricing to content providers that raise, welfare will lead to higher hookup fees for households.

This result has a possible political-economy implication: the ISP can credibly promise a lower household hookup fee should restrictions on charging content providers be lifted, which could create political pressure to lift those restrictions. Of course, this argument presumes a certain naïveté on the part of households because, in equilibrium, they capture no surplus and should, thus, be indifferent as to how residential ISPs' pricing to content providers is regulated. That noted, one could view Proposition 6 as a shorthand for an analysis in which household heterogeneity means some households would strictly benefit from restrictions being lifted because they would no longer be priced out of Internet access or would see their surplus increase.

We return now to the question of how, given the flexibility, a residential ISP would price-discriminate to content providers. If the ISP can discriminate continuously across content providers, then, following well-known techniques, it will impose the scheme that maximizes expected virtual surplus subject to (i) $\tau(\cdot)$'s being nonincreasing and (ii) the bandwidth constraint (14′). In terms of notation, it seeks to maximize

$$\underbrace{\sigma \int_{\underline{\theta}}^{\bar{\theta}} \alpha(\tau(\theta), \theta) f(\theta) d\theta}_{\eta} + \int_{\underline{\theta}}^{\bar{\theta}} \left(\pi \alpha(\tau(\theta), \theta) - \pi \frac{1 - F(\theta)}{f(\theta)} \frac{\partial \alpha(\tau(\theta), \theta)}{\partial \theta} \right) f(\theta) d\theta,$$

where $f(\cdot)$ is the density function associated with the distribution function $F(\cdot)$, subject to $\tau(\cdot)$'s being nonincreasing and the bandwidth constraint. Equivalently, we can express this program as maximizing the Lagrangean

$$\int_{\underline{\theta}}^{\bar{\theta}} \left((\sigma + \pi - \frac{\lambda \omega(p^*)}{\tau(\theta)}) \alpha(\tau(\theta), \theta) - \pi \frac{1 - F(\theta)}{f(\theta)} \frac{\partial \alpha(\tau(\theta), \theta)}{\partial \theta} \right) f(\theta) d\theta \tag{23}$$

subject to $\tau(\cdot)$'s being nonincreasing. It is straightforward to see that maximizing (23) will not, in general, lead to neutrality.

For example, if $\alpha(\tau, \theta) = \theta/\tau$, which satisfies Proposition 4, and $F(\cdot)$ is the uniform distribution on [0, 1], then maximizing (23) yields the solution[21]

$$\tau(\theta) = \begin{cases} \infty \text{ (exclusion)}, & \text{if } \theta \leq \dfrac{\pi}{\sigma + 2\pi} \\ \dfrac{\theta \sqrt{\omega(p^*)} \sqrt{\sigma^2 - \pi^2 - 2\pi^2 \log(\pi) + 2\pi^2 \log(\sigma + 2\pi)}}{\sqrt{2B}(\pi(2\theta - 1) + \theta\sigma)}, & \text{if } \theta > \dfrac{\pi}{\sigma + 2\pi} \end{cases} \tag{24}$$

Clearly, in this case, what is profit maximizing for the ISP, specifically departing from neutrality, is not welfare maximizing.

In the current network neutrality debate, much of the discussion concerns a finite number of tiers (as opposed to "continuous" discrimination). Even when limited in this way, a residential ISP can still have an incentive to discriminate in a welfare-reducing manner. To demonstrate this, suppose the ISP can offer two services: high-speed, denoted by an h subscript, and low-speed,

[21] The Mathematica program used to derive this solution is available from the authors upon request.

© RAND 2012.

denoted by an ℓ subscript. Let r (for rate) denote an arbitrary element of $\{\ell, h\}$. The quantity t_r is the time required, in equilibrium, to send all content of those content providers that subscribe to service r. In keeping with their names, $t_h < t_\ell$. Let s_r be the access fee that the ISP charges content providers for service r. Because, otherwise, there would effectively be only one service, $s_h > s_\ell$.

In deciding which, if any, service to purchase, a content provider chooses the largest of

$$0\,,\ \alpha(t_\ell, \theta)\pi - s_\ell\,,\ \text{ or } \alpha(t_h, \theta)\pi - s_h\,.$$

Let 0 denote "no service" and consider the services ordered $h \succ \ell \succ 0$. The following result follows immediately from Lemma 1 and Assumption 1 and the proof is, therefore, omitted.

Lemma 2. Let $\theta > \theta'$ and $r \succ r'$. Then a θ'-type content provider's preferring r to r' implies that a θ-type content provider prefers r to r'. Conversely, a θ-type content provider's preferring r' to r implies that a θ'-type content provider prefers r' to r.

If there is discrimination in equilibrium, then the previous lemma implies that there are two cutoff types, θ_ℓ and θ_h, with $\underline{\theta} \leq \theta_\ell < \theta_h < \bar{\theta}$, such that types $\theta \geq \theta_h$ purchase the h service, types $\theta \in [\theta_\ell, \theta_h)$ purchase the ℓ service, and types $\theta < \theta_\ell$ purchase no service. Observe that

$$s_\ell = \alpha(t_\ell, \theta_\ell)\pi \ \text{ and } \ s_h = (\alpha(t_h, \theta_h) - \alpha(t_\ell, \theta_h))\pi + s_\ell = (\alpha(t_h, \theta_h) - \alpha(t_\ell, \theta_h) + \alpha(t_\ell, \theta_\ell))\pi\,. \tag{25}$$

Because the analysis is greatly facilitated if we can solve the identity (8) for t_r, suppose that $\alpha(\tau, \theta)$ is multiplicatively separable as specified in (18). As before, $v(\cdot)$ is increasing and $\gamma(\cdot)$ is decreasing. Assume both functions are differentiable. To keep the notation manageable, define

$$\mathcal{I}(\theta_1, \theta_2) = \int_{\theta_1}^{\theta_2} v(\theta) dF(\theta) \ \text{ and } \ G(z) = g^{-1}(z).$$

Under tiering, the ISP's profit is

$$(1 - F(\theta_\ell))\ \underbrace{\pi v(\theta_\ell)\gamma\left(G\left(\frac{2\omega(p^*)\mathcal{I}(\theta_\ell, \theta_h)}{B}\right)\right)}_{s_\ell}$$

$$+ (1 - F(\theta_h))\underbrace{\pi v(\theta_h)\left(\gamma\left(G\left(\frac{2\omega(p^*)\mathcal{I}(\theta_h, \bar{\theta})}{B}\right)\right) - \gamma\left(G\left(\frac{2\omega(p^*)\mathcal{I}(\theta_\ell, \theta_h)}{B}\right)\right)\right)}_{s_h - s_\ell}$$

$$\underbrace{\sigma\left(\frac{BG\left(\frac{2\omega(p^*)\mathcal{I}(\theta_\ell, \theta_h)}{B}\right)}{2\omega(p^*)} + \frac{BG\left(\frac{2\omega(p^*)\mathcal{I}(\theta_h, \bar{\theta})}{B}\right)}{2\omega(p^*)}\right)}_{\eta}. \tag{26}$$

Define $\hat{\theta}$ as the θ that solves $\mathcal{I}(\underline{\theta}, \theta) = \mathcal{I}(\theta, \bar{\theta})$. It is readily seen that a solution must exist and, moreover, is unique. Suppose there is network neutrality (all content providers are served under a common service), but the ISP charges content providers an access fee $s = \pi\alpha(t^*, \underline{\theta}) = \pi v(\underline{\theta})\gamma(t^*)$. This is the most profitable situation, given neutrality and no exclusion, for the ISP. Observe that it is equivalent to the following trivial tiering arrangement: $B_\ell = B_h = B/2$, $s_\ell = s_h = v(\underline{\theta})\gamma(t^*)$, $\theta_\ell = \underline{\theta}$, and $\theta_h = \hat{\theta}$. We can now establish the following.

Proposition 7. Suppose the adjustment function, $\alpha(\tau, \theta)$, is multiplicatively separable in delay time and type (i.e., is given by (18)) with both functions differentiable. Then, the residential ISP will strictly prefer not to implement network neutrality.

© RAND 2012.

Proof. Consider two cases. In the first, $v(\underline{\theta}) > 0$. Observe that a feasible deviation from network neutrality is for the ISP to raise θ_h above $\hat{\theta}$ (maintaining the same bandwidth allocation between the two segments). The derivative of the first line of (26) with respect to θ_h is positive, whereas the derivatives of the second and third lines, given $\theta_h = \hat{\theta}$, are each zero. Hence, the ISP will wish to deviate from network neutrality.

As a second case, suppose $v(\underline{\theta}) = 0$. An option for the ISP is to eschew tiering but charge a positive access fee (i.e., $s > 0$) for a common class of service. Setting $s > 0$ is equivalent to raising the lowest content type served. As this will exclude some content providers, it violates network neutrality. Under such a single-tiered-service, the ISP's profit is given by (26), except $\theta_h = \bar{\theta}$. The derivative of profit with respect to θ_ℓ at $\theta_\ell = \underline{\theta}$ is

$$\pi v'(\underline{\theta})\gamma\left(G\left(\frac{2\omega(p^*)\mathcal{I}(\underline{\theta},\bar{\theta})}{B}\right)\right) - \sigma G'\left(\frac{2\omega(p^*)\mathcal{I}(\underline{\theta},\bar{\theta})}{B}\right)\underbrace{v(\underline{\theta})}_{=0} f(\underline{\theta}) > 0.$$

Hence, in this case too, the ISP will wish to deviate from network neutrality. *Q.E.D.*

An immediate corollary to Propositions 4 and 7 is the following.

Corollary 4. Suppose the adjustment function, $\alpha(\tau, \theta)$, is multiplicatively separable in delay time and type (i.e., is given by (18)) with both functions differentiable. Suppose too that the function defined by $t \mapsto t/\gamma(t)$ is a convex function. Then, an unconstrained ISP will *not* price in a manner that maximizes welfare.

An issue related to the above analysis is the following. Suppose full network neutrality were not an option; rather, suppose the policy choice was between allowing the ISP freedom to engage in second-degree price discrimination and allowing it to set a positive price for a common level of service. An earlier version of this article (available from the authors upon request) explored that choice in depth. The answer is that, depending on the exact assumptions, either policy could be the welfare-superior one. In large measure, this ambiguity reflects the usual ambiguity as to whether second-degree price discrimination on the basis of quality is welfare superior or inferior to linear pricing (see, e.g., Hermalin and Katz, 2007).

As remarked in footnote 6 above, a theoretical possibility is that the ISP could engage in perfect discrimination vis-à-vis the content providers. If the ISP can engage in perfect discrimination, then it will charge a type-θ content provider an access fee exactly equal to that content provider's profits; that is, $s = \alpha(\tau(\theta), \theta)\pi$. Consequently, the ISP's total profit (including from households) would be

$$\int_{\underline{\theta}}^{\bar{\theta}} (\pi + \sigma)\alpha(\tau(\theta), \theta) dF(\theta),$$

which matches expression (7), total welfare. The following result is, therefore, immediate.

Proposition 8. If the ISP can engage in perfect discrimination vis-à-vis all content providers, then the ISP will choose the bandwidth allocation (service prioritization) that maximizes welfare.

Proposition 8 is, of course, a well-known result in the two-sided-markets literature: if the platform captures 100% of the surplus from both sides of the market, the platform will act in the welfare-maximizing manner.

There are a number of reasons why one might wish to be skeptical of the practical implications of Proposition 8. First, such discrimination would require the ISP to possess and process a tremendous amount of information. With hundreds of millions of content providers,

© RAND 2012.

implementation of perfect discrimination is likely infeasible.[22] Possible evidence to its infeasibility is that, although ISPs have sought (or claimed) the regulatory right to engage in second-degree price discrimination via prioritization, to the best of our knowledge none has ever sought or attempted to engage in firm-specific (perfect) discriminatory pricing toward the content providers.

Second, current regulations would appear to prohibit ISPs from pricing in such a manner. For instance, firm-specific discrimination—essentially blocking a content provider from access to an ISP's residential customers unless the content provider paid the ISP 100% of the profits the content provider earns from transacting with the ISP's residential customers—seems extremely unlikely to pass the FCC's "reasonableness" test concerning discriminatory practices by ISPs.[23] Admittedly, regulations change, but permitting ISPs to engage in firm-specific discrimination seems not only not to be at the center of the current network neutrality debate but we see no evidence that it is even on the radar screen.

One reason it may not be part of the debate—and a third reason to question the practical importance of Proposition 8—is that were ISPs able to price in such a manner, the *dynamic* consequences would be fairly severe: because an existing or would-be content provider would face complete holdup by the ISPs, it would have no incentive to invest or innovate. Hence, the static efficiency of Proposition 8 ignores the significant dynamic inefficiency that such holdup would induce.[24]

5. Dynamic issues

■ The analysis to this point has treated bandwidth, B, as an exogenously given constant. In reality, the ISP determines B via its investments in capacity. In this section, we briefly consider the implications of various pricing regimes on the ISP's incentives to invest in bandwidth.

For convenience, we assume the marginal cost of bandwidth expansion to be a constant $\kappa > 0$. We assume there is no possibility of the residential ISPs shutting down, so we ignore fixed costs.

An unconstrained ISP faces the same optimization problem as (23), except now it is concerned with the true cost of bandwidth rather than its shadow cost; that is, the λ in (23) is replaced with κ. The solution must still satisfy the incentive-compatibility constraint that delivery times be nonincreasing in type (i.e., $\tau(\cdot)$ must still be nonincreasing).

Comparing overall welfare under the solution to this price-discrimination program to welfare when the ISP is constrained to offer a common level of service is impossible at a general level. We therefore limit attention to a set of tractable functional forms. Specifically, suppose that $\alpha(\tau, \theta) = \theta/\tau$ and F is the uniform distribution on $[0, 1]$. The suitably modified version of program (23) is, thus, to choose $\tau(\cdot)$ to maximize

$$\int_{\underline{\theta}}^{\bar{\theta}} \left(\left(\sigma + 2\pi - \frac{\kappa\omega(p^*)}{\tau(\theta)} \right) \theta - \pi \right) \frac{1}{\tau(\theta)} d\theta \,, \tag{27}$$

[22] As a simplifying assumption, we have assumed that households potentially interact with all content providers, which suggests that each household knows who all the content providers are. However, as we show in Section 6, this assumption can readily be relaxed; that is, each household potentially interacts with an idiosyncratic subset of all the content providers. In other words, our results do not require that households be more sophisticated than ISPs. no content provider needs to be known to all households.

[23] See FCC Report and Order 10-201, released December 23, 2010. One of the principles it sets forth (see paragraph 1) is "no unreasonable discrimination." See also paragraphs 6, 35, and 43 and Section III.C.

[24] That complete holdup, of the sort represented by perfect discrimination, would destroy all investment incentives is well-known (see, e.g., Williamson, 1975,1976; Tirole, 1986). The effect of partial holdup, as would result from imperfect (e.g., second-degree) discrimination, on investment incentives is more ambiguous; see Hermalin and Katz (2009) and the citations therein.

© RAND 2012.

subject to $\tau(\cdot)$ nonincreasing. Solving that program yields

$$\tau(\theta) = \begin{cases} \infty \text{ (exclusion)}, & \text{if } \theta \leq \dfrac{\pi}{\sigma + 2\pi} \\ \dfrac{2\kappa\theta\omega(p^*)}{(2\pi + \sigma)\theta - \pi}, & \text{if } \theta > \dfrac{\pi}{\sigma + 2\pi} \end{cases}. \tag{28}$$

It follows that the unconstrained ISP would build bandwidth:

$$B_{\text{un}} = \omega(p^*) \int_{\underline{\theta}}^{\bar{\theta}} \frac{\alpha(\tau(\theta), \theta)}{\tau(\theta)} dF(\theta) = \int_{\frac{\pi}{\sigma+2\pi}}^{1} \frac{((2\pi + \sigma)\theta - \pi)^2}{4\kappa^2\theta\omega(p^*)} d\theta = \frac{\sigma^2 - \pi^2 + 2\pi^2 \log\left(\frac{2\pi+\sigma}{\pi}\right)}{8\kappa^2\omega(p^*)}.$$

The situation in which the ISP is subject to network neutrality (nonexclusion and common service) is equivalent, in this example, to the situation in which there is no network neutrality constraint, but $\pi = 0$. Hence, its objective remains to maximize (27) (with π set to 0). From (28), it follows that the solution is $\tau(\theta) = 2\kappa\omega(p^*)/\sigma$ for all $\theta \in [0, 1)$. Straightforward calculations yield

$$B_{\text{n}} = \frac{\sigma^2}{8\kappa^2\omega(p^*)}.$$

Because $2\log(2) > 1$, it is readily seen that $B_{\text{UN}} > B_{\text{N}}$; that is, the ISP installs more bandwidth when unconstrained than when subject to network neutrality.

Welfare, including the variable cost of bandwidth, is, from (7) and (14′),

$$\int_{\underline{\theta}}^{\bar{\theta}} \left(\pi + \sigma - \frac{\kappa\omega(p^*)}{\tau(\theta)}\right) \alpha(\tau(\theta), \theta) dF(\theta).$$

which equals

$$\frac{(\pi + \sigma)(4\pi^2 + 3\pi\sigma + \sigma^2) - \pi^2(4\pi + 2\sigma)\log\left(\frac{2\pi+\sigma}{\pi}\right)}{8\kappa\omega(p^*)(2\pi + \sigma)} \tag{29}$$

if the ISP is unconstrained, and which equals

$$\frac{2\pi\sigma + \sigma^2}{8\kappa\omega(p^*)} \tag{30}$$

under network neutrality. Depending on the parameter values for π and σ, (29) can be greater or less than (30); that is, once the ISP's investment incentives are considered, network neutrality can still be welfare maximizing, but is not guaranteed to be so. Letting $\sigma = \zeta\pi$, it can be shown that (29) is not less than (30) if and only if

$$4 + 3\zeta - 2(2 + \zeta)\log(2 + \zeta) \geq 0.$$

Solving for the smallest ζ that satisfies that expression, we find that

$$\zeta = -2 + \exp\left(\tfrac{1}{2}(3 + 2\text{plog}(\exp(-3/2)))\right) \approx 1.314, \tag{31}$$

where plog: $\mathbb{R} \to \mathbb{R}$ is implicitly defined by $z = \text{plog}(z)\exp(\text{plog}(z))$.[25] To summarize, we have the following.

Proposition 9. Suppose the adjustment function, $\alpha(\tau, \theta)$, is θ/τ, the distribution of content types is uniform on [0, 1], and the ISP's marginal cost of additional bandwidth is a constant. Then an unconstrained ISP will install more bandwidth than an ISP subject to network neutrality. Welfare will be greater under network neutrality if the consumer surplus factor, σ, exceeds ζ times the gross profit factor, π, where the constant ζ is given by (31).

Proposition 9 indicates that there could be a fundamental distinction between efficient policy in a static setting and one in a dynamic setting. The greater bandwidth installed by an

[25] The function plog(·) is sometimes called the product-log function.

© RAND 2012.

unconstrained ISP is welfare enhancing *ceteris paribus*. On the other hand, there is still the static inefficiency that can arise from tiering. *A priori*, it is impossible to know which effect dominates.

Proposition 9 reaches a somewhat different conclusion from that reached by Choi and Kim (2010) concerning the effect of permitting price discrimination on the ISP's bandwidth investment incentives. Proposition 9 predicts that permitting price discrimination increases bandwidth unambiguously. In Choi and Kim's article, the effect is ambiguous; in particular, permitting price discrimination could *reduce* the ISP's incentive to build bandwidth. The reason for this difference in results is as follows. In Choi and Kim, the amount of content sent is independent of the ISP's capacity. Consequently, expanding capacity serves to increase speed on *both* the low-speed and high-speed service. Because the low-speed service is now faster, switching to low-speed service is more attractive for a high-type content provider than before. Hence, the amount of surplus the ISP can capture from selling high-speed service to such a content provider is reduced. This effect reduces the benefit the ISP gets from increasing capacity. This is a standard result for second-degree price discrimination on the basis of quality: improving the low type's quality reduces the rent that can be captured from the high type. Ambiguity arises in Choi and Kim because the high-speed service is also becoming faster. This increases a high-type content provider's willingness to pay for the high-speed service. As Choi and Kim observe, which of these two effects dominates is *a priori* ambiguous. In contrast, in our model, the amount of content traded is not fixed. The content providers' profits are increasing in the amount of content traded. price discrimination permits the ISP to capture some of the increased content-provider profit that greater bandwidth generates. Hence, the ISP's incentives to expand bandwidth are unambiguously greater when it can discriminate than when it cannot.

This discussion also indicates why the relative magnitudes of σ and π matter for the welfare conclusion. When σ is relatively large (exceeds $\zeta\pi$), the additional incentive to expand bandwidth that comes from being able to charge the content providers is relatively small; hence, the additional bandwidth vis-à-vis the amount it would build under neutrality is relatively small. So the additional-bandwidth benefit is correspondingly relatively minor and is outweighed by the static inefficiency of tiering.

6. Extensions

■ In this section, we briefly consider some extensions of our model.

□ **Unit pricing by the ISP.** In the United States, as well as many other countries, residential ISPs charge neither households nor content providers per unit of content transported (i.e., there are no per-byte or per-packet fees). Often, as modelled above, households pay a flat fee that entitles them to "all they can eat." In some places, there are limits on the amount that can be downloaded.[26] But even plans with limits do not charge on a per-unit basis. Nevertheless, it is worth briefly considering the implications for our analysis of per-unit pricing.

To start, suppose the ISP charges β per unit, where β does not depend on the content type. If the ISP charges households, then a household purchases $\alpha(\tau(\theta), \theta)\omega(p + \beta)$ from a θ-type content provider that is, itself, charging p per unit. The content provider will set its price to maximize

$$(q + p - c)\omega(p + \beta). \tag{32}$$

Alternatively, if the ISP charges the content provider directly, the content provider will set its price to maximize

$$(q + p - \beta - c)\omega(p). \tag{33}$$

[26] For instance, as of this writing, Rogers, a major Canadian ISP, has a service plan that provides unlimited use up to a monthly limit of 250 gigabytes of data download.

© RAND 2012.

Changing the control variable in (32) to $z = p + \beta$, it is readily seen that the two programs are the same. This is, of course, the usual result that statutory incidence is irrelevant to the determination of actual incidence. Let P^* denote the amount households pay, in total, in equilibrium. As just noted, it is independent of the statutory incidence.

It is straightforward to see that, for a β that is constant across content providers, the analysis of Section 3 carries through with P^* replacing p^* in the relevant expressions. To conclude, we have the following.

Proposition 10. Suppose the ISP charges either households or content providers a fee (β) per unit of content transferred. Propositions 1–5 remain true.

What per-unit fee would the ISP set? In answering this question, we limit attention to settings in which the ISP cannot charge content providers an access fee (i.e., $s \equiv 0$). Suppose, initially, that the per-unit fee cannot vary across content type. The ISP's profit is $\eta + \beta X$, where η is the household hookup fee and X is the total content traded; hence, the ISP's profit is

$$\begin{aligned}&\int_{\underline{\theta}}^{\bar{\theta}} \left(\alpha(\tau(\theta),\theta)\int_{\beta+p^*(\beta)}^{\infty}\omega(p)dp\right)dF(\theta)\\&\quad+\beta\int_{\underline{\theta}}^{\bar{\theta}}(\alpha(\tau(\theta),\theta)\omega(\beta+p^*(\beta)))dF(\theta)\\&\quad=\left(\beta\omega(\beta+p^*(\beta))+\int_{\beta+p^*(\beta)}^{\infty}\omega(p)dp\right)\int_{\underline{\theta}}^{\bar{\theta}}\alpha(\tau(\theta),\theta)dF(\theta),\end{aligned}\tag{34}$$

where $p^*(\beta)$ is the solution to maximizing (32) with respect to p.[27] It follows immediately from (34) that, given this form of pricing, the ISP does best under the welfare maximizing allocation of bandwidth (i.e., the allocation that maximizes the last integral in (34)). Its choice of β will maximize the expression in large parentheses in the last line of (34). To summarize, we have the folowing.

Proposition 11. An ISP that can charge a common fee (β) per unit of content but which cannot charge content providers an access fee (i.e., $s \equiv 0$) maximizes its profit by allocating bandwidth in the welfare maximizing manner.

If the ISP could set the per-unit fee as a function of the content (i.e., charge $\beta(\theta)$ per unit of type-θ content), then it is readily seen that its optimization program is essentially the same as (34):

$$\max_{\beta(\cdot)}\int_{\underline{\theta}}^{\bar{\theta}}\left(\beta(\theta)\omega(\beta(\theta)+p^*(\beta(\theta)))+\int_{\beta(\theta)+p^*(\beta(\theta))}^{\infty}\omega(p)dp\right)\alpha(\tau(\theta),\theta)dF(\theta).$$

Given any bandwidth division, pointwise optimization reveals that the ISP would set a common per-unit fee. But with a common per-unit fee, Proposition 11 implies the following.

Proposition 12. An ISP that can charge a fee ($\beta(\theta)$) per unit of content that varies by content type, θ, but which cannot charge content providers an access fee (i.e., $s \equiv 0$) maximizes its profit by allocating bandwidth in the welfare maximizing manner and by setting a *common* per-unit fee.

□ **Heterogeneous households.** A potential limitation to the analysis to this point is that it has assumed households are homogeneous. A more realistic model would recognize that households are heterogeneous in their preferences.

[27] An equivalent analysis can be done starting from (33).

© RAND 2012.

Two questions arise when the analysis is extended to allow for heterogeneous households:

(i) How to model household heterogeneity?
(ii) How to account for the consequent welfare loss that would arise from the residential ISP's monopoly pricing of *households'* Internet connections in an assessment of network neutrality?

The second question has to do with the following. With homogeneous households, the residential ISP's pricing to households is nondistortionary insofar as all households are served in equilibrium. If households are heterogeneous, then the usual monopoly distortion would arise: in order to capture more of the surplus of households that place a high value on connecting to the Internet, the ISP would end up excluding those households with a lesser value of connection. The usual deadweight loss associated with monopoly pricing attains. Moreover, there would be a negative access externality imposed on content providers—content providers lose profits they would have enjoyed had they been able to trade with the households that elect not to connect.[28]

The deadweight loss from too few households connecting would be reduced if the residential ISP were allowed to charge content providers. The reason is that, by connecting more households, the ISP can charge content providers more (the access externality). Hence, a result familiar to two-sided markets attains: the ability to charge the other side of the market (e.g., content providers) induces the platform to charge the one side (households) less. Although, as shown above, allowing the residential ISP to charge content providers can be welfare reducing in itself, we are in a second-best world and the theory of the second best applies. In particular, introducing the distortion of the ISP's being able to charge content providers ameliorates the distortion of the ISP's monopoly pricing to households. A departure from network neutrality could then be welfare improving in light of this amelioration effect.

This discussion raises, however, a philosophical point. Abandoning network neutrality is not inherently desirable; it simply may be desirable to offset another distortion. Hence, should the policy prescription be, drop neutrality or should it be to instead, to eliminate the root cause, which is the residential ISP's exercise of its monopoly position vis-à-vis the households? This, then, is the issue behind the second question.

Turning to the first question, there are many ways to model heterogeneous households in this context. One is that some households have an intrinsically greater value for Internet connections than others. This could be captured by multiplying the outer integral in (2) by a type parameter, ξ. An alternative means of capturing this would be to assume that the adjustment function varies across households (i.e., replace $\alpha(\tau, \theta)$ with $\alpha(\tau, \theta, \xi)$). Yet another way to model heterogeneous households is to assume different households want different content; hence, if $\Theta(\xi)$ is the measurable set of content types a type-ξ household may choose to consume, its utility is

$$y + \int_{\Theta(\xi)} \left(\int_0^{x(\theta)} m\left(\frac{x}{\alpha(\tau(\theta), \theta)} \right) dx \right) dF(\theta). \tag{2'}$$

The reader can readily envision various extensions and combinations of these formulations. What we will show, now, is that none of these ways of introducing heterogeneous households changes the results in Section 3 in any substantial way.

Suppose, first, that household utility is

$$y + \int_{\underline{\theta}}^{\bar{\theta}} \left(\int_0^{x(\theta)} m\left(\frac{x}{\alpha(\tau(\theta), \theta, \xi)} \right) dx \right) dF(\theta), \tag{2''}$$

where $\xi \in \Xi \subseteq \mathbb{R}_+$ denotes a household's type. Let $\Upsilon(\cdot)$ be the distribution function over Ξ. As a slight abuse of notation, we will sometimes write $\Upsilon(\widehat{\Xi})$ to denote the proportion of households

[28] See Taylor (1994) for an early discussion of access externalities. *Inter alia*, Economides and Tåg (2012) consider access externalities in an Internet context.

© RAND 2012.

with a type in set $\widehat{\Xi}$. Assume $\alpha(\tau, \theta, \cdot)$ is measurable with respect to Υ for all $\tau \in [0, \infty)$ and $\theta \in [\underline{\theta}, \bar{\theta})$. Define

$$A\big(\tau(\theta), \theta, \widehat{\Xi}\big) = \int_{\widehat{\Xi}} \alpha(\tau(\theta), \theta, \xi) d\Upsilon(\xi). \tag{35}$$

Assume that $\widehat{\Xi}$ is the set of households that connect to the Internet. Following the analysis of Section 2, it is readily seen that a given content providers demand is $A(\tau(\theta), \theta, \widehat{\Xi})\omega(p)$. It follows that all content providers price at the same p^* as in Section 2. Hence, a given content providers gross profit is $A(\tau(\theta), \theta, \widehat{\Xi})\pi$. It is readily seen that a given connected household's consumer surplus from trade with content provider θ is $\alpha(\tau(\theta), \theta, \xi)\sigma$ and, hence, aggregate consumer surplus from all trade with content provider θ is $A(\tau(\theta), \theta, \widehat{\Xi})\sigma$. It follows, therefore, that total welfare is

$$W = \int_{\underline{\theta}}^{\bar{\theta}} (\pi + \sigma) A(\tau(\theta), \theta, \widehat{\Xi}) dF(\theta). \tag{7′}$$

Straightforward calculations reveal that the identity between time and bandwidth allocation (i.e., expression (8)) is, in this context,

$$t_n \equiv \frac{1}{B_n}\omega(p^*) \int_{\Theta_n} A(\tau(\theta), \theta, \widehat{\Xi}) dF(\theta). \tag{8′}$$

Clearly, then, $A(\tau(\theta), \theta, \widehat{\Xi})$ acts as $\alpha(\tau(\theta), \theta)$ did in Section 3. We can conclude the following.

Proposition 13. Suppose household utility is given by (2 ″). Fix the set of household types that connect to the Internet. Propositions 1–3 and 5 remain true. In addition, if $A(\tau, \theta, \widehat{\Xi})$, as defined by (35), is multiplicatively separable into $\gamma(\tau)V(\theta, \widehat{\Xi})$ with $\partial V/\partial\theta > 0$ and $\gamma(\cdot)$ satisfying the conditions set forth in Proposition 4, then welfare neutrality is welfare superior to any division of the bandwidth and strictly superior to any division that either excludes some measurable segment of content or results in different groups of content providers having their content being sent with different transmission times in equilibrium.

Next, suppose that household utility is given by (2) except the outer integral is multiplied by the type parameter, ξ. Following the analysis of Section 2, it is clear that a connected household's demand for a given content provider's content is $\alpha(\tau(\theta), \theta)\omega(p/\xi)$. That content provider's total demand is

$$\alpha(\tau(\theta), \theta) \int_{\widehat{\Xi}} \omega(p/\xi) d\Upsilon(\xi) = \alpha(\tau(\theta), \theta)\Omega(p, \widehat{\Xi})$$

(note the implicit definition of Ω). It readily follows that all content providers will charge a common price, $\hat{p}$. Define $\hat{\pi}$ by

$$\hat{\pi} = (q + \hat{p} - c)\Omega(\hat{p}, \widehat{\Xi}).$$

A given household's consumer surplus from trade with a given content provider is

$$\alpha(\tau(\theta), \theta) \int_{\hat{p}}^{\infty} \omega\left(\frac{p}{\xi}\right) dp;$$

hence, aggregate consumer surplus from trade with that content provider is

$$\begin{aligned} &\alpha(\tau(\theta), \theta) \int_{\widehat{\Xi}} \left(\int_{\hat{p}}^{\infty} \omega\left(\frac{p}{\xi}\right) dp \right) d\Upsilon(\xi) \\ &\quad = \alpha(\tau(\theta), \theta) \int_{\hat{p}}^{\infty} \left(\int_{\widehat{\Xi}} \omega\left(\frac{p}{\xi}\right) d\Upsilon(\xi) \right) dp = \alpha(\tau(\theta), \theta) \int_{\hat{p}}^{\infty} \Omega(p, \widehat{\Xi}) dp. \end{aligned} \tag{36}$$

© RAND 2012.

Let $\hat{\sigma}$ equal the value of the last integral in (36). It follows, therefore, that all the analysis in Section 3 carries over to here, except with π, σ, and $\omega(p^*)$ replaced, respectively, by $\hat{\pi}$, $\hat{\sigma}$, and $\Omega(\hat{p}, \widehat{\Xi})$. We can conclude the following.

Proposition 14. Suppose household utility is given by (2), but with the benefit from the Internet scaled by household type (i.e., the outer integral multiplied by ξ). Fix the set of household types that connect to the Internet. Propositions 1–5 remain true.

Last, suppose that household utility is given by (2′); that is, different household types vary with respect to the content they may wish to acquire. Define

$$\Xi(\theta|\widehat{\Xi}) = \{\xi \in \widehat{\Xi} | \theta \in \Theta(\xi)\};$$

that is, $\Xi(\theta|\widehat{\Xi})$ is the set of *connected* household types that may wish to acquire content θ. Assume $\Xi(\theta|\widehat{\Xi})$ is measurable with respect to Υ for all $\theta \in [\underline{\theta}, \bar{\theta})$. The content demand for a household that wishes to trade at all with a given content provider is, from Section 2, $\alpha(\tau(\theta), \theta)\omega(p)$. That content provider's total demand is

$$\Upsilon(\Xi(\theta|\widehat{\Xi}))\alpha(\tau(\theta), \theta)\omega(p) = \mathcal{A}(\tau(\theta), \theta, \widehat{\Xi})\omega(p). \tag{37}$$

(Note the implicit definition of $\mathcal{A}$.) It is readily verified that all content providers price at the same p^* as in Section 2. It follows that a given content provider's gross profit is $\mathcal{A}(\tau(\theta), \theta, \widehat{\Xi})\pi$ and also that a given household's consumer surplus from trade with the content provider is $\alpha(\tau(\theta), \theta)\sigma$, where π and σ are the same as in Section 2. Aggregate consumer surplus from trade with a content provider is

$$\Upsilon(\Xi(\theta|\widehat{\Xi}))\alpha(\tau(\theta), \theta)\sigma = \mathcal{A}(\tau(\theta), \theta, \widehat{\Xi})\sigma.$$

It follows, therefore, that total welfare is

$$W = \int_{\underline{\theta}}^{\bar{\theta}} (\pi + \sigma)\mathcal{A}(\tau(\theta), \theta, \widehat{\Xi})dF(\theta). \tag{7′}$$

Straightforward calculations reveal that the identity between time and bandwidth allocation (i.e., expression (8)) is, in this context,

$$t_n \equiv \frac{1}{B_n}\omega(p^*)\int_{\Theta_n} \mathcal{A}(\tau(\theta), \theta, \widehat{\Xi})dF(\theta). \tag{8′}$$

Clearly, then, $\mathcal{A}(\tau(\theta), \theta, \widehat{\Xi})$ acts as $\alpha(\tau(\theta), \theta)$ did in Section 3. We can conclude the following.

Proposition 15. Suppose household utility is given by (2′); that is, different household types wish to trade with different content providers. Fix the set of household types that connect to the Internet. Propositions 1–3 and 5 remain true. In addition, if $\mathcal{A}(\tau, \theta, \widehat{\Xi})$, as defined by (37), is multiplicatively separable into $\gamma(\tau)\mathcal{V}(\theta, \widehat{\Xi})$ with $\partial\mathcal{V}/\partial\theta > 0$ and $\gamma(\cdot)$ satisfying the conditions set forth in Proposition 4, then welfare neutrality is welfare superior to any division of the bandwidth and strictly superior to any division that either excludes some measurable segment of content or results in different groups of content providers having their content being sent with different transmission times in equilibrium.

As noted, the introduction of heterogeneous households complicates a welfare assessment of discriminatory pricing (and corresponding service tiering) by ISPs vis-à-vis the content providers. The introduction of heterogeneous households also introduces the question of price discrimination by the ISP vis-à-vis the households themselves (e.g., residential ISPs could—and do—offer households faster speed for all the content they acquire or greater download allowances in exchange for higher monthly access fees). Unfortunately, an analysis of that question is beyond the scope of the current article.

© RAND 2012.

□ **Content providers with varying margins.** The analysis to this point has assumed that all content providers' margins are the same, holding price constant; that is, we assume the advertising rate, q, less the marginal cost of content production and transmission, c, is the same across the content providers (i.e., $\delta \equiv q - c$ is a constant). Here, we show that assumption can be partially relaxed, at least with respect to the analysis of Section 3.

Continue to assume a continuum of content providers of measure one. Let them now, however, be indexed by (θ, δ), where $\delta \in [\underline{\delta}, \bar{\delta}) \subset \mathbb{R}$. Assume, critically, that the two dimensions are distributed independently. As before, F is the distribution function over θ. Let $\Delta(\cdot)$ be the distribution function over δ. Making the content providers' types multidimensional greatly complicates any analysis of price discrimination by the ISP, and we do not attempt such an analysis in what follows.[29] In particular, we assume that any tiering is with respect to the θ dimension only (i.e., τ can be a function of θ but not δ).

Following the analysis in Section 2, a content provider with margin δ will set its price to maximize

$$(\delta + p)\omega(p);$$

let $p^*(\delta)$ denote the solution to that program. The content sold by a (θ, δ)-type content provider is, thus, $\alpha(\tau(\theta), \theta)\omega(p^*(\delta))$. Define

$$\pi(\delta) = (\delta + p^*(\delta))\omega(p^*(\delta)).$$

Define

$$\sigma(\delta) = \int_{p^*(\delta)}^{\infty} \omega(p)dp.$$

Welfare generated by a (θ, δ)-type content provider is

$$\alpha(\tau(\theta), \theta)(\pi(\delta) + \sigma(\delta)).$$

Hence, aggregate welfare generated by all $(\theta, \cdot)$ types is

$$\alpha(\tau(\theta), \theta) \int_{\underline{\delta}}^{\bar{\delta}} (\pi(\delta) + \sigma(\delta))d\Delta(\delta) = \alpha(\tau(\theta), \theta)(\Pi + \Sigma).$$

(Note the implicit definitions of Π and Σ.) Aggregate content traded by these types is

$$\alpha(\tau(\theta), \theta) \int_{\underline{\delta}}^{\bar{\delta}} \omega(p(\delta))d\Delta(\delta) = \alpha(\tau(\theta), \theta)\widehat{\Omega}.$$

(Note the implicit definition of $\widehat{\Omega}$) It is straightforward to verify that the analysis of Section 3 carries over to this case, with Π, Σ, and $\widehat{\Omega}$ in place of π, σ, and $\omega(p^*)$, respectively. To conclude, we have the following.

Proposition 16. Suppose content providers' margins vary independently of the household taste parameter (i.e., δ, the margin, and θ, the taste parameter, are independently distributed). Propositions 1–4 remain true.

If δ and θ are not independently distributed across content providers, then the previous analysis is no longer valid. That noted, it is not obvious that a particular dependence should exist between the tastes of the households and the production technology of the content providers or the scope for advertising. Hence, independence strikes us as a reasonable assumption as a first-order approximation to reality.

[29] For a discussion of the issues and complications that arise with screening over multiple dimensions, see Rochet and Choné (1998) or Basov (2010).

© RAND 2012.

□ **Generalization of the household utility function.** Our assumptions about household utility, expression (2), are essential to making the model tractable. Nevertheless, these assumptions are strong and it is therefore worth considering their importance for our analysis and the extent to which they can be relaxed.

An alternative specification to (2) is

$$\mathcal{U} = y + U\left(\int_{\underline{\theta}}^{\bar{\theta}} \left(\int_0^{x(\theta)} m\left(\frac{x}{\alpha(\tau(\theta),\theta)}\right) dx\right) dF(\theta)\right), \tag{38}$$

where $U(\cdot)$ is increasing and concave. Suppose $U(\cdot)$ is differentiable. If $U(\cdot)$ is not too curved, then a reasonable approximation to this last expression is

$$y + U'(\mathcal{M}) \int_{\underline{\theta}}^{\bar{\theta}} \left(\int_0^{x(\theta)} m\left(\frac{x}{\alpha(\tau(\theta),\theta)}\right) dx\right) dF(\theta) + U(\mathcal{M}) - \mathcal{M}U'(\mathcal{M}), \tag{2'''}$$

where $\mathcal{M}$ is a suitable constant around which to make a first-order Taylor expansion. It is clear that an analysis with $U'(\mathcal{M})m(\cdot)$ instead of $m(\cdot)$ will yield similar results to those derived above. Hence, (2‴) is equivalent to (2) and we can thus view our earlier analysis as an approximation to an analysis of the more general utility function given by (38).

In terms of economics, this discussion indicates that as long as the marginal utility from consuming a given content provider's content varies little when the consumption of all content changes (i.e., as long as $U'(\cdot)$ is close to constant), then (2) is a reasonable approximation of household preferences.

A related issue concerns the additive separability in (2). For disparate types of goods, additive separability is an accepted approximation in the literature. The approximation is less reasonable, however, for goods that are substitutes or complements. Even so, substitutes are not necessarily a significant issue: a plausible approximation is to model households as being of a type to consume from one or another content provider of similar content, but not from both (e.g., a household watches TV shows from Hulu or iTunes, but not both).[30] To the extent that is a reasonable approximation, we are effectively assuming households have preferences given by (2′). As shown in Proposition 15, such preferences yield similar results to those derived when preferences are given by (2). The issue of complementary content is less readily dismissed. Whether there are significant complementarities across the content of different content providers is an open question, but we see no obvious reason to suspect complementarities are so critical in this context that they render (2) or (2′) wholly invalid approximations of actual household preferences.

Lastly, there is the specification of marginal utility, expression (1). This specification generates the highly tractable demand function given in (5). Moreover, when combined with a quasilinear and additively separable utility function, expression (2), it justifies using consumer surplus as a measure of consumer welfare and provides tractable forms for both consumer surplus and overall welfare (expressions (21) and (7), respectively).

There are, of course, other specifications of demand besides (5) that would reflect the phenomenon that delays in content delivery should reduce demand for content, *ceteris paribus*. What we now show, via an example, is that similar conclusions to those reached earlier are still possible under such alternative specifications. To that end, assume

$$x(p,\theta) = \sqrt{\alpha(\tau(\theta),\theta)} - p\,. \tag{39}$$

Observe that we are returning to the assumption of a single dimension of type. As a convenient normalization, suppose $q - c \equiv 0$. Familiar calculations reveal that a content provider's profit

[30] An objection to assuming different household types consume one or the other similar content, but not both, is that it rules out the potential effect of prices on household choice. On the other hand, if we assume households have types, then many models of differentiated-goods competition would justify such segregation arising in equilibrium. Given we have, in fact, ruled out content-provider competition, such a justification is, admittedly, speculative.

© RAND 2012.

is $\alpha(\tau(\theta), \theta)/4$, consumer surplus is $\alpha(\tau(\theta), \theta)/8$, and the welfare generated by trade with this content provider is thus $3\alpha(\tau(\theta), \theta)/8$. The amount of content traded is $\sqrt{\alpha(\tau(\theta), \theta)}/2$.

Given (39), the Lagrangean associated with constrained welfare-maximization (the analog to expression (15)) is

$$\int_{\underline{\theta}}^{\bar{\theta}} \left(\tfrac{3}{8}\alpha(\tau(\theta), \theta) - \tfrac{1}{2}\lambda \frac{\sqrt{\alpha(\tau(\theta), \theta)}}{\tau(\theta)} \right) dF(\theta).$$

Pointwise optimization with respect to $\tau(\theta)$ yields the first-order condition

$$\frac{4\lambda\sqrt{\alpha(\tau(\theta),\theta)} + \frac{\partial\alpha(\tau(\theta),\theta)}{\partial\tau}\left(3\tau(\theta)^2 - 2\frac{\lambda\tau(\theta)}{\sqrt{\alpha(\tau(\theta),\theta)}}\right)}{8\tau(\theta)^2} = 0. \tag{40}$$

Suppose $\alpha(\tau, \theta) = \theta/\tau^3$; then the solution to (40) is

$$\tau(\theta) = \left(\frac{9}{10\lambda}\right)^2 \theta,$$

that is, welfare is enhanced by giving priority to *lower* θ content. Given that content-provider profit is still proportional to $\alpha(\tau, \theta)$, it remains true that the only feasible discrimination schemes are those in which priority is given to higher θ content. Hence, for this example, we arrive at the same cautionary result as in Proposition 5: namely, that among the *feasible* discrimination schemes, network neutrality is welfare maximizing. Our key conclusion is, therefore, not dependent upon our having assumed (1). Of course, as was true of our original analysis, under a different functional form for α, a feasible discrimination scheme could exist that is welfare superior to neutrality (such would, in fact, be true, given (39), if $\alpha(\tau, \theta) = \theta/\tau$).[31]

7. Concluding remarks

■ This article has analyzed the private and social benefits of allowing residential ISPs to exclude or provide tiered-service to content and application providers seeking to engage with the ISPs' residential subscribers. We have shown that under a reasonable model of household (end-user) behavior, the evaluation of various tiering or prioritization schema can be reduced to an assessment of which schema will permit the greatest flow of content (Proposition 1). The reason we can focus on the amount of content is that households optimally adjust their consumption of different content as the transmission speeds of that content vary.

There are practical and regulatory reasons to see the set of possible schema as being limited. In particular, the focus of much of the network neutrality debate has been on schema that are forms of second-degree price discrimination. Because content providers must be induced to play their part in any such price-discrimination scheme, the set of feasible schema is restricted. Hence, even if network neutrality—no exclusion and no discrimination—is not welfare maximizing generally, it could be welfare maximizing within this restricted set. In particular, if the elasticity of content demand with respect to transmission time does not increase with households' time sensitivity for the content, then network neutrality is welfare maximizing within the set of feasible schema (Proposition 5). If that elasticity is invariant with households' time sensitivity for the content, then network neutrality is welfare superior to all other schema (feasible or not), as shown in Proposition 4.

Endogenizing an ISP's bandwidth-building decision potentially alters some of these conclusions; that is, dynamic issues could make departing from network neutrality welfare superior to maintaining it. Using a limited version of our more general model, we showed that the static

[31] The existence of a feasible scheme that dominates neutrality does not imply that the scheme a profit-maximizing ISP would choose to impose would necessarily be welfare superior to neutrality; as noted in the discussion around Proposition 5, the ISP could "overdo it."

© RAND 2012.

efficiency of neutrality nevertheless outweighs dynamic issues if household utility significantly dominates as a share of overall welfare.

As with any model, a number of simplifying assumptions were imposed. As we showed in Section 6, many of our results continue to hold even relaxing those assumptions. As, however, was also made evident in that section, relaxing some assumptions generates other issues to be explored. Among these, a critical one could be an analysis of what happens when the ISP is engaging in price discrimination on both sides of the market, that is, unlike in the analysis in this article, exploring what happens if the ISP can also engage in discrimination on the house hold side of the market. Although additional study on issues such as these is warranted—and we recognize this article, is unlikely to be the final word on the topic—we nonetheless believe the various insights offered herein will serve to inform a debate that has been long on advocacy and short on analysis.

Appendix

This Appendix contains the proof of Proposition 4.

Proof. In light of Proposition 1, the result follows if we can show that content sent under neutrality is not less than (or exceeds) output under division.

From (9) and (20), content sent under neutrality is

$$Bg^{-1}\left(\frac{\omega(p^*)}{B}\int_{\underline{\theta}}^{\bar{\theta}} v(\theta)dF(\theta)\right). \tag{A1}$$

Define $\phi_n = B_n/B$; that is, ϕ_n is the proportion of the bandwidth allocated to the nth transmission group. Denote by $\mathcal{N}$ the class of groups that are allocated a positive amount of bandwidth (i.e., $\phi_n > 0$ for $n \in \mathcal{N}$). From (9) and (19), content sent under division is

$$\sum_{n\in\mathcal{N}} \phi_n Bg^{-1}\left(\frac{\omega(p^*)}{\phi_n B}\int_{\Theta_n} v(\theta)dF(\theta)\right). \tag{A2}$$

Because the ϕ_n sum to one, expression (A2) is an expected value. By Jensen's inequality, it follows that (A2) cannot exceed

$$\begin{aligned} Bg^{-1}\left(\sum_{n\in\mathcal{N}} \phi_n \frac{\omega(p^*)}{B\phi_n}\int_{\Theta_n} v(\theta)dF(\theta)\right) &= Bg^{-1}\left(\frac{\omega(p^*)}{B}\sum_{n\in\mathcal{N}}\int_{\Theta_n} v(\theta)dF(\theta)\right) \\ &\leq Bg^{-1}\left(\frac{\omega(p^*)}{B}\int_{\underline{\theta}}^{\bar{\theta}} v(\theta)dF(\theta)\right), \end{aligned} \tag{A3}$$

where the last inequality is strict if $[\underline{\theta}, \bar{\theta}) \setminus \bigcup_{n\in\mathcal{N}} \Theta_n$ is a set of positive measure (i.e., if a positive measure of content providers are excluded). The last term in (A3) is (A1) (i.e., content under neutrality). So no division can increase the amount of content sent relative to neutrality, and a division strictly reduces the amount sent if it excludes a positive measure of content providers. Finally, if, for $n, m \in \mathcal{N}$,

$$\frac{\omega(p^*)}{\phi_n B}\int_{\Theta_n} v(\theta)dF(\theta) \neq \frac{\omega(p^*)}{\phi_m B}\int_{\Theta_m} v(\theta)dF(\theta), \tag{A4}$$

then Jensen's inequality implies (A2) is strictly less than the first term in (A3). Multiplying the left side of (A4) by $\gamma(t_n)/\gamma(t_n)$ and the right side by $\gamma(t_m)/\gamma(t_m)$, it follows that inequality (A4) holds if and only if $g(t_n) = t_n/\gamma(t_n) \neq t_m/\gamma(t_m) = g(t_m)$. Because $g(\cdot)$ is strictly monotonic, (A4) holds if $t_n \neq t_m$. This establishes that a division, which results in content providers having their content being sent with different transmission times, is strictly welfare inferior to neutrality. *Q.E.D.*

References

Basov, S. *Multidimensional Screening*. Berlin: Springer Verlag, 2010.

Cheng, H.K., Bandyopadhyay, S., and Guo, H. "The Debate on Net Neutrality: A Policy Perspective." *Information Systems Research,* Vol. 22 (2011), pp. 60–82.

Choi, J.P. and Kim, B.C. "Net Neutrality and Investment Incentives." *RAND Journal of Economics,* Vol. 41 (2010), pp. 446–471.

© RAND 2012.

DURANTON, G. AND TURNER, M.A. "The Fundamental Law of Road Congestion: Evidence from US Cities." *American Economic Review,* Vol. 101 (2011), pp. 2616–2652.

ECONOMIDES, N. AND TÅG, J. "Net Neutrality on the Internet: A Two-Sided Market Analysis." *Information Economics and Policy*, Vol. 24 (2012), pp. 91–104.

FEDERAL COMMUNICATIONS COMMISSION. *Report and Order* 10–201, Washington, DC, 2010.

HERMALIN, B.E. AND KATZ, M.L. "The Economics of Product-Line Restrictions with an Application to the Network Neutrality Debate." *Information Economics and Policy,* Vol. 19 (2007), pp. 215–248.

———. "Information and the Hold-up Problem." *RAND Journal of Economics,* Vol. 40 (2009), pp. 405–423.

KRÄMER, J. AND WIEWIORRA, L. "Network Neutrality and Congestion-Sensitive Content Providers: Implications for Service Innovation, Broadband Investment and Regulation." NET Institute Working Paper 10–09, 2010.

ROCHET, J.C. AND CHONÉ, P. "Ironing, Sweeping, and Multidimensional Screening." *Econometrica,* Vol. 66 (1998), pp. 783–826.

TAYLOR, L.D. *Telecommunications Demand in Theory and Practice* Dordrecht: Kluwer Academic, 1994.

TIROLE, J. "Procurement and Renegotiation." *Journal of Political Economy,* Vol. 94 (1986), pp. 235–259.

———. *The Theory of Industrial Organization*. Cambridge, Mass.: MIT Press, 1988.

VARIAN, H.R. "Price-Discrimination and Social Welfare." *American Economic Review,* Vol. 75 (1985), pp. 870–875.

———. "Price-Discrimination." In R. SCHMALENSEE AND R. WILLIG, eds., *Handbook of Industrial Organization*, Vol. 1., 1st ed. Amsterdam: Elsevier, 1989.

WILLIAMSON, O.E. *Markets and Hierarchies, Analysis and Antitrust Implications: A Study in the Economics of Internal Organization*. New York: The Free Press, 1975.

———. "Franchise Bidding for Natural Monopolies—In General and with Respect to CATV." *Bell Journal of Economics,* Vol. 7 (1976), pp. 73–104.

© RAND 2012.