Aalto University
School of Business

The Economic Effects of Network Neutrality: A Policy Perspective

Economics
Master's thesis
Kasper Eerola
2015

Department of Economics
Aalto University
School of Business

Kasper Eerola / Department of Economics / Aalto University

MSc Thesis – April 2015

The Economic Effects of Network Neutrality: A Policy Perspective

Abstract

Network neutrality – regulation of Internet service providers (ISPs) to ensure equal treatment of all traffic – is becoming something many people have heard about. While the context is technical, network neutrality ultimately boils down to economics. The political weight of the subject is heavy, and the international debate is fierce. Still, surprisingly little rigorous research appears to be behind it. In this paper, I review economic literature on network neutrality and ISP regulation, covering both practical and theoretical implications for the broadband market. I define the degrees of network neutrality with more granularity than papers so far, evaluate the qualitative economic effects of regulation, and describe the broadband market, frameworks for modeling it, and its peculiar economic characteristics. In particular, I review and compare different theoretical modeling approaches and models' predictions of the welfare effects of different regulatory regimes. Throughout the paper, I incorporate economic literature from relevant areas into the analysis. I do not make definite policy recommendations, but I draw conclusions that are potentially of interest from a policy point of view.

My analysis would indicate that the complexity of the Internet ecosystem and interrelations between market participants make effective regulation difficult. There is no economic evidence that network neutrality generally increases total welfare. In fact, it turns out that from a well-rounded economic perspective, strong network neutrality appears in most cases as detrimental to both consumer surplus and total welfare. In certain scenarios, however, models predict that neutrality can increase static and dynamic efficiency. The results depend crucially on model specifications and parameters, which differ significantly across the literature. So far, there is no consensus among economists on the optimal level of ISP regulation. Market-driven solutions such as dynamic pricing might provide a way to circumvent the neutrality question.

Keywords: network neutrality, broadband market, Internet service provider, telecommunications policy, welfare

Tiivistelmä

Verkkoneutraliteetti – teleoperaattorien sääntely tietoliikenteen tasa-arvoisen kohtelun varmistamiseksi – on astunut käsitteenä julkisuuteen. Vaikka konteksti onkin tekninen, verkkoneutraliteetti viime kädessä redusoituu taloustieteeseen. Aiheen poliittinen painoarvo on suuri ja kansainvälinen keskustelu kiivasta. Tästä huolimatta sen takaa vaikuttaa löytyvän yllättävän vähän tieteellistä tutkimusta. Lopputyössäni tarkastelen taloustieteellistä kirjallisuutta verkkoneutraliteetista ja teleoperaattorien sääntelystä ja sen vaikutuksia laajakaistamarkkinaan käytännöllisestä kuin myös teoreettisesta näkökulmasta. Määrittelen verkkoneutraliteetin asteet hienojakoisemmin kuin aikaisemmat julkaisut, arvioin sääntelyn laadullisia vaikutuksia ja kuvailen laajakaistamarkkinaa, viitekehyksiä sen mallintamiseksi sekä sen eriskummallisia taloudellisia piirteitä. Kuvaan teoreettisia lähestymistapoja ja merkittävimpien mallien ennusteita sääntelymallien hyvinvointivaikutuksista. Liitän analyysini relevanttiin taloustieteelliseen kirjallisuuteen. En anna suoria politiikkasuosituksia, mutta teen johtopäätöksiä, jotka ovat mahdollisesti mielenkiintoisia politiittisesta näkökulmasta.

Analyysini perusteella vaikuttaa, että Internet-ekosysteemin monimutkaisuus ja toimijoiden väliset suhteet tekevät tehokkaasta sääntelystä vaikeaa. Taloustieteellistä näyttöä verkkoneutraliteetin hyvinvointia kasvattavista vaikutuksista ei ole. Tasapainoisesta taloudellisesta näkökulmasta katsottuna tiukka neutraliteettisääntely näyttää useimmissa tapauksissa sekä pienentävän kuluttajan ylijäämää että laskevan kokonaishyvinvointia. Joissakin skenaarioissa mallit toisaalta ennustavat neutraliteetin lisäävän staattista ja dynaamista tehokkuutta. Tulokset riippuvat rajusti mallin rakenteesta ja parametreistä, jotka vaihtelevat merkittävästi tutkimuksesta tutkimukseen. Toistaiseksi taloustieteilijät eivät ole päässeet yhteisymmärrykseen optimaalisesta teleoperaattorien sääntelyn asteesta. Markkinalähtöiset ratkaisut kuten dynaaminen hinnoittelu saattavat mahdollistaa neutraliteettikysymyksen kiertämisen.

Avainsanat: verkkoneutraliteetti, laajakaistamarkkina, teleoperaattori, telepolitiikka, hyvinvointi

Contents

Figures

Tables

1. Introduction

[We] each pay to connect to the Net, but no one can pay for exclusive access to me. – Tim Berners-Lee, inventor of the World Wide Web[1]

I would like to begin by saying what this paper is not: A statement as to whether or not or to what degree "network neutrality" ought to be regulated.[2] There are good arguments both ways – be they economic, technological, political, or ethical. In what follows, I aim for an objective, even if an economics-based analysis.[3]

That being said, in this paper I analyze the potential effects of *network neutrality*, or rather, degrees of neutrality, from an economic point of view. Simply put, network neutrality (also "net neutrality" or "Internet neutrality" or the "Open Internet") refers to the regulation of Internet service providers (ISPs) by requiring them to treat all traffic equally. The political weight of the subject is heavy, and the international debate is fierce. Still, surprisingly little rigorous research appears to be behind it all. I have chosen the research topic on the grounds that even with the ongoing developments, well-rounded economic analysis of ISP regulation remains scarce. Moreover, the modern broadband market in itself is somewhat sparsely studied in the economic literature. However, during the past decade, and especially the last few years, enough articles about the broadband market have been published so that some standard modeling frameworks have been established and a solid base for further research has been built. In addition, there is a range of surrounding literature applicable to the market in interesting ways. The contribution of my thesis to the existing stream of economic literature stems from the element of synthesis: I compare different economic approaches to network neutrality, define its forms with more granularity than papers so far, and present a review on the broadband market as a whole. While not making definite policy recommendations, I draw practical conclusions potentially of interest from a policy perspective. As developments in the field continue to unfold, I believe an up-to-date

[1] dig.csail.mit.edu. Accessed 29 Oct 2014.

[2] For a statement on the matter, see the President's Statement: www.whitehouse.gov. Published and accessed 10 Nov 2014.

[3] Full disclosure: I thank Telealan edistämissäätiö for a scholarship toward this independent thesis. In the same vein, I thank my supervisor Mikko Mustonen for helpful comments.

review on the background, current state, and models of this economically fundamental industry – yet unfamiliar or even mysterious to many – is valuable as (an economist's) common knowledge. Unfortunately, due to the broad ramifications of the neutrality question, limitations in the coverage of my particular study are unavoidable. My research questions are the following:

i. How to model the broadband market and formulate an economic framework for policy analysis?
ii. What peculiar economic characteristics does the broadband market exhibit, and how is the market developing?
iii. What are the qualitative economic effects of network neutrality and alternative regulatory regimes?

The remainder of the paper is organized as follows. In Section 2, I review the technological background of the broadband market and network neutrality, the forms of which I then proceed to define in detail. In Section 3, I describe the broadband market and characterize different economic forces at play in it. In Section 4, I analyze the potential implications of deregulation in light of these issues. In Section 5, I present theoretical modeling approaches and review the models' predictions of the welfare effects of neutrality regulation. The majority of models specify a two-sided market with an ISP monopoly or duopoly, two or more content or service providers, and a continuum of Internet users. Finally, I discuss the policy implications. Throughout the paper, I incorporate extant economic literature into the analysis. Section 6 concludes.

2. Background and Definitions

The Internet is perhaps the most fundamental building block of the modern information society. According to an estimate, the emergence of broadband Internet has generated up to 50% of the US GDP in 1999-2006 (Greenstein and McDevitt 2011). Consequently, the regulation of ISPs is related to a myriad of big questions. While the context is technical, the network-neutrality debate ultimately boils down to economics. As Hazlitt and Wright (2012) put it, "Whatever the engineering designs of networks or the interfaces between them, the terms of trade on which demanders and suppliers transact are economic. [...] They are the standard building blocks of markets: Property and contracts, layered upon a general legal regime enabling ownership, production, and trade." The big questions related to network neutrality include economic welfare (efficiency, prices, surpluses), competition policy (vertical integration, price discrimination, bundling and tying), technological innovation (incentives to invest in technology), telecommunications engineering, digital and intellectual property rights, privacy, freedom of speech, censorship, and equality, among others. I elaborate on the economic aspects over the course of this paper.

To establish a more cohesive framework for understanding the broadband market, a brief review on the technological background of the Internet and network neutrality might be justified. Then we can specifically define neutrality to be able to evaluate with consistency its meaning for the market. Additionally, I shed some light on why neutrality is debated in the first place, and summarize recent regulatory developments.

2.1 Technical Preliminaries

Logic of the Internet

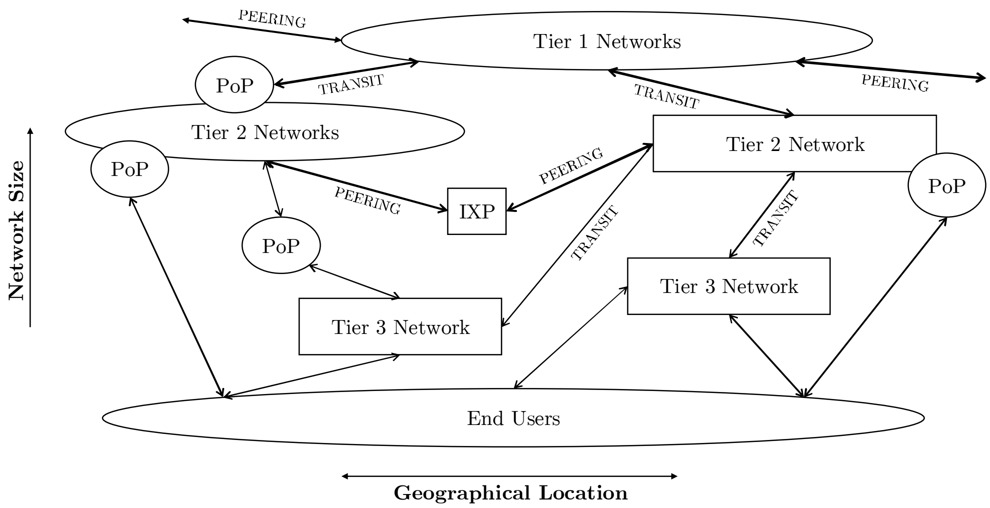
The Internet is a global network of private, public, commercial, academic, and government computer networks that has existed for about 20 years in the form we use it today. It was developed on top of the TCP/IP protocol from the US Department of Defense's ARPANET dating back to the 1960s and the WWW protocol invented by Tim Berners-Lee's at the turn of the 1990s. From the approximate 200 terabytes in 1994, Internet traffic has grown to an estimated 600 exabytes in 2014 – a 30-million-fold increase – and is expected to triple by 2018.[4] While the technological specifics of the systems are outside the scope of this study, Figure 1 shows a schematic characterization of the Internet topology.

[4] Cisco Visual Networking Index 2014. Accessed 21 Dec 2014.

An *end user* connects from a computer (or mobile device) through the cables in the *local loop* (or wirelessly) to her residential ISP's network. The local ISP providing the end user's "last-mile" broadband Internet access then forwards the data to higher-level networks, operated by *transit ISPs* and *carriers*, from which the ISP has purchased *transit* access, ultimately through the Internet *backbone* to their destination. End users include both residential and business broadband users. *Interconnect agreements* between ISPs allow them to connect to each other and data to be transmitted between an end user's computer and a *content/service provider*'s (CSP) server, or in the case of a peer-to-peer (P2P) connection, directly between end users. CSPs can be defined very broadly to include news sites, social media, online marketplaces, voice-over-IP (VoIP) applications, video and music streaming services, and any Internet content and services in between.

The backbone comprises principal data routes, hosted at large commercial or public data centers with Internet exchange and network access points. Data are transmitted between the largest ISPs and carriers in "Tier 1" networks, where the operators use each other's networks reciprocally as per the principle of *peering*. The principal data routes are optical (submarine) cables that carry virtually all the Internet traffic. Cables have been constructed and are operated by carriers and private telecommunications service providers (TSPs), from which ISPs lease capacity or purchase IRUs (indefeasible rights to use). As such, there are two layers of economic activity in the connectivity ecosystem: (1) Internet access provision by ISPs and (2) physical network operation by carriers (van Schewick 2007).

Figure 1. Schematic representation of the Internet topology. In Tier 1 networks, ISPs engage in peering, i.e. they utilize each other's networks reciprocally. Tier 2 ISP's purchase access to Tier 1 networks (transit), but may also have peering agreements between each other (secondary peering). Tier 3 operators purchase access to Tier 2 and Tier 1 networks. Points of presence (PoPs) connect ISPs and CSPs to networks, and Internet exchange points (IXPs), some of which are publicly maintained (public peering vs. private peering), connect networks to each other. Adapted from "Internet Connectivity Distribution & Core," user "Ludovic Ferre," Wikimedia Commons, September 2014.

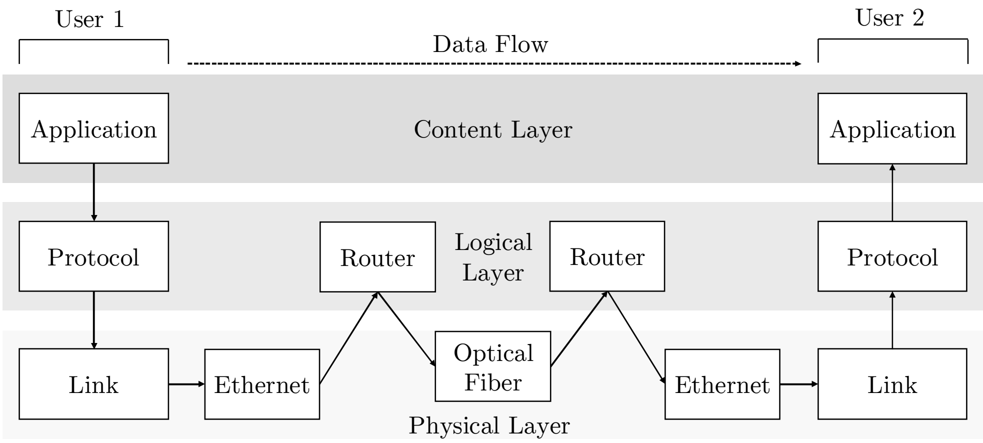


Routing of traffic on the Internet is a nontrivial problem. While the basic idea is that each router "maps" the surrounding network into a routing table and forwards each packet to the next hop toward the destination IP address, various economic, technological, and other private objectives are at play (Caesar and Rexford 2005). The transmission route of a packet is not always the one with the smallest hop count: The terms of interconnect agreements, revenue generation by transmitting through customer networks, and load balancing between networks influence the packet's path.

Another dimension of the Internet is its logical architecture as opposed to its geographical-topological structure described above. The Internet is usually treated in telecommunications engineering as a four-layer protocol suite. On the Internet, data streams are split into packets of roughly 1000 bytes (8000 bits). In *packet switching*, as opposed to circuit switching in GSM networks, multiple connections share the conduit simultaneously. The four abstraction layers are, from top to bottom, the application layer, transport layer, internet layer, and link layer. At the application layer, protocols such as HTTP standardize the interfacing methods and underlying protocols. At the transport layer, protocols such as TCP organize the transportation of the packets between users. The IP protocol at the Internet layer then specifies the structure of the packets themselves. The link layer comprises physical technologies, including Ethernet, for transmitting the bits and bytes within the packets between adjacent nodes in the network.

To simplify things a bit, it is enough for us to slice the system into three layers: (1) the *content layer*, (2) *logical layer*, and (3) *physical layer* (Ganley and Allgrove 2006). Figure 2 displays the logical structure of the Internet through these abstraction layers.

Figure 2. Schematic representation of the Internet architecture. Within an application or a web site, a data stream is sent from User 1's (or a CSP's) computer to User 2's receiving computer. Packets are transmitted through physical pipes and organized using protocols at the logical layer. Adapted from "IP stack connections," user "Cburnett," Wikimedia Commons, January 2015.



The Internet was designed to follow the so-called *dumb-pipe* principle. A "dumb pipe" refers to a network that has no intelligence of its own: It simply transmits bit streams as they come, treating each packet equally and not distinguishing between different types of data within packets. The dumb-pipe principle is closely related to network neutrality in the sense that no packet gets special treatment on the basis its content.

Quality of Service

Inherent to the Internet is a tradeoff between reliability and efficiency. The system is still not, nor will probably ever be, optimized for the latest commercial requirements. Scarcity of *bandwidth* is one of the central concepts in the economics of network neutrality. Bandwidth in the computer-networking context is defined as the net bitrate, channel capacity, or maximum throughput of a connection within a communication system. For our purposes, bandwidth can be simply thought of as the speed of the Internet connection. Since cables and routers in the networks have limited capacities, packets may need to be *queued* before they can be forwarded. *Congestion* occurs when the load is too high relative to the capacity, and *quality of service* (QoS) deteriorates, i.e. the connection speed decreases and there is *latency* (delay), jitter (variation in latency), or packet loss. A computer at the edge of the network typically slows down its transmission rate if it detects congestion. Still, the ISP can and often will manage congestion from within the network. One way to alleviate congestion would be to slow down a set of data streams to allow room for the rest. Prioritization insofar as it is technologically necessary and not harmful to consumers is sometimes tolerated by regulators, vague as current legislation is. Clearly, another solution to congestion is to add capacity. For technological reasons, mobile broadband generally has lower speed and higher latency than fixed broadband. Lately, 4G technologies such as LTE-A have started to challenge and surpass fixed connections in speed if not latency.

2.2 Network Neutrality

The term "network neutrality" was coined by Tim Wu (2003). Network neutrality can be defined as follows.

<u>Definition. Network Neutrality (NN)</u>. The regulatory principle that *all Internet traffic be treated by the ISP equally and without regard to content, source, or destination.* More specifically, this leads to two corollaries.

i. Demand-side neutrality: The ISP cannot *discriminate* or *prioritize* or *filter*[5] packets based on the origin, destination, or content.
ii. Supply-side neutrality, i.e. the *zero-price rule*: The ISP cannot charge content and service providers a *termination fee* for access to its customers.

As a prophylactic *ex-ante* rule, network neutrality contrasts with *ex-post* case-by-base regulation. It is compatible with ordinary *user tiering*, i.e. offering different bandwidth and QoS options to customers (Krämer et al. 2013). Data discrimination and termination pricing would often but not necessarily be interrelated. Crucially, the zero-price rule only concerns the relations between residential last-mile ISPs and CSPs

[5] From here on, I use "discrimination" and "prioritization" interchangeably. I also reserve the right to use "neutrality" and "non-neutrality" quite liberally when the context so allows. Finally, I use "demand side" ("supply side") or "retail market" ("wholesale market") to refer to the end-user (CSP) side of the broadband market.

separate from each other, not the relations between directly interconnected ISPs and CSPs. While part of the Internet ecosystem, *interconnection* or *access prices* charged by Tier 1 or 2 ISPs to Tier 3 ISPs or by ISPs at the end of the network to CSPs are not the central target of neutrality regulation. Sizable settlements on direct interconnection points between CSPs and ISPs already happen (3.2).

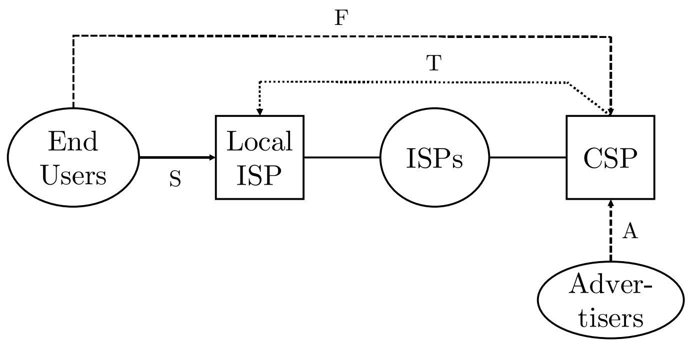
Degrees of Network Neutrality

In most discussions and studies, network neutrality is treated as a straightforward yes-no binary question or is vaguely defined. As Gans (2014) remarks, the reality is subtler. Gans classifies the degrees of network neutrality with more granularity than most papers, into (1) *strong network neutrality* where the ISP cannot discriminate in its prices to end users or CSPs, (2a) *weak content-provider neutrality* where the ISP cannot discriminate in its price to CSPs, (2b) *weak consumer neutrality* where the ISP cannot discriminate in its price to end users, and (3) *no regulation.* As it turns out, even this level of granularity can be insufficient. So far, I have defined a strict form of network neutrality, the NN regime. However, multiple degrees of regulation can be considered as natural generalizations of both parts of the NN definition.

On the demand side, the first regulatory possibility is that no data discrimination or traffic shaping is allowed, with the exception of special circumstances such as illegal content. The second possibility is to allow the ISP to discriminate based on the traffic class such as voice or video, but not based on individual CSPs. For example, the ISP is allowed to prioritize voice over video, but not Skype over Google Hangouts. Under a laissez-faire policy, discrimination based on specific CSPs is also allowed. Bandwidth *throttling* and *blocking*, i.e. intentional slowing down or blocking of a traffic class or service even in the absence of congestion, can also be allowed. The prohibition of data discrimination without explicit regard to pricing practices is termed as a *no-exclusivity rule* by Kourandi et al. (2014). According to the rule, the ISP can charge CSPs termination fees but cannot contract with them on the exclusivity of content, i.e. sign exclusivity contracts. Content can be defined as exclusive when the ISP lets the content of only one or a few CSPs within a content class through to its end users, and denies access to the rest. Finally, the ISP may be required to offer minimum QoS to end users, a neutral slow lane to end users, or a zero-price slow lane to CSPs (as analyzed in e.g. Economides and Hermalin [2012] and Peitz and Schuett [2013]).

On the supply side, the zero-price rule is in effect a special case of price-cap regulation. Funnily enough, price caps as a generalized form of ISP regulation have been largely neglected in the literature. First, the regulator can cap the termination fee the ISP is allowed to charge CSPs. Alternatively, the termination fee must be set uniformly across CSPs without discrimination. Another alternative is to allow termination fees to differ across CSPs but require them to be "fair" by reflecting the given CSP's data intensity and the ISP's capacity provisioning costs. These considerations lead the regulator to a choice of a supply-side pricing policy in conjunction with a demand-side network policy. Recall that the ISP charging a termination fee is not generally the same one that provides upstream connectivity to the CSP. Rather, it is the local ISP, that is, the residential access ISP, which is in position to charge CSPs for last-mile access to end users. Figure 3 illustrates the relations between market participants.

Figure 3. Structure of the broadband market. Under supply-side non-neutrality, the last-mile ISP can charge the CSP a termination fee (T) for access to its customers, in addition to charging end users a subscription fee (S) for Internet access. Depending on regulation, the termination fee may be negotiated between the ISP and CSP and may differ across CSPs. Under demand-side non-neutrality, the end user's broadband connection and subscription fee may be differentiated based on QoS or access level (Section 4.1). The CSP generates revenue from membership fees (F) or, increasingly commonly, advertisements (A). Not shown in the figure are interconnection payments between ISPs or between the CSP and ISPs.



Now what we have is a 3×3 matrix of possible combinations of discrimination and price regulation policies, amounting to a set of nine both discrimination and price regulation-spanning regimes in total; Table 1 lists these regimes. Accounting for the alternative and complementary network and pricing policies listed in the table footer, we end up with way over 100 theoretically possible regimes in total. To complicate things a bit, a whole another issue is how the policymaker finds an appropriate level for a price cap if it deems one socially beneficial. Alas, I cannot analyze the implications of all feasible regimes in detail. But I do discuss some of them in more depth along with methods for constructing more specific policy instruments later (Section 5.3), at which point we have the appropriate context for a policy analysis.

Table 1. ISP regulation regimes. The table below cross-tabulates ISP regulation regimes with respect to the degree of freedom in prioritization and pricing. On one hand, the ISP may be allowed to (i) "manage" traffic based on the traffic class (video, audio etc.) or (ii) discriminate data based on the specific CSP (Netflix, Skype etc.). On the other hand, the ISP may be allowed to set freely (i) the subscription fee, (ii) termination fee, or (iii) both. Data discrimination is inclusive of traffic-class prioritization. Extensions to the regimes are listed below the table. Expanded upon Krämer et al. (2013).

Pricing Policy / Network Policy		ISP Free to Set...?		
		None	Subscription Fee	Subscription Fee + Termination Fee[3]
ISP Free to Prioritize...?	None	Network Regulation (NR)	Network Neutrality (NN)	Network Regulation w/ Two-Sided Pricing[4] (NRT)
	Traffic Class[1,2]	Network Management w/ Price Regulation (NMR)	Network Management w/ Zero-Price Rule (NMZ)[*]	Network Management w/ Two-Sided Pricing[4] (NMT)
	CSP[1,2]	Data Discrimination w/ Price Regulation (DDR)	Data Discrimination w/ Zero-Price Rule (DDZ)	Network Non-Neutrality (NNN)

[1] Alternative network policies: (1) Throttling OK / (2) not OK (see remark next page)
[2] Complementary network policies: (i) minimum QoS; (ii) neutral slow lane for end users; (iii) zero-price slow lane for CSPs
[3] Alternative pricing policies: (A) Discriminatory termination pricing:
(a) unregulated termination fees; (b) "fair" termination fees reflecting ISP's capacity provisioning costs; (c) non-zero termination fee cap; (d) total price cap on sum of subscription and termination fees
(B) Uniform termination pricing: (a) uniform termination fee across CSPs, i.e. weak content-provider neutrality (Gans 2014); (b) uniform termination fee with cap; (c) uniform termination fee with total price cap
[4] Regimes also referred to as "no-exclusivity rule" (Kourandi et al. 2014)
[*] Closest to *status quo*

I have thus decomposed the neutrality question into two practically related but conceptually separate issues: Data discrimination and price capping. This contrasts with treating the degrees of neutrality as price discrimination (cf. Gans 2014), a more general of an issue than broadband-market-specific. Technically, data discrimination can be considered either product differentiation or second-degree price discrimination. Hermalin and Katz (2007) treat neutrality as a product-line restriction (Section 5.1). For brevity, I refer to either product differentiation or price discrimination or their combination as “differential pricing.”

One should in principle pay attention to the distinction between network management and throttling as defined earlier. The objective of network management is QoS optimization, and it should not harm consumers overall. Throttling is here taken to mean prioritization independent of the network congestion state. Surprisingly, most studies do not acknowledge this distinction between congestion-based and strategic prioritization. Throttling can be implemented at both the traffic-class and CSP levels. In practice, the distinction between congestion-based prioritization and throttling is a fine line. Data-discrimination regimes can be perhaps be assumed to imply throttling, for it is currently somewhat unclear if slowing down traffic from a particular CSP can be justified as a means of congestion avoidance. Such ambiguity disappears in the event of “paid prioritization” and exclusivity contracts, which are strategic in nature. At any rate, the distinction has regulatory implications and makes modeling more challenging, which is a natural reason for the lack of granularity in theoretical models.

As one last technical remark, the traffic-class and CSP-level data discrimination regimes can be observed to correspond to the abstraction layers from Section 2.1. Under the network-management regimes, the ISP can use prioritization techniques at the logical layer for congestion avoidance and network optimization, but under data-discrimination regimes it can also apply content-layer prioritization. Neutrality at the logical layer is thus equivalent with the dumb pipe principle. While most discussions have concerned content-layer prioritization, it has been argued[6] it is actually lower-layer prioritization that endangers the dump-pipe nature of the Internet and can indirectly lead or has already led to deliberate and systematic discrimination of certain

[6] Network Neutrality. p2pfoundation.net. Accessed 16 Nov 2014.

applications and, consequently, users based on their traffic profiles. With the constant evolution of the system, deviations from the pure dumb-pipe architecture are emerging. Protocols now specify the *DiffServ* ("differentiated services") packet header field that allows for traffic management based on the QoS requirements of data streams. The ISP might still use methods overriding DiffServ-based priority designation, as it is up to the CSP at the edge of the network to implement and hence not directly controlled by the ISP.

Status and Developments

If the definition of network neutrality is tricky, its legal status is perplexing; I cannot cover it comprehensively here. The *status quo* is somewhere between the NN and NMZ regimes in Table 1. Hitherto, no widespread, explicit, CSP-level data discrimination and no termination fees has been the *de facto* policy in the broadband markets. Compared to telephone network operators,[7] ISPs are not generally regulated as strictly, and termination fees are largely unregulated (Vogelsang 2014). In the EU, ISPs are regulated mainly by national authorities, and in the US by the Department of Justice and the Federal Communications Commission (FCC). In Finland, fixed broadband providers need not obtain an operating license.[8] The broadband market is not widely price regulated, with some notable exceptions. In Germany, the broadband market is *ex-ante* cost-basis regulated and in Austria retail-minus wholesale-regulated (Götz 2013). As far as data discrimination goes, "network management," "bandwidth management," or "traffic shaping" is something ISPs already do to a degree. Disclosure and transparency of these practices is increasingly required.[9] Some regulators have enforced vertical separation on a case-by-case basis (Jamison 2012). Authorities can generally try to control the market *ex post* through competition laws. In particular, "unfair" exclusion of content or termination pricing by an ISP with a high degree of market power could possibly fall under Article 102 in EU competition law, which includes, for example, setting unfair prices and "applying dissimilar conditions to

[7] The EU regulates roaming charges, i.e. termination fees, between telephone operators.
[8] Communications Market Act. Ministry of Transport and Communications, Finland, 2011.
[9] Directive 2002/22/EC.

equivalent transactions with other trading parties" (cf. price discrimination) in its definition of *abuse of a dominant position.*

A Body of European Regulators for Electronic Communications (BEREC) and European Commission investigation report[10] paints a rather interesting picture of the state of traffic prioritization in EU-area broadband markets. BEREC and the Commission found that at least 21% of fixed broadband users and 36% of mobile broadband users in EU markets are affected by P2P throttling or blocking, and at least 21% of mobile broadband users are affected by VoIP throttling or blocking. At least Vodafone and Telefonica have planned to deviate from the zero-price rule.[11] US ISPs have been publicly suspected of discriminating traffic, most prominently retail market leaders Comcast and Verizon. Comcast has been accused of throttling BitTorrent and VoIP traffic; Verizon has been blamed for specifically discriminating Netflix and Amazon cloud service traffic.[12] Traffic prioritization, while not necessarily explicit, and in most cases of the low-level variety, would appear to be a very real phenomenon.

In response to an increasing exposure to non-neutrality during the past few years, countries have started to come up with legislation packages targeted to uphold neutrality. While still not explicitly enforced at the EU level,[13] and while EU member states' legislations are not harmonized, neutrality has entered the EU legislative agenda. Directive 2002/22/EC classifies Internet access as a "universal service." The relevant EU legislation was updated in 2009 when the Telecoms Package was adopted, and has ever since given member countries the power to specify minimum QoS requirements for broadband. Finland was the first one to jump the wagon in October

---

[10] BEREC BoR (12) 30. A View of Traffic Management and Other Practices Resulting in Restrictions to the Open Internet in Europe. BEREC report, 2012.
[11] www.telecoms.com. Accessed 11 Oct 2014.
[12] www.nbcnews.com. Accessed 8 Oct 2014. For a more comprehensive look into what sorts of "manipulation" ISPs have been accused of, see e.g. wikipedia.org.
[13] Directive 2009/136/EC reads: "[ISPs must] inform subscribers of any change to conditions limiting access to and/or use of services and applications, where such conditions are permitted under national law in accordance with Community law."

2009.[14] The Netherlands was the first EU country to go all the way to enact neutrality regulation in 2011. The EU legislation was to be updated again after the Commission voted in favor of an amendment package[15] that outlaws throttling and blocking and was expected to become effective in 2015. In March 2015, the European council decided to reassess the rules amid proposals to allow prioritization of "specialized" services with high QoS requirements.[16]

In the US, the FCC has traditionally taken more of an *ex-ante* approach to ISP regulation (Shin 2014). An interpretation of a 2010 court ruling[17] is that bandwidth throttling, blocking, or discrimination is not allowed in the presence of other viable means of dealing with congestion. Later in 2010, the FCC established a set of regulations approaching neutrality with the FCC Open Internet Order 2010, which (a) enforced transparency and (b) outlawed blocking and "unreasonable" discrimination. An eye-catching "detail" in the Order was that the second part of the order only applied in to fixed, not mobile broadband connections (Hazlett and Wright 2012; Maxwell and Brenner 2012). The regulatory momentum in the US was reversed after a 2014 court ruling that rejected the FCC's authority to apply the latter part of the order to ISPs, as they are classified under information services rather than common carriers.[18] In May 2014, the FCC launched a public comment period that garnered comments on ISP regulation from four million people. US President Barack Obama subsequently made an official statement to the FCC urging it to place ISPs under Title II of the Communications Act of 1934,[2] which would reclassify both fixed and mobile broadband as a telecommunications service and effectively preserve neutrality, viz. ban blocking and throttling. A milestone was reached in February 2015 when the FCC voted in favor of new regulation guidelines, grounded in Title II, outlawing paid prioritization.[19] However, space is given to network management, which means that gray areas may

---

[14] Decree of the Ministry of Transport and Communications on the minimum rate of a functional Internet access as a universal service (732/2009).
[15] Connected Continent legislative package. Accessed 17 Sept 2014.
[16] www.wired.co.uk. Accessed 7 March 2015.
[17] Comcast Corp. v. FCC: 600 F. 3d 642 (D.C. Cir. 2010); 08-1291 (2010).
[18] Verizon v. FCC: 740 F.3d 623 (D.C. Cir. 2014); 11-1355 (2014).
[19] FCC's Open Internet rules. www.fcc.gov. Accessed 26 Feb 2015.

remain in the legislation. In any event, the matter is far from settled, lawsuits have already been filed by ISPs, and the debate shows no signs of calming down in the near future.

The Debate

The network-neutrality debate has become fierce over the last years, in the US in particular. By now, it is a soup of emotions, confusions, and misinterpretations. As Zhu (2007) noted already years ago, "[T]he legal community originated and popularized the debate, which has since fallen victim to political and ideological polarization. [...] If the industry giants and Congress were actually neutral to this "neutrality" debate, they should have found a middle ground by now. If legal scholars understood the technicalities of the internet, they could have reached that middle ground as well." Why is there such a heated debate on neutrality in the first place? The answer is at least fourfold. The stakeholders are numerous, but include above others (1) ISPs, (2) CSPs, (3) end users, i.e. the majority of developed countries' population, and (4) the regulator. This means differing and conflicting interests.

ISPs. Deregulation has been suggested to give rise to financial gains to ISPs. The potential benefit to ISPs is thought to come from *paid prioritization*, that is, prioritization of affiliated and sponsored content over other content. Deregulation would also create new possibilities for differential pricing. Some ISPs have said they have no plans to implement paid prioritization.[20] In general, ISPs have vouched for deregulation, although not necessarily to the fullest extent. For example, a legislative framework proposal by Google and Verizon would allow for low-level discrimination (cf. Table 1), i.e. "network management" based on the traffic class, and high-level discrimination in the case of wireless connections (cf. the Open Internet Order).[21] The main argument by ISPs in favor of prioritization is technological: Due to the explosive growth of traffic, network optimization is required to maintain QoS and deliver a better customer experience. Moreover, ISPs maintain that regulation prevents them from

[20] Verizon response to US Senate. publicpolicy.verizon.com. Retrieved 3 Nov 2014.
[21] googlepublicpolicy.blogspot.com. Accessed 1 Nov 2014.

recovering their infrastructure costs. ISPs have also proposed that revenue or cost savings from prioritization be used toward new infrastructure, eventually benefiting consumers. Lately, Google, which is rolling out its Fiber broadband service, has reportedly started to support Title II regulation because it would give the company access to utility infrastructure such as local phone and cable lines.[22]

CSPs. Many CSPs and Internet application companies are neutrality proponents.[23] The aforementioned benefit from non-neutrality to ISPs is seen to come at a cost to CSPs, which would possibly have to not only pay a base termination fee but also compete to secure a competitive access to customers. The "gatekeeper" positions of ISPs could even enable them to auction off access to customers and extract much of the CSPs' surplus in the process. This would threaten the development of the content markets.

End users. There is no general-public consensus on the neutrality matter. Among many of those who support neutrality, there is a fear that the deregulated Internet experience would look nothing like today's Open Internet. The dystopia has been characterized as a "Tiered Internet," an Internet where ISPs have the power to provide tiered broadband services with different levels of access to the Internet. This could result in Internet "fragmentation" (Section 4.2), where end users end up isolated from each other due to differentiated access to the net as per ISPs' exclusivity contracts with CSPs. Allowing data discrimination could result in this sort of tiering or versioning. Subsequently, there are concerns about digital and intellectual property rights (freedom to share and reach content), privacy ("deep-packet inspection," eavesdropping), freedom of speech and censorship (filtering out content), and equality (customer discrimination).

Despite the ongoing debate, it is also probable that the majority of end users are uninformed about the concept of neutrality to begin with. A 2014 online survey

[22] blogs.wsj.com. Accessed 2 Jan 2015.

[23] Support for net neutrality has been most visibly demonstrated by the "Save the Internet" and "Battle for the Net" initiatives. An Internet Slowdown Day was organized on 10 September 2014, during which CSPs and software firms including Netflix, Reddit, Mozilla, Vimeo, and Tumblr slowed down their services or displayed a symbolic "loading" symbol on their websites. 40,000 sites reportedly participated.

indicates that 58% of US users are uninformed, 22% support neutrality, and 20% oppose it.[24] A natural language processing analysis of millions of comments during the FCC public comment periods would indicate that, excluding one libertarian group's million anti-neutrality letters, over 99% of individual comments were pro-neutrality[25] – selection bias is probable, however. A Google Trends analysis would indicate that the number of informed users has spiked in November 2014 and again in February 2015, with most informed users residing in the US and Canada; most informed European users are located in the UK, the Netherlands, and Germany.[26]

The regulator. In theoretical terms, the regulator can be seen as a *social planner* aiming to maximize total welfare. However, the regulator is also interested in the allocation of welfare and who the gainers and losers are under different regulatory regimes. In addition to static efficiency, the regulator has to contemplate the incentive effects and dynamic consequences of regulation. In practice, the policymaker uses more mundane arguments. The European Commission motivates neutrality regulation in support of the openness of the Internet and in prevention of (1) unfair traffic management practices, (2) weakening of the competition, (3) decline of innovation, and (4) potential degradation of QoS.[27] "Unfair traffic management practices" refer to (paid) prioritization.

3. Description of the Broadband Market

In this section, I describe the economics at play in the broadband market. I start by laying out some basic notions related to modeling the market. I try to focus on the broadband market rather than cover online markets in general (who could?). I discuss inter-ISP relations on the supply side, interrelations between ISPs and CSPs, and the supply and demand for Internet content insofar as these might affect the retail market.

[24] www.google.com. Accessed 18 Dec 2014.
[25] sunlightfoundation.com. Accessed 30 Dec 2014.
[26] www.google.com. Accessed 4 Jan 2015. I use web searches for "net neutrality" as a proxy for the fraction of informed users.
[27] Net Neutrality challenges. ec.europa.eu. Accessed 18 Oct 2014.

In Sections 4 and 5, then, I evaluate the effects of regulation in light of the characteristics of the broadband market.

3.1 Modeling the Market

I start by briefly presenting some fundamental concepts typically used in modeling the broadband market and, subsequently, the economic effects of network neutrality: (1) two-sided markets, (2) queueing systems, and (3) a natural monopoly, duopoly, and oligopoly.

Two-Sided Markets

The broadband market is a *two-sided market* in which two groups connect via a platform. The ISP maintains the platform, namely, Internet connectivity, which end users and CSPs use to interact with each other. The reader can refer back to Figure 3 for the structure of the broadband market, with the ISP lying in between end users and CSPs. In addition to broadband, other two-sided markets include operating systems, credit cards, shopping malls, gaming consoles, and stock exchanges (Rochet and Tirole 2006; Rysman 2009). Two-sided markets are characterized by dependence of parties upon a platform and by *network effects* (or "network externalities" or "group externalities"), in the presence of which the value of the platform to a user depends on other users (Armstrong 2006; Rysman 2009). The relative sizes of the group externalities affect the prices the platform operator charges at each side. The positive network externalities associated with telecommunications networks can be used as an argument for promoting universal service (Sidak 2006), at which public policies such as the Finnish "Broadband 2015" plan aim. The platform, e.g. a credit card, allows the seller and consumer to conduct the transaction and only has value if it is widespread enough.

Two-sided pricing or "double charging" is common in two-sided markets, but sometimes the platform does not find it optimal (Economides and Hermalin 2012). MasterCard charges merchants for each transaction but gives benefits to cardholders. On the other hand, Google Play taxes both the users and app developers by pocketing 30% of sales revenue.[28] In this light, termination fees by ISPs would not seem that out of place. Platforms in a two-sided market may enjoy special attention from the regulator concerned about customer discrimination; an example analogous to throttling by ISPs is the recent court case against Visa and MasterCard following their blocking of payments to Wikileaks.[29]

While two-sided markets have been relatively widely studied, research in the particular context of the broadband market is still rather sparse. In addition to the aforementioned papers, notable works on the mechanics of two-sided markets include Amelio and Jullien (2012), Caillaud and Jullien (2003), Eisenmann et al. (2006), Hagiu (2006), Parker and van Alstyne (2005), Rochet and Tirole (2003), and Weyl (2010). The interconnectedness of ISPs differentiates broadband from other two-sided platforms, as the customers (CSPs) directly connected to one platform (ISP) get access to all platforms (Musacchio et al. 2009). Under non-neutrality, this does not necessarily apply.

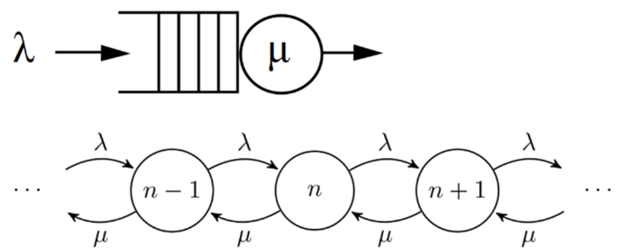
Queueing Systems

The broadband market is a peculiarity among two-sided markets also due to the existence of congestion (Economides and Hermalin 2012). Internet networks are essentially queueing systems. In most of the network neutrality studies, congestion in the network is modeled in some more or less mathematical form. The ISP's capacity provisioning, costs, and economic behavior depend on the demand for bandwidth at a given time; demand for bandwidth in turn is reflected as data streams and packets in the system. Congestion can be costly, and the ISP will generally act to avoid it. Most models specify a monopolistic or representative ISP, whose networks can most abstractly be treated as a single pipe that forwards packets. The simplest approach is to model the network as an M/M/1 queue with only one server (router). Customer (packet) arrival times follow a Poisson process, i.e. waiting times are exponentially distributed. The server uses the FIFO policy, where packets are forwarded in the order

[28] google.com. Accessed 13 Jan 2015.
[29] rt.com. Accessed 13 Dec 2014.

they arrive (cf. the dumb pipe). The number of packets in the system (in the trillions or so) is then a stochastic process with state space $\{0, \ldots, n-1, n, n+1, \ldots\}$, which can be normalized with respect to some reference state. Most neutrality studies employ the M/M/1 specification. One can generalize the M/M/1 system to e.g. the M/M/c system, where c denotes the number of servers, or all the way to the G/G/k system, where arrival and service times have arbitrary distributions. Continuous, real-valued analogs to queueing processes are studied in the field of fluid models. In a broad context, these sorts of models are not usually worth the extra complexity, and hence not germane to the analysis. Figure 4 illuminates the logic of the M/M/1 system.

Figure 4. M/M/1 queueing system. The system consists of a queueing node with a waiting area (top left) and a service node (top right), and evolves according to the state transition diagram (bottom). At a given moment, there are n customers/packets in the system. λ is the mean arrival rate and μ is the mean service rate; similar notation is usually used in network neutrality models. Hence, $\lambda/\mu \equiv \rho$ is the utilization rate of the server (load), $\lambda/(\mu-\lambda) \equiv \bar{n}$ is the expected number of packets in the system, $\rho\bar{n} \equiv L$ is the expected queue length, $1/(\mu-\lambda) \equiv W$ is the expected total time spent in the system, and $\rho W \equiv w$ is the expected waiting time. Congestion occurs if, for some time span, $\lambda > \mu$, i.e. load exceeds capacity and packets start to accumulate in the waiting area. Figure sources: Users "Gareth Jones" and "Tsaitgaist," Wikimedia Commons, December 2014.



In the language of economics, modeling the network as a queueing system means treating the demand for bandwidth stochastically and incorporating congestion into the model. Congestion-inclusive frameworks can also describe regimes where throttling is not allowed, i.e. traffic is only prioritized in response to congestion. Models differ in whether the ISP is bound by a capacity constraint. Typically, the capacity of mobile networks is lower than that of fixed networks. Choi et al. (2014) in particular make this explicit in their model.

Assuming deterministic demand and no congestion can be insufficient for a realistic model; congestion, demand volatility, and uncertainty are essential features of the broadband market. A subsequent question is how to model the subjective disutility to the customer from waiting. Traditionally this *delay cost* has been assumed constant (Afèche 2006). (It can be argued that this is not realistic; perhaps after a certain waiting time disutility begins to increase faster.) Krishnan and Sitaraman (2012) estimate that after a two-second startup delay, each one-second delay increment increases video stream abandonment rates by 5.8%. Maister (1985) provides more insight into the psychology of waiting lines, very well applicable to virtual ones. Kleinrock (1967) shows that the optimal bribe – the extra payment for a "priority lane" to reduce the waiting time – is monotonically increasing in the customer's "impatience factor" for all Poisson-arrival single-server queues.

Natural Monopoly, Duopoly, and Oligopoly

Due to substantial economies of scale (quantity produced) and scope (product mix), telecommunications networks are often considered a natural monopoly. Natural monopoly can be defined as a scenario in which

$$C(Q) < C(q_1) + C(q_2) + \ldots + C(q_k), \quad (1)$$

where $C(Q)$ is the cost of producing output $Q \equiv \sum_{i=1}^{k} q_i$ of a homogenous good (Joskow 2007). Thus, in the case of a homogenous good, the situation falls under this definition of a natural monopoly whenever there are economies of scale over a relevant range of total output. A broadband connection is not far from a homogenous good, although it can be differentiated in speed or price or be bundled with other services. Much more differentiation would be possible under non-neutrality. It should be recognized that there are alternative approaches to defining a natural monopoly. In any case, in the same way than with highways or power transmission lines, it usually makes no economic sense to firms to build two next to each other. Even though there is competition at the macro level, locally the end user's alternatives are limited, especially so in periphery areas.

Most natural monopolies are regulated in some ways. The firm may also be state-owned. Monopolies or dominant firms are closely monitored by competition authorities for abuse of a dominant position, defined in Article 102 in EU legislation. The argument for natural-monopoly regulation can be condensed into prevention of socially costly market failures that arise from the poor economic performance of naturally monopolistic industries (Joskow 2007). In general, industries that exhibit natural monopoly characteristics, such as the broadband market, tend to become concentrated. Then, the usual warnings against a monopoly and weak competition apply. The policy instruments in turn are heterogeneous. The regulator can control prices, entry, and terms and conditions of service through price or profit ceilings and floors and operating licenses. Price regulation is often cost-based, and pricing may be either linear or nonlinear. One can set either $P = AC$ (linear pricing) or $P = MC$ plus fixed fee S covering fixed costs (two-part tariff). In the telecommunications industry, S would represent a large fraction of the total tariff. Subject to a break-even constraint, *Ramsey pricing* (Ramsey 1927) gives the socially optimal price choice for the monopolist:

$$\frac{P-C}{P} = \frac{\lambda}{1+\lambda}\frac{1}{\varepsilon}, \tag{2}$$

where λ is the shadow cost and ε is the price elasticity of demand. Hence the optimal price is inversely proportional to the elasticity of demand. The theoretical optimality of Ramsey prices would also apply to a monopolistic upstream Tier 1 or 2 ISP when downstream competition is perfect (Vogelsang 2003). As we know, it is usually not, in which case optimal pricing becomes more elaborate.

Another cost-based regulation scheme is a "yardstick"-based price cap (Shleifer 1985), where firm i sets

$$p_i = \frac{1}{N-1}\sum_{j \neq i} c_j\,, \tag{3}$$

where N is the number of locally monopolistic firms and c_j is firm j's marginal cost. In words, the optimal price is equal to average marginal cost to other firms operating in the same product market, acting as a cost benchmark. The most obvious challenge in cost-plus regulation schemes is incentivizing firms to keep costs in line. Cost-frontier- and performance-based incentive regulation schemes have been devised to this end, where cost leaders act as the benchmark instead of a simple industry average. A quite complex derivative of the yardstick principle is the revenue-cap benchmark regulation of Finnish electricity distribution firms.

When not monopolistic, broadband markets usually exhibit oligopolistic characteristics. Oligopolies, as potentially dominant firms, are monitored in the EU under Articles 101 and 102; Article 101 outlaws *collusion*, tacit or explicit. The two main theoretical approaches to modeling oligopolies are Cournot (quantity) competition and Bertrand (price) competition. Through an undercutting argument, Bertrand competition with only two firms can be seen to lead to an equilibrium with prices equal to those in perfect competition. In very general terms, the Cournot (Bertrand) model is the suitable one when capacity adjustment is difficult (easy). In the short run, ISPs can purchase transit and make small-scale capacity adjustments; large-scale infrastructure investments can take years. Since with both models a duopoly setting yields results in most cases easily generalizable to an oligopoly with N firms, a duopoly is the most popular alternative to a monopoly in broadband market models. It is also quite realistic in our context. In some models the game is of the Stackelberg type, where one firm makes its move first. The Cournot and Bertrand models can also be used concurrently, as is in a sense done in e.g. Njoroge at al. (2009), where two ISPs first set quality ("quantity") levels and then compete in prices. Quality can be taken to mean QoS, i.e. features such as the connection speed, or, in the non-neutral world, access level. In general, dynamic games such as the Stackelberg game can be solved by finding the subgame perfect Nash equilibria (SPE), most commonly using backward induction.

Particularly popular in the broadband-market context is the Hotelling (1929) model, in which firms are located on a line, construed to reflect either the geographical location or product characteristics. The setting can be monopolistic, monopolistic competition, duopolistic, or oligopolistic. A continuum of consumers is uniformly distributed on the line, again reflecting either a geographical map or a "preference space." Firm and consumer locations can be thought to reflect product properties and preferences, respectively. The unit "transportation cost" from moving along the line can be taken to reflect the degree of product differentiation (Choi and Kim 2010). In a non-neutral network, an ISP's location might reflect the particular content offering its broadband connection enables consumers to reach. The Hotelling formulation is applied in some form in most of the mathematical neutrality-related models.

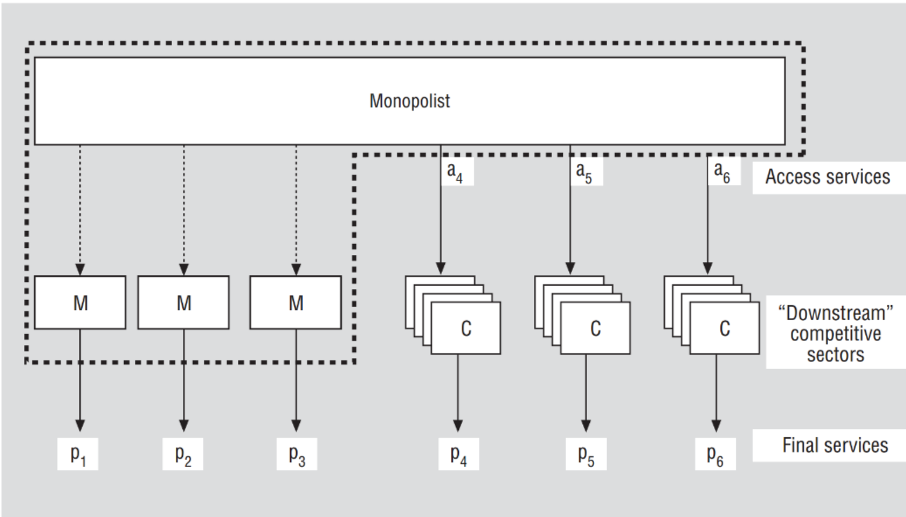
3.2 Market Structure

The underlying supply-chain structure of the broadband market can be seen to be reminiscent of other markets with manufacturers (CSPs), wholesalers (transit ISPs and carriers), and retailers (residential ISPs) (Yoo 2006b). As described earlier, the market has the tendency to be locally monopolistic or oligopolistic. The barriers to entry are high due to the high upfront investment cost. In most legislations, in the EU in particular, the local loop is *unbundled*: ISPs share last-mile access from the local exchange point to the end user's premises and are free to compete for broadband provision to the user. Alternatively, the incumbent ISP can grant entrants *bit-stream access* to its equipment installed at the end user's premises – a "handover point" this low in the topological hierarchy is not mandated at the EU level, however (Leal 2014). It has been argued that mandatory unbundling facilitates competition, but this is not always true, as the incumbent ISP can still control much of the local traffic, and having to share access can hinder incentives for market entry (Cambini and Jian 2009; Hogendorn 2007; Wallsten and Hausladen 2009). Indeed, a rule of thumb is that an end user has one to three ISPs to choose from. A handful of large ISPs sometimes enjoy substantial market power in a geographically vast market area.

Characteristic to the broadband and inter-ISP markets has been a strong hierarchy. While local ISPs may have significant market power over end users, larger transit ISPs and backbone carriers may have significant power over local ISPs, which have to purchase network access from them. Traditionally, access prices in the EU and US have been cost-based with small common markups (Vogelsang 2006) and are regulated in some areas (Bourreau and Lestage 2013). Still, small players may not have much negotiation leverage. The situation is referred to as the *one-way access problem*

(Armstrong 2002; Vogelsang 2003). Figure 5 delineates this kind of a market structure. During the last years, however, the industry has started to evolve toward a "mesh," with also smaller ISPs directly connecting between each other (Besen and Israel 2013). Power over of the principal data routes remains concentrated nonetheless.

Figure 5. One-way access problem. A monopolistic upstream network operator provides network access to smaller downstream operators that in turn provide the final services. The monopolist may be vertically integrated and provide final services directly in the retail market. Source: Access Pricing in Telecommunications. OECD Competition Committee report, 2004.



Large ISPs can be vertically integrated in two senses of the term. On one hand, the same firm can operate both as a local residential ISP and as a global carrier. In the US, AT&T and Verizon operate both as last-mile and backbone providers. TeliaSonera is a leading residential access ISP in the Finnish market and the second-largest Tier 1 carrier in the world.[30] On the other hand, the ISP can not only provide Internet access but also act as a CSP or be closely affiliated to CSPs. Further, ISPs can be horizontally integrated and act as telephone operators, for example. A large and integrated telecommunications firm can offer a broadband, cable television, streaming, and telephone subscription in the same bundle. Vertical integration does not necessarily reduce welfare and can solve issues such as double marginalization. Bandyopadhyay et al. (2010) find that vertical integration of an ISP can be beneficial in the short term but hurt competition in the content market.

[30] TeliaSonera International Carrier. research.dyn.com. Accessed 12 Nov 2014.

The broadband market is characterized by high switching costs. Switching an ISP can be both expensive and time-consuming to the end user due to new equipment needed with the new connection, the hassle of terminating the old contract, delay in the activation of the new connection, and so on. High switching costs aggravate the power an ISP can have over end users, and may result in vendor lock-in situations. The situation is more difficult in periphery areas. In some countries, such as Finland, there are government plans to require or incentivize ISPs to connect properly to periphery households, as it may not otherwise be economically profitable. In general, coverage and penetration of high-speed broadband access is lower in segregated areas.

The broadband industry is highly analogous to electricity distribution: CSPs are comparable to electricity production firms and ISPs are comparable to electricity distribution firms. Like electricity distribution, fixed costs are high relative to marginal cost. Like electric power, bandwidth is not storable in the sense that any exceed supply at a given moment would contribute to inventories for the future; the output has to, at the least, equal the demand at all times, or else there will be blackouts and dropped connections. Another similarity between the markets is the prevalence of a zero-price rule: Local electric distribution operators charge end users but not appliance manufacturers (Hemphill 2008). In the electricity context two-sided pricing would admittedly seem impractical at the very least. The crucial difference between electricity distribution and telecommunications is that usually the principal national electrical grid is operated by a single transmission system operator and local endpoints of the distribution network by private distribution system operators. The analogous policy in the telecommunications market would then be that the principal data routes were controlled by national operators. The Internet is not centrally governed, however, and the backbone has been largely privatized.

Perhaps more tangibly, a telecommunications network can be thought of as a physical road (Crocioni 2011), a classic case of a *negative externality*. Each car contributes to congestion, a social cost not internalized by drivers. As a result, the amount of traffic may be too large at the societal level. The textbook solution to this “tragedy of the commons" is to impose a Pigouvian tax equal to the difference of the estimated social cost and private cost. The ISP – the road operator acting as a local government – would like to tax the drivers, CSPs. Prioritization plays a role in the road traffic analogy, as well, in the form of fast lanes for taxis and buses.

Finally, one could reach out to draw an analogy to postal services, which often differentiate delivery options based on not only package size and weight but also destination and content type.

Interconnection

Behind the scenes, in the "wholesale" Internet interconnection market, business relations between residential ISPs, carriers, and CSPs can get messy. ISPs face a decision problem between peering and transit agreements. Peering differs from transit in that under a peering agreement between ISP A and B, ISP A has no obligation to terminate ISP B's traffic to or from a third party (Jahn and Prüfer 2008). While in a peering agreement the ISPs use each other's networks reciprocally, peering is not necessarily free for both parties. Peering agreements where no settlements are paid are sometimes referred to as "settlement-free peering" or "bill-and-keep peering," whereas those involving settlements are referred to as "paid peering" or just "peering" (Jahn and Prüfer 2008). Transit comes in many forms, as well: Full transit (access to all routes), partial transit, and access to specific routes. In settlement-free peering, the loads the ISPs exert on each other's networks are usually quite symmetric. Under asymmetry, the larger network will theoretically prefer a reciprocal fee on peering, set equal to cost (Carter and Wright 2003).

When an end user streams video from Netflix to her computer, the video stream goes through carriers, e.g. Cogent, before reaching the end user's, e.g. a Comcast customer's, residential end node. When many of its customers start using Netflix and increase their traffic volume, Comcast may have to purchase additional capacity from Cogent and incur additional costs. Else, Comcast's gateways may become congested and QoS degrade for its customers. Nothing, however, prevents Comcast and Netflix from signing an interconnect agreement where both invest in a new interconnection point directly connecting Netflix servers to a Comcast network, thus bypassing Cogent networks altogether. The benefit to Comcast is reduced costs (since transit is relatively expensive) and less congestion, and the benefit to Netflix is an improved customer experience.[31] Today, it is not rare anymore for CSPs to pay ISPs for enhanced network access. While these direct interconnect agreements do not fall under the scope of the mainstream neutrality debate, the case becomes relevant from the debate's perspective if it turns out that Comcast has either (a) throttled Netflix traffic or (b) threatened to throttle Netflix traffic to secure an interconnection deal where Netflix pays an interconnection fee to gain direct access to Comcast customers. The latter scenario could then in effect constitute termination pricing. While the terms of actual Comcast–Netflix and Verizon–Netflix agreements from February and April 2014 remain undisclosed, the FCC is investigating the matter.[32] Netflix streams reportedly sped up by some 65% on average for Comcast customers after the deal.

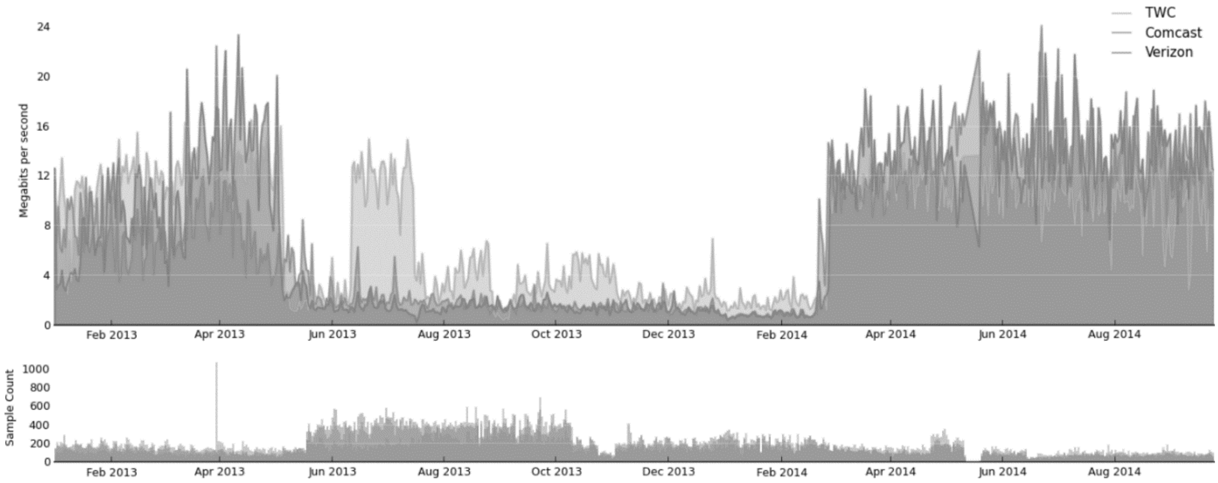
Although happening behind the scenes, ISP interconnection agreements indeed have real effects on the QoS observed by end users. Recalling the one-way access problem, QoS by residential ISPs may be constrained by the carrier network through which they access the backbone: Congestion can occur on the carrier side, as empirically illustrated by Figure 6. Notably, QoS degradation is not always as much a result of a true technical limitation as of inter-ISP business relations.

A CSP may itself operate as a platform in a two- or multi-sided market. These kinds of CSPs can be called *content network platforms* (CNPs), and they act as intermediaries between end users and other CSPs (Mialon and Banerjee 2014). Hence also content markets have hierarchical structures. Fuelled by the explosive growth of Internet traffic with high QoS requirements, another particular business model has emerged: A *content delivery network* (CDN) that interconnects a CSP to backbone ISPs to ensure high-performance delivery of content, acting as an intermediary between the CSPs and ISPs. The use of a CDN can be more cost-efficient for the CSP than directly connecting to

[31] QoS also depends on the capacity of the CSP's own servers. The source can be self-limiting, i.e. the data rate may be constrained at the CSP's end.
[32] gigaom.com. Accessed 13 Nov 2014.

multiple ISPs – the latter strategy is called *multi-homing*. The largest, most data-intensive CSPs, Netflix and Google, have recently invested massively in their own in-house CDNs.[33] There can be substantial economies of scale, as Netflix and YouTube account for approximately 50% of US end users' Internet traffic. Other leading CDNs include Akamai, Limelight, and Level 3.

Figure 6. Carrier as a bottleneck. TWC, Comcast, and Verizon end users in the NYC area experienced substantial QoS degradation (decreased median download throughput) when connecting through a Cogent network during May 2013 – February 2014, after which Cogent apparently rerouted traffic, increased capacity, or allocated more bandwidth to the three ISPs. Cablevision customers connecting through the same network experienced no QoS degradation. Source: ISP Interconnection and its Impact on Consumer Internet Performance. Measurement Lab Consortium report, 2014.



Characteristics of the Finnish, EU, and US Markets

The Finnish Communications Regulatory Authority (FICORA, fin. Viestintävirasto) oversees ISPs in Finland. The main last-mile fixed broadband technologies in use in Finland in 2013 were, in order of popularity, xDSL, cable modem, Ethernet (optical fiber), FTTH ("fiber to the home"), and housing cooperative broadband. The share of optical fiber technologies is increasing while that of DSL is declining. In the EU, FTTx represented less than 5% of the market in 2011, compared to about one-half in e.g. Japan and South Korea.[34] The number of mobile broadband connections in Finland

[33] www.businessinsider.com. Accessed 8 Dec 2014.
[34] Insights on the European Telecoms Market: Analysis, Forecasts and Commentary. Telecoms Market Research report, 2011.

increased by 600% between 2008 and 2013, and the majority of new broadband subscriptions are wireless.[35] As of June 2014, there were 1.32 mobile broadband subscriptions (more than in any other country in the world) and 0.31 fixed broadband connections per person in Finland.[36] Despite high broadband penetration, the Finnish broadband market exemplifies the "one to three ISPs to choose from" rule as an oligopoly where three market leaders share most of the market. In 2013, 85% of the fixed broadband connections and 99% of the mobile broadband connections in Finland were provided by three ISPs: Elisa, TeliaSonera, and DNA. Smaller ISPs in Finland include 24 local operators under the Finnet group with a 12% total market share in fixed broadband and a number of small, local, or specialized players. Close to 200 Finnish firms or subsidiaries have submitted a telecommunications notification to FICORA, many of these inactive.[37] The fact remains that the Finnish market is the most concentrated in the EU (Calzada and Martínez-Santos 2014). In this light, it is surprising that the PPP-adjusted prices in Finland are below the EU average; the history of municipal ownership, advanced technological infrastructure, and government support schemes provide partial explanation. In 2008, the Finnish government initiated a "Broadband 2015" project to ensure high-speed broadband access in sparsely populated areas. Expected public subsidies total €130m.[38]

Anecdotal evidence suggests that the US broadband market does not function as well as the EU markets. The limitations in US end users' ISP choice are severe: Less than 10% of US consumers can choose between more than two ISPs in the case of a 10 Mbps connection; the percentage is much lower for faster connections.[39] This may in part be due to the lack of local-loop unbundling in the US since around 2005 (Hogendorn 2007). Title II reclassification would not directly reinstate local-loop unbundling, either. The US wireless broadband industry is highly concentrated (Rosston and Topper 2010).

[35] Toimialakatsaus 2013. Finnish Communications Regulatory Authority report.
[36] OECD Broadband Portal. www.oecd.org. Accessed 14 March 2015.
[37] www.viestintavirasto.fi. Accessed 14 Jan 2015.
[38] www.viestintavirasto.fi. Accessed 29 Dec 2014.
[39] NTIA State Broadband Initiative. Accessed 30 Dec 2014.

The situation is aggravated by general customer dissatisfaction; Comcast won the Consumerist's "Worst Company in America" award in 2014.[40] The proposed merger of Comcast and Time Warner Cable, which share more than two-thirds of the US broadband cable market and approximately 40% of the US broadband market as a whole, would make the broadband product market still more concentrated (although the companies maintain that they operate in separate geographical markets). As it happens, the merger has direct relevance to the neutrality debate, as Comcast has agreed to extend their commitment to the Open Internet Order to span the whole of Comcast-TWC as a merger remedy, although only until 2018.[41] Relative to their EU peers, US consumers have paid more for an equivalent connection.[42] For example, in Kansas City, MO, in 2013, the least expensive 10 Mbps connection reportedly cost $112 per month.[42] For comparison, in Turku, Finland, an equivalent connection cost €20 per month. However, Kansas City is better off now with Google Fiber available there along with a dozen other locations in the US. The Fiber is a noteworthy development in the US market because 5-Mbps access is free after a $300 construction fee, and 1 Gbps costs $70 per month.[43] The FCC's National Broadband Plan has since 2010 subsidized infrastructure investment to improve broadband access in the US, but a 2015 report finds that "broadband [at least 4 Mbps downstream] is not being deployed to all Americans in a reasonable and timely fashion."[44] OECD does report an over 100% wireless broadband penetration in the US.[36] OECD gives $44 as the PPP-corrected US average monthly price in 2012 for fixed connections over 2.5 Mbps in speed – the sixth most expensive (behind Turkey, Spain, Chile, Norway, and Luxembourg) in the 34-country sample. In Finland, the average price was $26.[45]

---

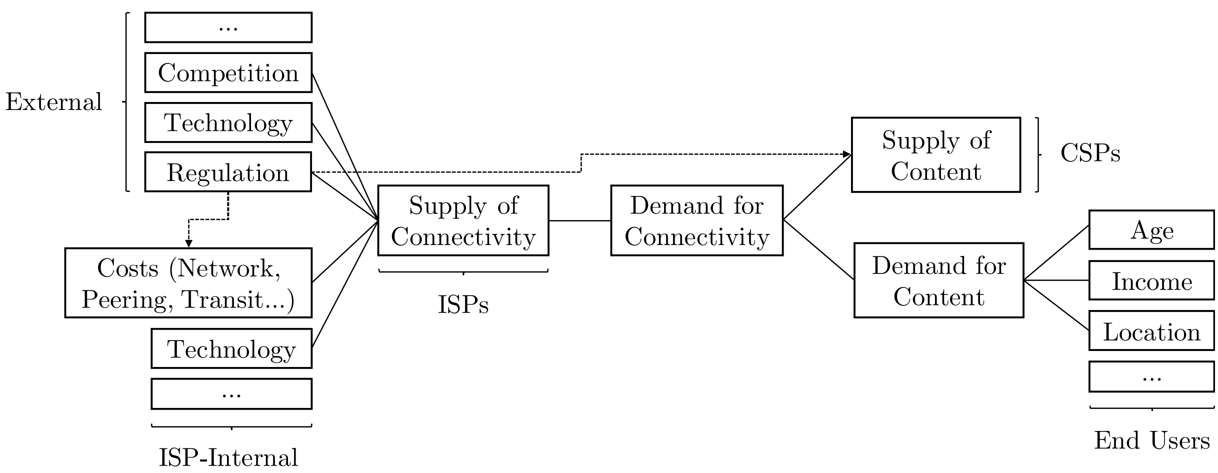
[40] consumerist.com. Accessed 12 Jan 2015.
[41] Comcast response to US Senate. www.franken.senate.gov. Retrieved 19 Dec 2014.
[42] The Cost of Connectivity. Open Technology Institute, New America foundation report, 2013, 2014. Accessed 30 Dec 2014.
[43] fiber.google.com. Accessed 2 Jan 2015.
[44] arstechnica.com. Accessed 8 Jan 2015.
[45] OECD Communications Outlook 2013.

3.3 Supply and Demand for Connectivity

The supply and demand for broadband services, viz. Internet connectivity, are related to factors on the ISP side as well as on the end-user side. Supply is determined by both ISP-internal factors such as cost structure and technological developments. Moreover, the ISP's external environment, including the competitive and regulatory landscape affect supply decisions. Network neutrality has also direct effects on the ISP's revenue and costs.

On the end-user side, the demand for connectivity depends on a number of factors. Theoretically, the fundamental factor behind the demand for connectivity is the demand for Internet content. The ISP is merely the intermediary between end users and content, and would add no value if there were no CSPs on the Internet. As a typical firm acts both as a set of end-users (Internet access at workplace, cloud systems, the industrial Internet etc.) and a CSP (company websites, online services etc.), it contributes to the demand for connectivity through both its downstream demand for broadband access and upstream supply of content. Market prices of broadband connections ultimately determine "quantities" demanded when moving along end users' demand curves. Figure 7 breaks down an economic framework for the broadband market.

Figure 7. Framework for the broadband market. Supply and demand drivers determine broadband prices, QoS, penetration, and coverage.



Supply

At the general computer-networking level, data transport services can be categorized into (1) guaranteed or "dedicated," (2) best-effort, and (3) flexible bandwidth-sharing services as in Afèche (2006). Most retail contracts and Internet interconnect agreements fall under the best-effort category. Indeed, retail contracts list the connection speed as "up to x megabits per second." Dedicated bandwidth usually comes with service-level agreements (SLAs). A typical household connection is asymmetric with less upstream than downstream bandwidth allocated to it by the ISP (e.g. a download-to-upload speed ratio of 10:1), due the fact that a typical end user uploads less than downloads.

It is extremely unlikely that all users use 100% of their bandwidths simultaneously. Therefore, analogously to banks', insurance companies', or airlines' operation relying on the law of large numbers – banks do not expect everyone to withdraw their holdings simultaneously; insurance companies do not expect everyone to have an accident at the same time; airlines expect a fraction of people to cancel their tickets – the ISP can *oversubscribe* by allocating users nominal bandwidths under the expectation that aggregate load averages out to a level below the sum of nominal allocations. Hence, the allocation is merely virtual, as in reality packet switching allows data streams to be statistically multiplexed together, that is, intermingled to reduce slack. The oversubscription ratio (or "contention ratio") may range from something like 5:1 to anywhere over 100:1 depending on the case. Oversubscription is becoming more challenging with many users streaming video, which requires high and constant QoS in contrast to "bursty" web surfing traffic that averages out across users. The fact that the aggregate load on the network evens out to a degree does not mean it would be deterministic: Far from it, large stochastic fluctuations remain. Even with oversubscription, or precisely because of it, networks tend to be lightly utilized relative to their theoretical capacity – this is *overprovisioning*. Utilization rates around 5% on average and around 25% at peak loads (Afèche 2006) reflect the redundancy in the networks.

An overarching characteristic of the broadband business is that fixed costs dominate the total cost structure (Lyons 2013). Fixed costs include the sunk cost of upfront investment in network infrastructure, operating and maintenance costs, and fixed interconnection contracts. Sunk investment is made continually over time (Sidak 2006). The marginal cost of transmitting a packet is small but nonzero, for there are peering and transit, maintenance, and other costs that depend on the amount of data transmitted. As the time horizon is stretched, fixed costs such as infrastructure investment become variable. In the short run, peak loads exceeding the network capacity force the ISP to purchase external capacity. Strictly speaking, marginal cost as a function of the total amount of data transmitted by the ISP can jump at the point where initial capacity runs out and the ISP has to purchase more. Public estimates are difficult to come by, but two of these put the ISPs' average marginal cost from fixed broadband traffic in the order of €0.01 per GB,[46;47] and another puts the transit price in the order of €1 per Mbps per month[48] (note the units) depending on location. (Also note that, due to oversubscription, in practice 1 Mbps to the ISP translates to more than 1 Mbps to an end user.) For mobile traffic, costs are likely to be an order of magnitude higher. Costs have decreased drastically over the past years especially with wireless technologies and will continue to do so. In the case of mobile broadband in the EU, the prices ISPs charge end users for additional data over a data cap range somewhere from €0.1 per MB to €1 per MB; the unit price is generally lower than this when one purchases a larger bundle at once.

If traffic increases above the ISP's initial capacity, capacity needs to be added through transit or peering (in the short run), or equipment and infrastructure installations (in the long run). Large infrastructure projects, e.g. optical fiber installations, can cost hundreds of millions and take years to complete. Peering calls for larger volumes than transit and becomes cost-efficient at the point where the unit price of peering – inclusive

[46] Delivering High Quality Video Services Online. Analysys-Mason report for Ofcom, 2009. Retrieved 2 January 2015.

[47] The Cost of Incremental Internet Transit Bandwidth in the Local Access Cloud. Lemay-Yates Associates report for Netflix, 2011. Retrieved 2 January 2015.

[48] www.drpeering.net. Accessed 21 Jan 2015.

of the cost from the extra load on the network from the other ISP's traffic – decreases to equal the unit price of transit.[48] Transit is usually priced per megabits per second (Mbps) or gigabits per second (Gbps) per month, sometimes based on allocated bandwidth regardless of whether the customer ISP uses it all. Transit can be priced *ex ante* or *ex post*. Among larger ISPs metered services are reportedly more common.[48] Metering is typically based on "burstable billing," most typically the 95-percentile method, where the transit provider calculates the average bandwidth consumption in five-minute samples and discards the top 5% of the samples for anomalies.[49] Unsurprisingly, this has provoked strategies where the customer utilizes maximum bandwidth for 5% of the time and minimally for the rest of the time. In an unmetered contract, the ISP purchases fixed "transit bundles" for lump-sum fees and, consequently, near-zero intra-bundle marginal costs. Transit often comes with significant volume discounts (Faratin et al. 2008). Pricing may be regional and customized with different sorts of side deals. The customer often has to commit to a certain bandwidth and contract period at once. A typical transit agreement stipulates best-effort delivery. Overall, not much public information is available on interconnection pricing practices.

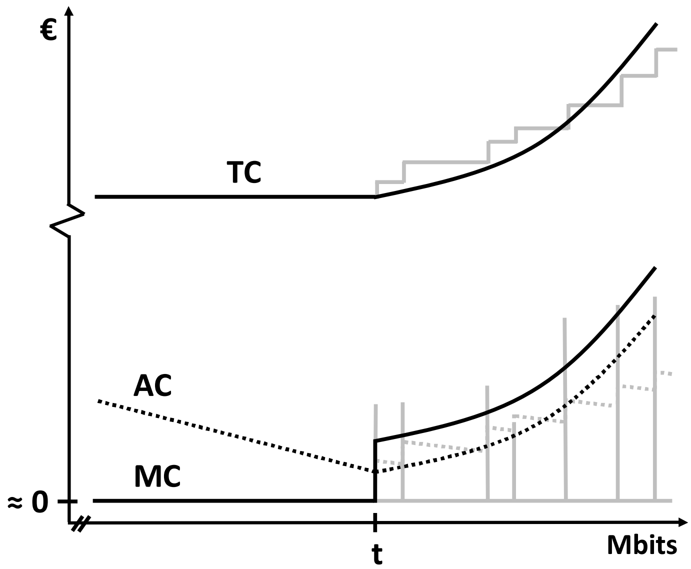
The ISP's overall cost profile depends heavily on the ISP's positioning in the interconnection ecosystem. An estimation of per-user infrastructure cost in wireless networks by Johansson et al. (2007) gives a constant per-user cost up to a certain volume of downloaded data; after this "congestion threshold," the cost increases exponentially. Figure 8 roughly represents a hypothetical ISP's marginal, average, and total cost functions with respect to the volume of data transmitted.

A significantly easier choice when modeling the market is to stick to the assumption of constant or even zero marginal cost; this is done in most of the neutrality-related theoretical models. In reality, marginal cost factors in the ISP's routing and pricing decisions. An essential realization, following Laffont and Tirole's (2000) point, is that marginal cost can be calculated by taking the first derivative of the total cost function with respect to not only the amount of data transmitted – which can be measured as

[49] wikipedia.org. Accessed 22 Jan 2015.

average bandwidth usage over some time interval or some other metric – but for example the number of active users or QoS. Moreover, the choice of units and the time scale crucially affect the look of the cost functions. What the most meaningful cost accounting method or cost formula is when calculating variable costs in telecommunications remains an open question. Still, an understanding of the cost structure in the broadband industry is essential for modeling and regulation purposes.

Figure 8. Marginal cost, average cost, and total cost functions for an ISP. Marginal cost (MC), average cost (AC), and total cost (TC) are plotted against the total amount of data transmitted. In the short run, transit and peering is variable but large-scale infrastructure remains fixed. If the values on the horizontal axis are updated over time as infrastructure is gradually upgraded in response to increasing Internet traffic, the functions can be interpreted to apply also to the long run. Marginal cost is close to zero up to a congestion threshold (t), on which it starts to increase. When transit is priced in larger bundles relative to traffic volumes, cost functions are more "discrete" (gray graphs).



Demand

The technical specifics of a connection are not typically of interest to the end user who is, at best, concerned with the general QoS requirements: Bandwidth, latency, and loss (Afèche 2006). Specifying customers' requirements is relevant to modeling the market because QoS metrics have an effect on consumer utility and the demand for connectivity. Congestion in the network is reflected as a decrease in QoS and is thus of interest for modeling end-user demand.

Rosston et al. (2010) empirically evaluate the demand for broadband Internet services in the US, and choice-experiment on US consumers' willingness to pay for improvements in service levels. First, they find that at least the income level, age, geographical location, and online "skills" shape residential end users' marginal willingness to pay for a faster connection. The difference between the perceived value of a "very fast" and a "fast" connection, vaguely as they are defined, is found to be insignificant, whereas the difference in value between a fast and a "slow" connection is significant. In other words, the expected marginal utility from connection speed is decreasing and would seem to be zero with faster connections. Higher-income households value a fast connection more than lower-income households do. Young households value speed and reliability more than older ones. Likewise, the level of education correlates positively with the willingness to pay for speed, but negatively with the willingness to pay for reliability. Urban households value reliability more than rural ones. Finally, better online "skills" imply a higher willingness to pay for speed. Nurski (2014) estimates that increasing the connection speed of a given broadband service (out of multiple options) by 1% increases the market share for the given service by 1.5% in the UK. The estimated UK price elasticity of demand is -3.4, so the demand for connectivity appears to be quite elastic.

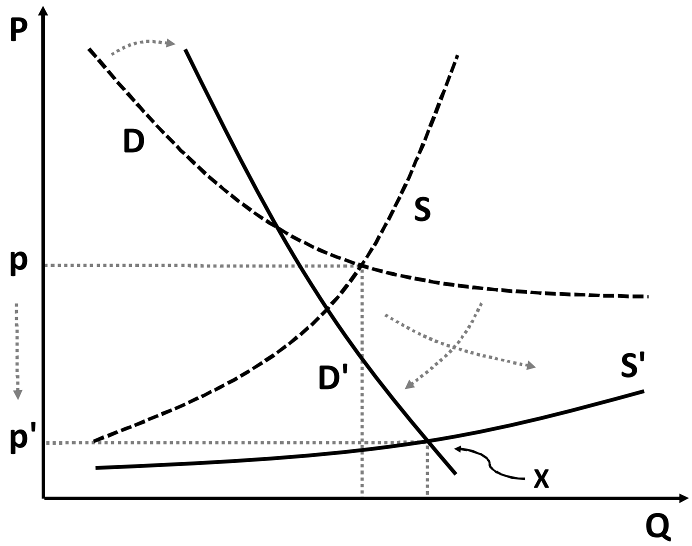
3.4 Supply and Demand for Content

Internet content can be thought of as the fundamental factor of the demand for connectivity. However, while content and connectivity are two distinct commodities, they are only meaningful products in relation to each other, and, as such, it is problematic to isolate the demand for either one from the other (Hande et al. 2009). Even so, the first thing to note about the supply and demand dynamics of Internet content is that the supply of additional content may not result in end users increasing their aggregate demand for content or connectivity substantially; supply does not necessarily create its own demand as Say's law says. While we can interact with multiple online services simultaneously, at least with current technology we are able to multitask only to a degree. Even as content on average becomes more and more data-intensive, now it is the number of hours in a day that I would argue is the limiting factor in the demand for content in the aggregate (or rather, content that is consumed "actively" as opposed to background processes). Additional content, provided it has demand in the first place, merely makes users reallocate their time and traffic; if one gets a free option to choose between four video streaming services instead of three, one does not simply watch 33% more films, and even if does, it is at the cost of time available for other types of content. In a way, content can create a nearly zero-sum game.

If Internet content is construed as a distinct set of non-physical and substitutable goods, the market for it is flooded. At the risk of creating an arbitrary dichotomy, even though the Internet permeates our lives in ever more ways, is it still perhaps possible to conceptualize it as in some sense separate from our physical activities? Be that as it may, it may not be meaningless to analyze the supply and demand for Internet content as a whole. The supply-demand diagram for content drawn in Figure 9 provides some straightforward economic intuition into why the average price of online content is close to zero: The supply of content has exploded and the price has dropped dramatically. Revisiting the theoretical framework of Figure 7, the demand for Internet content, *if measured through volumes*, can be empirically meaningless for the demand for connectivity if we are in a saturated zone where a change in the total amount of content is not relevant anymore. In other words, the marginal utility from the volume of content or total number of CSPs is decreasing, and we may have reached a point where the value of an additional unit of content is infinitesimal. Instead of the amount of content *per se*, we should think of content *innovation* as the driver of the demand for content and connectivity.

Most of the network neutrality studies model the demand for connectivity through the demand for content. When this framework is chosen, the mathematical formulation of the demand has crucial effects on the theoretical results.

Figure 9. Evolution of the supply and demand for Internet content.[50] The time allocated to content is used as the numéraire for the quantity of content demanded, as time can be thought of as a more general metric than e.g. the amount of transmitted data. Over the last 15 years or so, the supply of Internet content and services has mushroomed, which is reflected in the supply curve as a massive shift to the right (S → S'). Meanwhile, the aggregate demand for content has become inelastic (D → D') due to the fact that at the end of the day consumers have time constraints that bound the demand. In the end state (x), the average price of content has dropped close to zero (p → p').



At the individual level, end users' content and service preferences tend to be "sticky." For the sake of convenience, we stick to a specific service even when there are numerous, functionally equivalent or superior alternatives available. For the ISP, the demand profiles of different services are of high relevance because it can make different kinds of deals with CSPs and bundle its connections with content. Under non-neutrality its possibilities are wider.

In the economic dimension, Internet traffic can be divided into CSP-to-CSP and CSP-to-end user traffic, the former being insignificant in volume relative to the latter. From an end user's point of view, traffic can be classified into online activities such as web surfing, Skyping, video and audio streaming, and gaming. Ads, a main source of revenue to CSPs, are often considered a nuisance by end users. While Anderson (2003) provides a framework of the broadcasting industry that associates the social marginal benefit of

[50] An equivalent diagram has been uploaded by Albert Wenger to ShowMe. It could be a good thing as he has economics and IT degrees from Harvard and MIT.

advertising with advertisers' willingness to pay for visibility, inefficiency still arises in many scenarios, as the monopoly platform does not fully internalize the nuisance costs.

From an ISP's point of view, traffic can be classified into (1) *sensitive traffic*, (2) *best-effort traffic*, and (3) *undesired traffic*.[51] Sensitive traffic has high QoS requirements with respect to latency, jitter, and loss, and includes video conferencing, VoIP, and gaming. QoS is not deemed as crucial for best-effort traffic, such as P2P or email traffic. Video streaming belongs somewhere in the neighborhood of sensitive and best effort. In 2013, video constituted two-thirds of the volume of Internet traffic; the share is predicted to increase to over 80% by 2018.[4] Undesired traffic includes spam and other malicious or illegal content. Under non-neutrality and exclusivity contracts, it would include traffic from CSPs excluded by the ISP.

3.5 Retail Pricing

Retail Internet access pricing has evolved from time-metered pricing in the days of dial-up Internet into primarily *flat-rate pricing* schemes: The ISP typically charges the end user a flat monthly fee. The fee is typically dependent on the connection speed, not the total volume of data transmitted over a period. Calzada and Martínez-Santos (2014) empirically verify for the EU area the fact that price correlates with speed. An alternative pricing method is *usage-based pricing*. Pure usage-based pricing is nowadays rare. A pricing scheme that was becoming more uncommon but has been reintroduced by some ISPs is to charge the customer for a bundled amount of data and set an overage charge or restrict the bandwidth for usage exceeding the monthly data cap. This is a two- or three-part tariff technique with a fixed part and usage-based parts. The idea of nonlinear pricing is to capture consumer surplus more efficiently by increasing the quantity demanded with a lower unit price and then extracting the leftover surplus with a fixed fee, partially used to cover sunk infrastructure investment costs. Nonlinear pricing is theoretically efficient under preference heterogeneity, and

[51] wikipedia.org. Accessed 19 Nov 2014. Note that in contractual terms, the vast majority of traffic is delivered on a best-effort basis.

has a special role in the management of capacity-constrained resources such as broadband (Beckert 2005). Nonlinear pricing is more widely used in the case of mobile broadband subscriptions, among which metered subscription models are still sometimes seen.

Even though resale of subscriptions is not practically possible in the broadband market and hence price discrimination cannot be arbitraged away by consumers, first-degree price discrimination is not particularly prominent among ISPs. Likewise, third-degree price discrimination, i.e. charging different prices for the same services based on the sales channel, customer location, age, or something else to that effect, is not widely practiced. On the other hand, second-degree price discrimination is almost universally practiced in the form of volume discounts, as again empirically verified by Calzada and Martínez-Santos (2014). For example, 10 Mbps might cost €20 per month, whereas 100 Mbps costs €50 per month. Under non-neutrality, service tiering based on data discrimination can be thought to correspond to second-degree price discrimination.

There is room for some ingenuity in bandwidth-based pricing, as well. At least in Finland it is typical for an ISP to offer a discounted price for the first month to year as a part of, say, a 24-month fixed-term contract. In the US, the monthly subscription fee reportedly drifts upward a dollar or so per month for some of a certain large ISP's customers. If true, this can in fact be construed as first-degree price discrimination where the ISP "tests" the user's valuation.

Bundling and *tying* of broadband and content services have gained much popularity during the last years. The economic intuition of bundling is that offering two or more products in a single bundle can allow for more effective extraction of consumer surplus due to consumers' differing willingness to pay for the products. Further, bundling can help in customer lock-in. The benefits from bundling are one of the reasons for why ISPs have expanded vertically into the online content, cable television, and mobile telephone markets. Bundled services that large ISPs offer with broadband connections range from streaming services and cable television channel bundles to mobile phone subscriptions and almost any value-added services, products, and content in between.

Some large ISPs are cable operators, including Comcast and Time Warner Cable. It has been speculated that the new rise of data caps is in part be due to vertically integrated ISPs shielding their cable services against competitors' streaming services (Lyons 2013). Choi (2010) finds that tying in a two-sided market can be welfare-enhancing if consumers can multi-home. With broadband, this is not usually the case, however.

4. Empirical Effects of Network Non-Neutrality

In this section, I evaluate the meaning of hypothetical deregulation for ISPs and the other market participants at a practical level, and consider the directions the broadband market might take. My approach is necessarily speculative to some degree, as non-neutrality at least in more extreme forms is still an imaginary scenario. By "deregulation" and "non-neutrality," I mainly refer to the NMT, NNN, and DDZ regimes in Table 1, under which either termination pricing or data discrimination or both are deregulated.

The imposition of a more neutral regime (NR, NN, NMR, NMZ) would effectively prevent the ISP from taking two kinds of actions: (1) the *extraction* of rent from CSPs and (2) the *exclusion* of CSPs from end users' Internet content streams (Hemphill 2008). Under non-neutrality, the one-way access scenario is in a way expanded. The framework of Figure 5 can now be reinterpreted to include CSPs. The position of a CSP is analogous to small ISPs having to pay Tier 1 and 2 ISPs for access to networks connecting them to end users. However, in this case the local ISP is lifted upward in the hierarchy. In effect, neutrality regulation can in part be construed as a limitation on the vertical restraints ISPs can place on CSPs (Rosston and Topper 2010).

Before continuing, it should be reminded it is conceivable that non-neutrality would not be optimal for the ISP in all scenarios. An ISP could capitalize on other ISP's non-neutrality and find a market niche in offering neutral connectivity – prevalence of non-neutrality could create market pressure for neutrality. More generally, non-neutrality might foster "network diversity" and hence increase the number of dimensions in which networks compete (Yoo 2006a). That being said, deregulation definitely has upside potential for ISPs. What would non-neutrality mean in practice, then? While speculative, it is sensible to anticipate certain general directions the broadband market might take in the event it happened to be deregulated.

4.1 Prioritization and Pricing

The deregulated ISP is likely to be incentivized to discriminate packets based on their source, destination, and content (e.g. Economides and Tåg 2013; Lee and Kim 2014; van Schewick 2007; Wu 2003). Data discrimination can be hypothesized to benefit the ISP through two distinct mechanisms:

i. Direct revenue from paid prioritization and exclusivity contracts.
ii. Extraction of consumer surplus through differential pricing.

The practical implementation of prioritization at any layer is not a real problem for ISPs. A variety of technical methods including scheduling algorithms such as weighted fair queuing can be used to prioritize packets. Privacy issues arise (but are not discussed further here), as the ISP might use *deep-packet inspection* (DPI) to access the payload of a packet to determine its contents, in contrast to non-intrusively reading the DiffServ field.

As a side note, again decomposing the concept of QoS into bandwidth and latency, the distinction between these two parameters turns out to have business relevance. It is possible for the non-neutral ISP to implement discrimination through either parameter. The ISP can, for instance, offer a connection with high speed (e.g. 100 Mbps downstream) but high latency for low-priority content (e.g. a 10-second waiting time before transmission), or a connection with minimal latency (no waiting time) but low speed (e.g. 1 Mbps) for low-priority content.

Termination Pricing

Under the NRT, NMT, and NNN regimes, i.e. without the zero-price rule, the ISP might then consider charging CSPs termination fees for access to its customer base, regardless of whether it prioritizes data or strikes exclusivity deals. Termination fees can be thought of as analogous to roaming charges by mobile network operators (Jullien and Sand-Zantman 2014). In Economides and Tåg's (2012) model, the profit-maximizing ISP will charge CSPs a positive fee if they value access to additional end users more than end users value access to additional CSPs. By charging termination fees, an ISP with a high degree of market power can extract CSP surplus, as CSPs are forced to pay the ISP for access to a significant number of their customers. Under uncapped termination fees, the higher the degree of the ISP's market power over a CSP, the higher a fee it is able to charge. This bargaining-power ratio is explicitly included in e.g. Choi and Kim's (2010), Altman et al.'s (2011), and Hanawal and Altman's (2013) models, and is in the latter ones seen to influence the parties' preferences between the neutral and non-neutral regimes, with the intuitive outcome that a higher ISP-to-CSP bargaining power ratio makes the ISP prefer the non-neutral regime and the CSP prefer the neutral regime.

Even if charging positive termination fees is allowed, termination pricing need not be completely deregulated. As listed in Table 1, termination pricing can be discriminatory or non-discriminatory (a uniform termination fee). At first glance, the ISP would appear to be better off with discriminatory termination pricing – which, if unregulated, allows differential pricing tactics also on the supply side – and to be likely to prefer discriminatory fees. The CSP would appear to be the worse off the higher the termination fee. However, the effect of termination fees on ISP and CSP surpluses is not necessarily this straightforward. Also, end users may or may not be the better off the lower the fee. I come back to the welfare implications of termination pricing in Section 5.2.

Research on interconnection pricing in telecommunications can potentially be applied in evaluating termination fees if one assimilates CSPs with downstream ISPs as in the expanded one-way access problem. When applying research about telephony to broadband, one has to remember that the two do not share all their characteristics, although they are highly analogous. In mobile telephony, most studies on "off-net" termination rates (for calls terminating in competitors' networks) point toward a lack of threat of excessive rates under deregulation, though Tangerås (2014) finds that the threat exists if there are income effects in end-user demand.

Paid Prioritization and Exclusivity Contracts

Under the DDR, DDZ, and NNN regimes, prioritization can be based on not only the traffic class but also the specific CSP. Under the last non-neutral regime, while not necessarily dependent on payment flows between CSPs and the ISP, prioritization would most likely be closely linked to contracts in the forms of paid prioritization and exclusivity contracts. In paid prioritization, the ISP prioritizes the delivery of affiliated content over other content, and the affiliated CSP pays the ISP for the expedited delivery of its content. For the ISP, the direct benefit is the monetary flow from the CSP; for the CSP, the benefit is better access to and QoS for end users, viz. potential customers. If the basic connection offered by an ISP has very low QoS, the CSP resorting to the basic subscription is in danger of losing customers who do not have the patience to wait when connecting to the CSP with such slow speed, especially if competitors have fast-lane access. (Recall Krishnan and Sitaraman's [2012] estimate of a 6% abandonment rate per one-second delay increment for video streams.)

The other revenue channel from prioritization is more effective extraction of surplus from end users – differing in their valuation of access – in the form of differential pricing. The ISP needs to balance the two revenue streams from prioritization, i.e. direct payments from CSPs versus the extra surpluses extracted from end users through differential pricing, as these are not always compatible. Creating scarcity of access on the supply side to extract revenue from CSPs restricts the possibilities with the service offering and differentiation on the demand side. Moreover, cost minimization through network optimization plays a role. Paid prioritization might range from "soft" QoS differentiation across CSPs to strict exclusivity contracts where the ISP only lets content from an exclusive partner through. The latter would only be possible under the NNN regime. The ISP would be able to auction off access, i.e. exclusive contracts or the best QoS, to extract maximum revenue from CSPs. The specific form of the auction would depend on the situation. Access auctions would naturally not be possible under a uniform or fair termination fee policy. Under a neutral slow lane policy, the ISP would be required to offer end users a basic subscription with neutral, non-discriminated access to the Internet. Complementarily, a zero-price slow lane with acceptable QoS for CSPs can be required. If the definition of "acceptable" is low enough, it may turn out to be necessary for a bandwidth-hungry CSP to pay the last-mile ISP for sufficient QoS in the first place.

Concerns about prioritization have been directed at vertically integrated ISPs in particular, and it is true that vertically integrated ISPs are especially incentivized to prioritize content (Economides and Hermalin 2012; Wang and Sun 2012; Waterman and Choi 2012). This follows from the fact that in-house content can yield more expected revenue than outside affiliated content due to customer lock-in, data, and synergy reasons.

Tiering and Versioning

Data discrimination in connection to differential pricing can be called *tiering* or *versioning*. More specifically, tiering can be considered a form of either product differentiation or second-degree price discrimination, depending on whether the tiered services the ISP offers under non-neutrality differ from each other more than superficially in cost. We can distinguish between two types of tiering the non-neutral ISP can practice: (1) *QoS tiering* and (2) *access tiering*. QoS tiering can be considered more closely product differentiation than access tiering can, as the cost of QoS is likely to be more significant than the cost of allowing access to CSPs. QoS tiering in the non-neutral world should not be confused with ordinary user tiering. Now the ISP can tier its broadband services according to QoS for different traffic classes and individual CSPs as opposed to equally for all of a given end user's traffic. The price of the connection can depend on both the average speed across all CSPs and the priorities or bandwidths given to different CSPs as per the terms of the end user's contract. QoS tiering need not imply truly dedicated bandwidths; connectivity is still supplied on a best-effort basis, provided that certain average QoS with specific CSPs is more or less delivered as advertised.

In addition to QoS tiering, the non-neutral ISP might consider access tiering where customers pay separate fees for access to different CSPs or bundles of CSPs. The ISP might, for instance, offer as the entry-level service a connection that gives access only to basic content. Bundles that provide wider access to the Internet can be made available as more steeply priced options. The first complaint[52] filed following the FCC Open Internet Order nicely gives the idea of tiering as it has already actually been practiced. Mobile operator MetroPCS, now part of T-Mobile, tiered its services as follows: (A) $40 base plan with unlimited GSM talk, text, Web browsing, and YouTube, (B) $50 plan which adds access to services such as Netflix and Skype, but with these capped at 1 GB per month, and (C) $60 plan without the data cap. In sum, tiering can be based on (i) preferential access and exclusivity contracts with CSPs or on differential pricing of (ii) access to CSPs and (iii) QoS to CSPs. As such, tiering can take the form of offering fast and slow lanes and wide and narrow access to the Internet.

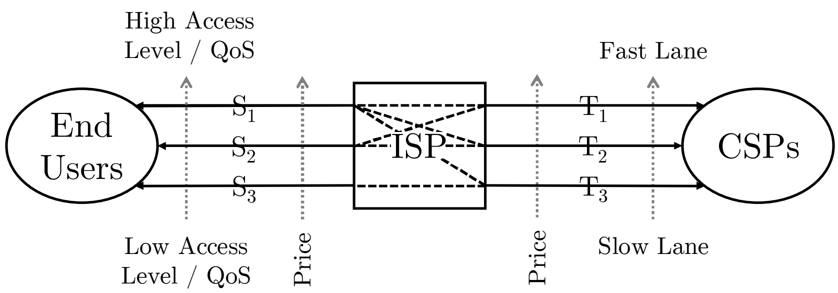
Tiering on the demand side implies similar tiering on the supply side. The ISP devises its tiering and versioning strategy to optimize not only the direct revenue from paid prioritization and exclusivity contracts but also extraction of end-user surplus through differential pricing. Pricing the tiered services on both sides then becomes quite difficult. Optimizing its product mix and prices on either side, the ISP maximizes profit subject to end users' and CSPs' supply and demand profiles, its interconnection costs, and other factors. The general profit-maximizing pricing model could perhaps be something along the lines of a two-sided two- or three-part tariff such that end users pay monthly subscription fee $S + f(a, q, d)$, where S is a common base fee and f is a pricing function. a is the access tier, $q = (q_1, \dots, q_i, \dots, q_k)$ is QoS with traffic classes or specific CSPs, and $d = (d_1, \dots, d_i, \dots, d_k)$ is the volume of data transmitted by the end user within traffic classes or through specific CSPs; $\frac{\partial f}{\partial a} > 0$, $\frac{\partial f}{\partial q_i} > 0\ \forall i$, and $\frac{\partial f}{\partial d_i} \geq 0\ \forall i$. $\frac{\partial f}{\partial d_i} = 0\ \forall i$ translates to no usage-based price components. Analogously, a CSP pays termination fee $T + g(p, c)$, where T is a common base fee and $g(c)$ is the pricing

[52] MetroPCS letter to FCC. www.freepress.net. Retrieved 10 Dec 2014.

function reflecting the priority of the CSP's traffic lane (p), the ISP's capacity provisioning costs (c), and the parties' negotiation power balance. The ISP should set the differential prices to satisfy the incentive compatibility and individual rationality constraints: End users (CSPs) with high valuation should get more utility from a high-tier service net of the subscription fee (termination fee) than from a low-tier service, and hence prefer the high-tier service. Naturally, they should also stand to gain from transacting in the first place.

Tiering can be combined with data caps, where the end user pays for each additional megabyte or gigabyte after a predefined threshold. Provided the ISP charges over marginal cost, it is probably not in its interest to do much to prevent the user from transferring the extra bytes. But the ISP might consider offering, say, YouTube a deal where YouTube clips do not count toward the user's data cap. As noted by Claffy and Clark (2014), so far at least Facebook has paid certain mobile carriers for providing their customers free access to Facebook. On the whole, it can readily be seen that the possibilities for the ISP in tiering and versioning are practically limitless. Figure 10 sketches a simplified pricing framework under non-neutrality.

Figure 10. Tiering under non-neutrality. The ISP can differentiate the connections based on access level and QoS, and set subscription fees $S_1 > S_2 > S_3$ and termination fees $T_1 > T_2 > T_3$ accordingly to maximize profit.



4.2 Fragmentation

Internet *fragmentation* is a hypothesized large-scale consequence of network non-neutrality. Network neutrality has not been studied extensively from the fragmentation point of view, with D'Annunzio and Russo (2013), Hill (2012), Kourandi et al. (2014), and Lee and Wu (2009) as some noteworthy exceptions. Fragmentation refers to isolation of Internet services and users from each other due to access and QoS tiering by ISPs. We can distinguish between two levels of fragmentation: *Content* fragmentation and *physical* fragmentation (cf. Internet layers). The neutrality debate is mainly concerned with content fragmentation, and non-neutrality is unlikely to pose a direct threat of physical fragmentation. Fragmentation at both layers can also be caused by external factors, such as governmental censorship.

Remember that the non-neutral ISPs under the DDR, DDZ, and NNN regimes may be able to auction off exclusive access for a CSP to their end users. This is possible if the expected revenue from access to the ISP's customers, over which the ISP may have a monopoly, is high enough for CSPs. More often than not, the ISP has bargaining power over the CSP. However, this is not always true, and an opposite situation may actually emerge if the value of a CSP is high enough in the eyes of ISPs. Then, the ISPs may end up competing for an exclusive arrangement with the CSP (Lee and Wu 2009; Lotfi et al. 2014). Lotfi et al. (2014) model this inverted setup game-theoretically, and in their outcome – while the CSP is able to control the ISP to a degree – the ISP still retains its power and can actually extract some of the CSP's surplus. Funnily enough, the endgame is the same at both market extremes: Fragmentation of access to content.

When ISPs have bargaining power over CSPs, deregulating data discrimination and exclusivity contracts is prone to increase the risk of content fragmentation because ultimately there is no guarantee that the ISPs providing broadband services with differentiated access to content coordinate between each other so that each piece of content is reached by each end user. Mandating a neutral slow lane to have ISPs offer at least one connection with full, neutral access to CSPs would alleviate the problem. A minimum-QoS standard would similarly contribute to the prevention of fragmentation. Finally, a zero-price and full-access slow lane for CSPs would enable even the smallest CSPs to access to all end users. Even so, fragmentation can in practice still occur if either the neutral connection for end users or the zero-price slow lane for CSPs has untenably low QoS, and if better connections are not accessible (in terms of price) to all end users – then not all end users are truly able to interact with everyone else. Retail price regulation (NR, NMR, DDR) would, in turn, mitigate this problem. Also, the minimum-QoS standard should accordingly be set high enough both in absolute terms and relative to the average and fastest connections available.

A fragmented Internet stands in contrast to the idea of the Open Internet, and is harmful to both consumer surplus and total welfare (D'Annunzio and Russo 2013; Kourandi et al. 2014). The dystopia of fragmentation is that the largest ISPs actually collude with each other and the largest CSPs to exclude or discriminate against the rest of the CSPs' traffic. Content markets would concentrate, and end user would have a limited number of CSPs to choose from, and perhaps at low QoS depending on the user's purchasing power. Due to market pressure from end users, however, this is somewhat unlikely, and were it to happen, competition authorities would likely react (Hill 2012). A related but not directly non-neutrality-related dystopia is that the exclusion might extend into the inter-ISP market so that the largest ISPs end up forming exclusive peering relationships between the other, excluding smaller ISPs. As with data discrimination, vertical integration may cause concerns in regard to fragmentation: If vertically integrated backbone operators (e.g. AT&T or Verizon) decided to peer exclusively with each other, other backbone providers might imitate this strategy, and the end state would be both an oligopolistic and a vertically integrated Internet backbone that divides the Internet into detached archipelagos (Hill 2012).

The zero-price rule in itself may be neither sufficient nor necessary to prevent fragmentation (D'Annunzio and Russo 2013; Lee and Wu 2009; Kourandi et al. 2014). As the rule only concerns supply-side pricing, ISPs adhering to the rule will still be able to tier their services, although purely based on differential pricing tactics rather than direct payment flow from CSPs. More generally, the degree of neutrality and the level of fragmentation are not necessarily directly proportional to each other, and the relation may not be monotonic. Introducing termination fees does not necessarily lead to fragmentation, either (D'Annunzio and Russo 2013).

4.3 Innovation and Investment

Innovation at the Core

Investment in network infrastructure will not break even below a population density threshold (Götz 2013). Since it is not profitable to extend the infrastructure into all rural areas, governments have initiated universal-access agendas to subsidize network expansion. Investment can be made toward broadband coverage (geographical areas covered), penetration (percentage of people covered), and QoS. While the duplication of infrastructure in a regulated network industry can have a positive consumer-surplus effect (Krämer and Vogelsang 2014), for firms at the investment stage it often makes no sense. However, the natural-monopoly characteristics of the networks can make co-investment and joint infrastructure projects economically rational for ISPs. Then again, infrastructure cooperation facilitates tacit collusion in the retail market, and the consumer-surplus effect of co-investments is ambiguous (Krämer and Vogelsang 2014).

In contrast to the regulator's aspirations, coverage, penetration, and QoS are not of interest to the ISP as such, but only if investment in them pays back. An interesting question arises about the dynamic consequences of network neutrality: How does neutrality regulation affect ISPs' incentives to invest in infrastructure? It has been widely established that competition in both the retail broadband market and interconnection market increases broadband investment, although there is a tradeoff between coverage and penetration (see Götz [2013] for a review). Neutrality can affect ISP investment through either direct changes in ISP profit or changes in the competitive landscape. Intuitively, non-neutrality is more likely than not to increase profits. Some ISPs have maintained that deregulating neutrality and allowing more differentiated services would enable them to use any extra profits to invest in new infrastructure, benefiting the society in the end. If this happens, deregulation can be dynamically efficient. An opposing argument is that neutrality increases the level of competition between ISPs and therefore stimulates investment in infrastructure. From studies about access-price regulation, we know it is prone to reduce the ISP's incentives to invest in quality-enhancing technology (Cambini and Jiang 2009; Kotakorpi 2006).

More generally, Cambini and Jiang's review on the literature on telecommunications investment incentives under price-cap regulation concludes that stringent price caps can reduce investment in infrastructure. However, a price cap set relatively high may not have such an effect.

Some of the mathematical models to be summarized in Section 5.2 provide predictions of how ISPs' investment incentives differ between the neutral and the non-neutral regimes. Taken together, the models support the view that the price or network regulation of the more neutral regimes has the unwanted side effect of inhibiting investment and degrading dynamic efficiency. Although over-investment is not socially optimal, either, most of the models predict that under-investment is the more probable outcome under non-neutrality. Under neutrality, the investment level would in more cases be closer to the social optimum. Under the NMR, NMZ, and NMT regimes, where the line is drawn at throttling, a possibility that cannot be discounted is that the ISP decides to operate at the capacity limit to justify prioritization, which would de-incentivize capacity investment. The ISP could also create artificial scarcity for similar purposes.

In comparison to the regulated case, non-neutral ISPs have more freedom to innovate "at the core" of the network. Innovation might mean network optimization through new technologies, which need not be used for "malevolent" practices from the ISP side, but could help bring a better online experience to end users. Moreover, excess regulation can be argued to threaten the natural and continual evolution of the Internet and hence unnecessarily limit the possibilities it brings about. As Yoo (2006b) argues, neutrality might hinder innovation at the core: "Allowing network owners to employ different protocols can foster innovation by allowing a wider range of network products to exist. Conversely, compulsory standardization can reduce consumer surplus by limiting the variety of products available. In the words of two leading commentators on network economics, 'market equilibrium with multiple incompatible products reflects the social value of variety.'"

Innovation at the Edge

Ultimately, content on the Internet is the source of value to end users. (In marketing, applications fundamental to the value of the platform are called "killer applications.") The effects of neutrality regulation on CSP innovation – innovation at the edge of the network – are crucial to the overall welfare implications. Unfortunately, only a fraction of the theoretical models give predictions of the content-innovation dynamics under different regulatory regimes. According to those mathematical models that address this question, the prediction of the effect of neutrality regulation on CSP innovation is more ambiguous than the near-consensus prediction of a negative effect on ISP investment.

It is possible that termination pricing hinders competition among CSPs and creates barriers to entry, the idea being that the more powerful, incumbent CSPs may be able to buy their way to end users over a market entrant, even when the entrant would have the more innovative and data-efficient service. Start-ups and small CSPs that rely on online presence and have not started to generate revenue might become heavily burdened by termination fees. Small CSPs are then prone to suffer more from non-neutrality (Bourreau et al. 2013). This would hinder entrepreneurship and have a negative impact on competition in online markets (Sydell 2006). An opposing point of view is that tiering can encourage CSPs to employ higher QoS in tailoring services to end users' demands (Kulick and Weisman 2010). The tremendous diversity of CSPs means that non-neutrality affects them in divergent ways, making ubiquitous conclusions difficult to draw. CSPs more reliant on access to customers might suffer more. Industry-wise, especially affected by termination fees, tiering, and fragmentation may be "infopreneurs" whose whole business model of gathering and selling information on the Internet is based on wide access. Another group of CSPs that might suffer relatively much are those with a long-tail revenue model, catering to niche markets through a dispersed online presence (Lee and Wu 2009).

The rationality of "extorting" CSPs depends on the relative power of the ISP and the effects of surplus extraction on CSP participation. As we have seen, the absolute amount of content on the Internet is huge. But since the demand for new, diverse, and innovative content is the main driver of the demand for connectivity, ISPs' competitiveness relies more on content innovation than on the amount of content in itself. Remembering once again that it is dependent on CSPs that add value to the ISP's product, the broadband connection, the non-neutral ISP has to balance between short-term gains and long-term effects on CSP participation and innovation.

5. Theoretical Effects of Network Neutrality

In this section, I examine the theoretical predictions of the welfare effects of network neutrality. First, I review five notable mathematical models that take differing angles to attack the puzzle. For the most part, I stick to my objective of presenting the basic layouts of the models, excluding any extensions that may be made in the latter parts of the papers. In the succeeding subsection, I tabulate the welfare predictions of all the prominent models. Lastly, I discuss the policy implications.

5.1 Models of the Market

Network neutrality only began to emerge as a topic in economics 10 years ago. The vast majority of the formal models are from the last five years. Capturing all the aspects of network neutrality in a model is highly difficult if not downright impossible. The biggest weakness of most models is the simplistic formalization of the regulatory frameworks that loses the subtleties of the issue. An example is the lack of distinguishing between discriminatory and non-discriminatory termination pricing. This poses a danger of omitting dynamics arising in a two-sided market, even when the model is mathematically advanced. Another important factor affecting the real-world market outcome is the diversity of CSPs. This is also reflected in the models, in which the asymmetry of CSPs is by far the most commonly encountered parameter type affecting the welfare results (e.g. Altman et al. 2014; Cheng et al. 2011; Choi and Kim 2010; Economides and Tåg 2013; Guo et al. 2013; Kourandi et al. 2014; Krämer and Wiewiorra 2012; Mialon and Banerjee 2014; Musacchio et al. 2009; Njoroge et al. 2013;

Reggiani and Valletti 2012). In many models, the welfare effect also depends on the location on the price axis, that is, the termination or subscription fees.

The majority of the models in the literature specify a monopolistic ISP or a duopoly, two or a continuum of CSPs, and a continuum of end users (cf. the Hotelling model). On the other hand, the mathematical approaches to modeling the demand for connectivity or utility from content differ significantly across authors. Most models give qualitative rather than quantitative predictions of economic efficiency under neutrality and non-neutrality, with certain exceptions such as Guo et al. (2012), Lee and Kim (2014), and Nurski (2014). I believe that the limitations in each model make a “meta-analysis” of the models all the more important for policy analysis, however sophisticated each model may be in their own right.

Economides and Tåg (2012). Network Neutrality on the Internet: A Two-Sided Market Analysis

Economides and Tåg’s model is a natural starting point, attractive in its relative mathematical elegance and tractability. The authors first treat the ISP as a monopolistic platform operator between end users and CSPs. They specify a Hotelling model with a continuum of both end users and CSPs with differing valuation parameters. The two groups exert positive externalities on each other. Utility to end user i from connecting to the Internet is

$$u_i = v + bn_{csp}^e - tx_i - p, \tag{4}$$

and the marginal user indifferent between connecting and not connecting to the Internet using the given connection is located at

$$x_i = \frac{v + bn_{csp}^e - p}{t}. \tag{5}$$

The marginal CSP indifferent about entering the market is located at

$$y_j = \frac{an_u^e - s}{f}. \tag{6}$$

Here, p is the subscription fee and s the termination fee. v is the base intrinsic value to users from connecting to the Internet, b is the marginal value of an additional unit of content to the user, and t is their unit transportation cost. f is a parameter denoting the common part of CSPs' fixed costs, and a is the marginal value of an additional user to CSPs. n_{csp}^e and n_u^e denote the expected number of CSPs and users, respectively. Assuming fulfilled expectations, we have $n_u = x_i$ and $n_{csp} = y_i$, n_u (n_{csp}) denoting the number of users (CSPs) participating in the market. Solving the system of simultaneous equations (5) and (6), Economides and Tåg proceed to write the demand for connectivity as a function of the subscription and termination fees:

$$n_u(p,s) = \frac{f(v-p) - bs}{ft - ab}. \tag{7}$$

Analogously, the supply of CSPs is given by

$$n_{csp}(p,s) = \frac{a(v-p) - ts}{ft - ab}. \tag{8}$$

The monopolist chooses p and s simultaneously to maximize profit. The monopolist's problem is

$$\max_{p,s} \Pi_{isp}(p,s) = (p-c)n_u(p,s) + sn_{csp}(p,s), \tag{9}$$

which yields

$$p = \frac{(2ft - ab)(v + c) - b^2c - a^2v}{4ft - (a+b)^2} \tag{10}$$

and

$$s = \frac{f(a-b)(v-c)}{4ft - (a+b)^2}. \tag{11}$$

End-users' surplus is calculated as

$$CS(p,s) = \int_0^{n_u(p,s)} [v + bn_u(p,s) - tx - p]\mathrm{d}x, \tag{12}$$

and CSPs' total profits are

$$\Pi_{csp} = \int_0^{n_{csp}(p,s)} [an_u(p,s) - fy - s]\mathrm{d}y\,. \tag{13}$$

Due to the positive network effects, the social planner would actually want to set $p < c$ and $s < 0$, i.e. a subscription fee below marginal cost, and a negative termination fee. An intuitive insight from the analysis is that under non-neutrality, the ISP, which aims to maximize its profits, will charge CSPs a positive fee if they value access to additional users more than users value access to additional CSPs. Therefore, the relative size of the cross-group externalities, a/b, is of interest, as it influences the prices the ISP charges. Neutrality in the form of the zero-price rule is reflected in the model as the simple constraint $s = 0$.

Economides and Tåg evaluate the static welfare effects of the zero-price rule, namely, user, ISP, CSP, and total surplus. They also extend the analysis to a duopolistic setting between two ISPs, which yields outcomes similar to the monopoly case. Overall, their findings would indicate that in most cases neutrality regulation increases total surplus through the increase in the total CSP profits. For most parameter values, however, end users are actually better off under non-neutrality. As the authors put it, "The intuition is that in [non-neutral] monopoly, consumers benefit from a lower subscription price since the monopolist has incentives to attract more consumers to generate extra revenue from charging content providers." This model does not consider congestion or data discrimination.

Hermalin and Katz (2007). The Economics of Product-Line Restrictions with an Application to the Network Neutrality Debate

From 2007, Hermalin and Katz's study is the very first mathematical model of network neutrality, to my knowledge. The authors treat neutrality as a form of product-line restriction. Specifically, under neutrality, the ISP is not allowed to tier its broadband service based on different QoS to different CSPs. In the model, there is an ISP monopoly and continuums of end users and CSPs with unit masses. CSPs come in two types, $\theta \in [0, \bar{\theta}] \equiv \Theta$, with end users valuing the latter type more than the former, distributed according the cumulative distribution function $F(\cdot)$ and probability density function $f(\cdot)$. The ISP, who does not observe the type of a given CSP but only knows the general distribution, charges end users subscription ("hookup") fee h and CSPs interconnection fee $p(q)$; q denotes QoS, and $q = 0$ is used to indicate that the CSP is not connected to the platform. When connected, $\underline{q} > 0$ is the minimum QoS the ISP provides.

Hermalin and Katz specify two alternative business models for CSPs: A membership-fee model, where CSPs charge end users membership fee t, and an ad-supported model, where advertisers pay a per unit of end-user demand to CSPs, which offer their service to end users free of charge. A fee-charging CSP sets

$$t = \arg\max_{t} \theta q(t-k)d(t)\,, \tag{14}$$

and a CSP's maximal profit is

$$\Pi_{csp} = \theta q \rho\,, \tag{15}$$

where $\rho = (t^* - k)d(t^*)$ with the membership-fee model, and $\rho = (a-k)d(0)$ with the ad-supported model.

End-user utility is quasi-linear, given by

$$U = \int_{\Theta}\int_{0}^{x(\theta)} u\left(\frac{z}{\theta q(\theta)}\right) f(\theta)\mathrm{d}z\mathrm{d}\theta + y\,, \tag{16}$$

where $x(\theta)$ is end-user demand for each type-θ CSP and y is the amount of the composite good consumed. End-user surplus from consuming all CSPs' content is given by

$$CS = \sigma \int_{\Theta} \theta q(\theta) f(\theta)\mathrm{d}\theta\,, \tag{17}$$

where $\sigma = \int_{t^*}^{\infty} d(z)\mathrm{d}z$ with membership-fee CSPs, and $\sigma = \int_{0}^{\infty} d(z)\mathrm{d}z$ with ad-supported CSPs.

The timing in the model goes so that the ISP first chooses h and $p(q)$, after which CSPs choose q and t. Finally, end users observe h, q, and t, and decide whether to connect and how much to consume content from each CSP. Noting that to satisfy the profit-maximization, incentive-compatibility, and individual-rationality criteria, the unrestricted, non-neutral ISP sets $p[q(\theta)] = \rho\theta q(\theta) - \int_0^{\theta} \rho q(\tau)\mathrm{d}\tau$, Hermalin and Katz write down the ISP's problem as

$$\max_{q(\theta)} \int_0^{\overline{\theta}} \{(\rho + \sigma)\theta q(\theta) - \int_0^{\theta} \rho q(\tau)\mathrm{d}\tau - c[q(\theta)]\} f(\theta)\mathrm{d}\theta \\ \mathrm{s.t.}\, \underline{q} \leq q(\theta') \leq q(\theta) \tag{18}$$

using also the fact that end users are homogenous so that the ISP captures their entire surplus with h.

Neutrality regulation means restricting the ISP to offering only a single quality level, q_r. Now, the authors remark, "[T]here is a marginal [CSP] type just indifferent between connecting and not. Rather than view the [ISP's] problem as one of choosing an optimal quality and price, we can view it as one of choosing an optimal cutoff type and quality." The restricted, neutral ISP's problem then becomes

$$\max_{q,\theta} \int_{\theta}^{\overline{\theta}} [\rho\theta q + \sigma\tau q - c(q)] f(\tau)\mathrm{d}\tau\,. \tag{19}$$

Hermalin and Katz identify three welfare-effect channels for a product-line restriction. First, the *exclusion effect* reduces total welfare by reducing the number of active CSPs, *ceteris paribus*. Second, the *reduced-quality* effect reduces total welfare by reducing the highest available QoS. Third, the *improved-quality effect* increases total welfare by increasing QoS for some CSP types. The net welfare effect of neutrality regulation is ambiguous. However, the authors show that for neutrality regulation to improve welfare, the marginal CSP type should enjoy much higher QoS than what it would get under non-neutrality. The conclusion is that, for conservative parameter values, neutrality regulation as a product-line restriction tends to reduce total welfare.

Musacchio et al. (2009). A Two-Sided Market Analysis of Provider Investment Incentives with an Application to the Net-Neutrality Issue

Musacchio et al.'s model differs from the rest of the literature in that it models monopolistic competition among many ISPs. The authors include multiple ISPs in their model to describe a potential free-riding scenario among non-neutral ISPs where "an ISP can increase [its] price to [CSPs] and enjoy the additional revenue this increase causes, while the downside of inducing the [CSP] to invest less has to be borne by all of the ISPs." Musacchio et al. treat end users, ISPs, and CSPs as countable rather than continuously distributed on an interval. There are M CSPs and N ISPs. Each ISP T_n is connected to a set of end users, U_n, has monopoly power over them, and charges them subscription fee p_n for each "click," which acts as the unit of demand from the end-user side. CSPs receive ad revenue a per click. Under non-neutrality, each ISP charges each CSP C_m termination fee q_n per click. Infrastructure and innovation investment by the CSP and ISP are denoted by c_m and t_n, respectively. c_m^v (t_n^w) is the value of CSP C_m's (ISP T_n's) investment perceived by the average end user. In the special symmetric case where the total CSP (ISP) investment is split equally between each CSP (ISP), i.e. $c_m = c/M$ ($t_n = t/N$), the total rate of clicks to CSP C_m is

$$D_m = \left(\frac{1}{M^v} c^v t^w\right) e^{-p/\theta}, \tag{20}$$

and the total rate of clicks through ISP T_n is

$$B_n = \left(\frac{1}{N} M^{1-v} c^v t^w\right) e^{-p/\theta}, \tag{21}$$

where θ is a parameter reflecting end users' price elasticity, $e^{-p/\theta}$ being a normalization factor. More generally, $D_m = \sum_n R_{mn} \equiv \frac{c_m^v}{c_1^v + \cdots + c_m^v} B_n$. Both D_m and B_n are increasing and concave in t and c. The difference between them is that the latter is increasing in M but the former does not increase with N. In other words, end users value variety in content, whereas the number of ISPs does not add to their utility.

CSP surplus is

$$\Pi_{C_m} = \sum_{n=1}^{N} (a - q_n) R_{mn} - \beta c_m, \tag{22}$$

where $\beta > 1$ represents the CSP's opportunity cost. ISP profit is

$$\Pi_{T_n} = (p_n + q_n) B_n - \alpha t_n, \tag{23}$$

where $\alpha > 1$ represents the ISP's opportunity cost.

Like Economides and Tåg, Musacchio et al. define neutrality as the zero-price rule. Under neutrality, ISPs first choose (t_n, p_n) simultaneously; under non-neutrality, they can set (t_n, p_n, q_n). In the second stage, CSPs choose c_m. Solving for the symmetric Nash equilibrium, Musacchio et al. arrive at

$$p_n = p_0 \equiv \frac{\theta N(1-v)}{N(1-v)+v} \tag{24}$$

under neutrality. Under non-neutrality, ISPs set

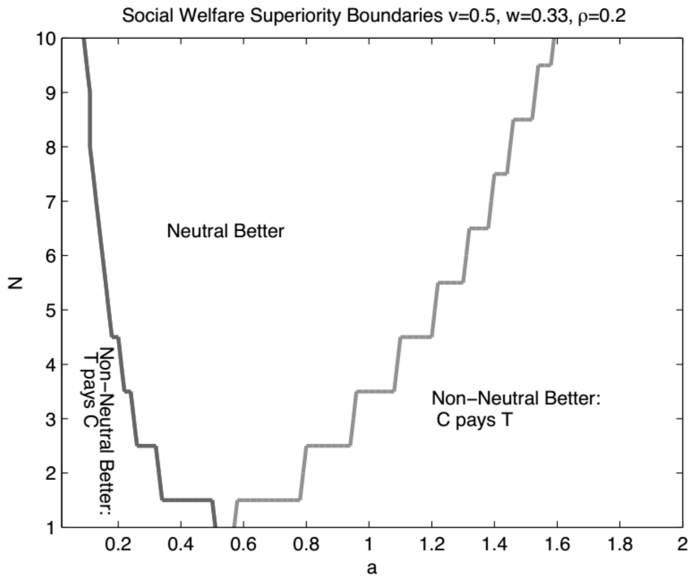
$$p_n = p \equiv \theta - a \tag{25}$$

and

$$q_n = q \equiv a - \theta \frac{v}{N(1-v)+v}. \tag{26}$$

The authors solve analytically for c_m, t_n, click rates and CSP, ISP, and end-user surplus, which they calculate by taking the integral of the total click rate function, the proxy for end-user demand, from the equilibrium price to infinity. They find that the ISP free-riding scenario is realized, and ISPs tend to overcharge CSPs in the non-neutral equilibrium. This reduces CSP investment. Still, as CSPs make their investment decisions only after ISPs have committed to their pricing decisions, they retain some surplus. The surpluses depend on N and the relative values of a and θ, i.e. the ratio of ad revenue per click to the price sensitivity of end users. In Figure 11, Musacchio et al. plot the dependence of the relative welfare levels under the two regimes on N and a. It can be seen that the total welfare effect of neutrality is in general ambiguous. For the intermediate range of a/θ, the neutral regime is welfare-superior. Musacchio et al. leave congestion and data discrimination out of their model.

Figure 11. Total welfare superiority boundaries. In Musacchio et al.'s model, the neutral regime is welfare-superior for intermediate ad-revenue values, whereas the non-neutral regime dominates for small or large ad-revenue values. Source: Musacchio et al. (2009).



Choi and Kim (2010). Net Neutrality and Investment Incentives

Contrary to Economides and Tåg and Musacchio et al., Choi and Kim approach neutrality from the data-discrimination perspective to evaluate the effects of QoS tiering on both static welfare and the agents' investment incentives. Their setting corresponds to the NNN regime but with no throttling allowed. They specify a monopolistic ISP, two CSPs, and a continuum of end users. The ISP charges end users, whose total mass is one, subscription fee a. End users demand content from CSP 1 and CSP 2 with market shares σ_1 and σ_2 at request rate λ. End users are dispersed over a Hotelling line with CSP 1 at one end and CSP 2 at the other. The unit transportation cost is denoted by t as in Economides and Tåg's model. Users' utility function is simply $u(\lambda) = v$. CSP i generates revenue stream r_i and incurs marginal cost c_i per request (or "click" as in Musacchio et al., if you will); i's markup is $m_i \equiv r_i - c_i$, and $m_1 \geq m_2$. This means that the two CSPs can be asymmetric in either their revenues or costs or both. Incorporating congestion in their model, Choi and Kim denote the ISP's capacity as μ. The service time is exponentially distributed with a mean of $1/\mu$. The authors treat μ as exogenous in the short run and endogenous in the long run. The network is modeled as an M/M/1 system, so in the neutral network the waiting time is $w \equiv \frac{1}{\mu-\lambda}$. In the non-neutral network, priority-lane packets wait for $w_1 \equiv \frac{1}{\mu-\lambda_1}$ and slow-lane packets for $w_2 \equiv w_1 \frac{1}{\mu-\lambda_1}$. We then have $w_2 > w > w_1$ when $\mu > \lambda$.

Choi and Kim start with a symmetric scenario where the CSPs split the market equally. The game to be solved for an SPE by backward induction goes so that, in the non-neutral case, the ISP first contracts on the priority lane with one of the CSPs, after which it sets the subscription fee. Then, end users choose their CSPs. The ISP delightingly straightforwardly maximizes $\pi_m = a$, although subject to the condition that the market is wholly covered. The ISP's profit, that is, the subscription fee in the neutral equilibrium is

$$\pi_m^* = a^* = v - \frac{1}{(\mu - \lambda)} - \frac{t}{2}, \tag{27}$$

and CSP i's profit is

$$\pi_i^* = \frac{m_i}{2}\lambda. \tag{28}$$

In the non-neutral case, the ISP maximizes $\pi_m = \tilde{a} + f$, where f the termination fee extracted from CSP 1, assumed to be the one to get the priority lane. Choi and Kim do not place restrictions on the trading or auction mechanism that determines f, but use θ to represent the ISP's bargaining power. It is derived that $f = (m_2 + \theta\Delta_m)(2\tilde{x} - 1)\lambda$, where $\tilde{x}$ denotes the indifferent user's location, and $\Delta_m \equiv m_1 - m_2$. The ISP's maximal profit under non-neutrality is

$$\tilde{\pi}_m^* = v - \frac{1}{(\mu - \tilde{x}\lambda)} - t\tilde{x} + f. \tag{29}$$

Getting the priority lane also leads to a higher market share for CSP 1 so that $\sigma_1 > \frac{1}{2} > \sigma_2$. Since customers' switching from CSP 2 to CSP 1 forms a positive feedback loop, Choi and Kim make the assumption of sufficient differentiation of content so that $\sigma_2 > 0$ and that an interior solution exists. CSPs' profits under non-neutrality are given by

$$\tilde{\pi}_i^* = \begin{cases} m_1\tilde{x}\lambda - f, & i = 1 \\ m_2(1 - \tilde{x})\lambda, & i = 2 \end{cases}. \tag{30}$$

When $m_1 = m_2$, i.e. the CSPs are symmetric, their profits are equal and independent of θ.

The output of Choi and Kim's basic model as far as static efficiency is concerned is that while subscription fee a decreases under non-neutrality, f may compensate for the loss to the ISP. The authors find that $\tilde{\pi}_m^* > \pi_m^*$, i.e. the ISP prefers the non-neutral regime, if m_1 and m_2 are relatively high. If the CSPs' markups are asymmetric, the ISP's preference between the two regimes depends on its bargaining power θ. The low-markup CSP 2 is always worse off under non-neutrality. If CSP 1's markup exceeds that of CSP 2 substantially enough, CSP 1 benefits from the priority lane compared to the neutral case. The higher θ, the larger the markup differential has to be for CSP 1 to prefer the non-neutral regime. End users are better off under non-neutrality. The authors remind that "an individual consumer's surplus increases linearly with the distance of her location from the marginal consumer who is indifferent between the two [CSPs]." Finally, due to there being no demand effects from pricing, the total welfare effect from a change of regime is composed of the CSPs' markups and users' transportation and delay costs. The markups are higher, as are transportation costs, while delay costs are unchanged. The irrelevance of delay costs for static efficiency reflects the fixed capacity constraint – binding under both regimes – meaning that the average service time is unchanged. If the markup differential is large enough relative to transportation costs, non-neutrality is welfare-superior in the static universe.

Moving to a dynamic framework, Choi and Kim evaluate the dynamic effects on ISP investment and CSP innovation. To see the ISP's incentives to invest in capacity expansion under the two regimes, they do away with the exogeneity of μ, with respect to which they now take the derivative of the ISP's profit functions. Evaluating $\frac{d\pi_m}{d\mu}$ and $\frac{d\pi_m}{d\mu}$ (taking into account the ISP's capacity provisioning costs), Choi and Kim identify two channels through which capacity expansion affects the ISP's profits: (i) the *network access fee effect* on the demand side and (ii) the *rent-extraction effect* on the supply side. Under neutrality, capacity expansion increases the connection speed in a uniform fashion, allowing the ISP to charge a higher subscription fee. The network access fee effect is ambiguous in the non-neutral case, where capacity expansion benefits CSP 2 customers with the slow lane relatively more, but on the other hand reduces the marginal CSP 1 customers' transportation costs. The rent-extraction effect refers to the fact that as capacity increases, the benefit from the priority lane for CSP 1 becomes smaller, thus having a negative effect on the ISP's investment incentives under non-neutrality. On the other hand, when CSPs' dynamic behavior is included in the equation, full rent extraction is not optimal for the ISP in the long run anyway. The net investment effect is ambiguous, but the identification of the effect channels in the model would allow testing for the net effect with different parameter values. *A priori*, it would seem that a positive investment effect is more likely than not under neutrality.

Choi and Kim also address CSPs' innovation incentives, representing investment in markup-increasing technology as increasing and convex in the magnitude of the resulting increase in the markup. CSPs invest to the point where the marginal benefit from the markup increase equals marginal cost. The non-neutral case is complicated by the fact that the benefit depends on whether or not the CSP gets the priority lane, which is only known *ex post*. The pure strategy equilibria conditions under non-neutrality are such that while the individual CSPs' strategies can differ across the equilibria, total CSP investment remains constant. The takeaway is that, under non-neutrality, the high-markup CSP 1 will choose to invest less than under neutrality in innovative technology when the ISP's bargaining power θ is relatively high, whereas the low-markup CSP 2 will always invest less than under neutrality. Hence, neutrality regulation would seem beneficial in terms of CSP innovation.

Njoroge et al. (2013). Investment in Two-Sided Markets and the Net Neutrality Debate

Njoroge et al. take a dynamic modeling approach to zoom in on the effects of neutrality regulation on ISP investment and CSP innovation. They use CSP market coverage (or CSP participation) as a proxy for CSP innovation. In their game-theoretic formulation, Njoroge et al. specify an ISP duopoly with continuums of end users and CSPs. User i's utility function in connecting to the Internet has the form

$$U_i(\varphi(i)) = max\{0, R + \theta_i F_i(y_{\varphi(i)}, y_{\varphi(-i)}, \bar{y}, a, r_\alpha, r_\beta) - p_{\varphi(i)}\}. \quad (31)$$

$\varphi: [0, f] \to \{\alpha, \beta\}$, where $f \in [0,1]$ is the mass of a continuum of users, maps users' ISP choices; $\varphi(i)$ is the ISP user i chooses out of the two. In words, a fraction of 0 to f of the users chooses ISP α while the rest choose ISP β. r_α and r_β (q_α and q_β) are then the realized masses of users (CSPs) joining platform α and β, respectively. R is users' common reservation utility, and θ_i is user i's individual preference parameter, drawn from a uniform distribution on $[0, f]$ – again a standard Hotelling formulation. F_i is user i's utility from using ISP $\varphi(i)$, $y_{\varphi(i)}$ is QoS with ISP $\varphi(i)$, $y_{\varphi(-i)}$ is QoS with the other ISP, and r_α (r_β) is the fraction of CSPs which connects to ISP α (ISP β). $p_{\varphi(i)}$ is the subscription fee charged by ISP $\varphi(i)$. $[\bar{\gamma} - a, \bar{\gamma} + a]$ is the support (roughly, the range of possible values) of the uniform distribution of CSP quality, so $a/\bar{\gamma}$ can be interpreted as the coefficient of variation in quality.

CSP j's profit from connecting to platform $\hat{\varphi}(j)$ is

$$v_j = g(\gamma_j, \gamma_{\hat{\varphi}(j)})q_\alpha + g(\gamma_j, \gamma_\beta)q_\beta - w_{\hat{\varphi}(j)}, \tag{32}$$

where the first two terms are the gross revenue, and $w_{\hat{\varphi}(j)}$ is the termination fee charged by ISP $\hat{\varphi}(j)$. $g(\gamma_j, \gamma_{\hat{\varphi}(j)})$ represents ad prices, increasing in both QoS and the CSP's quality.

The profit of ISP $z \in \{\alpha, \beta\}$ is

$$\pi_z = p_z q_z + w_z r_z - I(y_z), \tag{33}$$

where $I(y_z)$ is the investment cost to achieve QoS y_z, increasing and convex in QoS.

In the six-stage game, as compared to Musacchio et al.'s and Choi and Kim's two- and three-stage games, ISPs first simultaneously choose their QoS, after which they set the termination fees. Next, CSPs make their ISP choices. Then, the ISPs simultaneously proceed to set subscription fees to users, who decide which platform to join after observing the subscription fees. Finally, users choose which CSPs to use. Neutrality is represented in the model so that CSP j connected to platform $\hat{\varphi}(j)$ has equal access to both platforms' users and, conversely, user i connected to platform $\varphi(i)$ has access to all CSPs. This is not true in the non-neutral case. There, the CSP connecting to ISP z

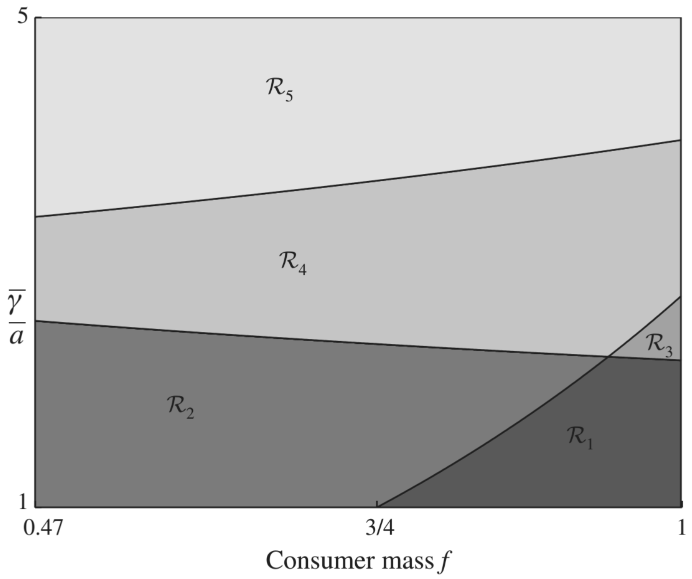
has to pay for an enhanced quality level $\gamma_z > \gamma_{basic}$ (priority lanes) or, in the "walled-garden" scenario, access in the first place: $\gamma_{basic} = 0$. In the neutral walled garden, ISPs offer only one type of connectivity at QoS γ_z. The non-neutral walled garden is equivalent to a uniform termination pricing policy (weak content-provider neutrality) under the NNN regime, whereas the priority-lanes regime is equivalent to a discriminatory termination pricing policy.

Similarly to Choi and Kim's model, Njoroge et al. solve the game for the SPE by backward induction. In the priority-lanes case, they do this numerically for the first two stages. Congestion is explicitly if rather simplistically accounted for in their model, reflected as a decrease in consumer utility as the number of CSPs connected to the platform increases *ceteris paribus*. In contrast to Economides and Tåg's model, users gain utility from the quality, not quantity of CSPs. Recall how this modeling choice actually resonates with Section 3.4's analysis of the supply and demand for Internet content. Here, ISPs can also add utility-increasing features to the platforms. The aforementioned tradeoff between surplus extraction using differential pricing tactics on end users and expropriation of profit from CSPs in the form of termination fees becomes apparent in this model.

Njoroge et al. give predictions of the effects of neutrality regulation on ISP investment, CSP participation, end-user surplus, ISP surplus, and total welfare. The numerical welfare results of the augmented scenarios in the latter part of their paper should not be mistaken for quantitative effect-size predictions: The cardinal utility levels are not meaningful as such, but only through their ordering. They conclude that the non-neutral ISPs invest more than their neutral counterparts do because of increased market power and profit, which increases end-user, CSP, and total surplus at the end of the day. However, the flipside of the coin is that increased market power allows the ISP to raise termination fees, which decreases CSP participation and reduces end-user, CSP, and total surpluses. The net welfare effect of neutrality regulation depends on the relative sizes of the effects, but has the tendency to be negative. Figure 12 displays

Njoroge et al.'s modeling outcomes with different parameter values insofar as CSP participation is concerned.

Figure 12. CSP market coverage. Njoroge et al. graph CSP market coverage as a function of the *reciprocal* of CSP heterogeneity ($\bar{\gamma}/a$) and the demand for connectivity as measured by the mass of end users (f). *Lighter* regions reflect *higher* CSP participation. In regions R_1 and R_2, the CSP market is uncovered, whereas in regions R_3, R_4, and R_5, all CSPs connect to at least one of the platforms. In R_5, all CSPs connect to both platforms. The level of CSP participation is inversely proportional to end-user participation and CSP heterogeneity: High consumer participation and CSP heterogeneity induce the ISPs to extract consumer and CSP surplus by differentiating their QoS and raising the termination fees they charge the more inelastic CSPs, hindering the CSPs' incentives to participate in the market. Source: Njoroge et al. (2013).



5.2 Efficiency and Welfare

The theoretical models related to network neutrality provide predictions of the welfare effects of neutrality. They consider short-run static efficiency and long-run dynamic efficiency. More specifically, economic efficiency can be divided into three classes: (i) *static efficiency*, (ii) *investment efficiency*, and (iii) *innovation efficiency* (Faulhaber 2011). Investment and innovation efficiency are dynamic in nature, as they are only realized over time. These are more difficult to model than static efficiency, a fact manifested as only a portion of the models providing results on the dynamic effects. Table 2 presents a "meta-analysis" of the welfare predictions of virtually all the published models I have been able to find. In total, I include the predictions of 28 formal models from 2007 to 2014. The effect categories evaluated are end-user, ISP, CSP, and social surpluses, which correspond to static efficiency, and ISP investment and CSP innovation, which reflect dynamic efficiency.

Looking at Table 2, it would appear that the predictions of the welfare effects of neutrality regulation, as defined by either a termination-fee cap or no paid prioritization, are mixed. When the predictions are aggregated over the models, network neutrality can be perceived to affect the economic welfare of agents in the following way:

End-user surplus: Ambiguous/negative

ISP surplus: Negative

CSP surplus: Positive

Total welfare: Ambiguous/negative

ISP investment: Negative

CSP innovation: Ambiguous.

Table 2. Welfare effects of network neutrality. The table lists 28 models' predictions of the welfare effects of network neutrality as defined by either a termination-fee cap ("Z") or no paid prioritization ("D,Z"). "+" ("–") denotes an increase (decrease) in the metric under neutrality for at least some parameter values and no decrease (increase) for any parameter value. "0" denotes no change in the metric. "?" denotes that the metric is not evaluated. "⋆" denotes that the change in the metric is ambiguous. "+⋆" ("–⋆") denotes that while the change in the metric depends on parameter values, it generally tends to be positive (negative). The "NN OK?" field has "Yes" ("No") if the overall stance of the article is supportive of (against) network neutrality, or "?" if no such stance can be identified. Adapted, expanded, and corrected from Altman et al. (2012). Sources:

ARWHX = Altman et al. (2014)
B = Baranes (2014)
BKV= Bourreau et al. (2014)
C = Cañón (2009)
CBG = Cheng et al. (2011)
CJK1/2 = Choi et al. (2014a)/(2014b)
CK = Choi and Kim (2010)
DR = D'Annunzio and Russo (2013)
EH = Economides and Hermalin (2012)
ET1/2 = Economides and Tåg (2012)/(2013)
G = Gans (2014)
GBCY = Guo et al. (2010)
GCB1/2 = Guo et al. (2012)/(2013)
HCCR = Hande et al. (2009)
HK = Hermalin and Katz (2007)
JH = Jamison and Hauge (2008)
KKV = Kourandi et al. (2014)
KW = Krämer and Wiewiorra (2012)
MB = Mialon and Banerjee (2014)
MSW = Musacchio et al. (2009)
NOSW = Njoroge et al. (2013)
PS = Peitz and Schuett (2013)
RV = Reggiani and Valletti (2012)
S = Shrimali (2008)
WS = Wang and Sun (2012)

Model	Market Structure		NN Definition	Congestion	End-User Surplus	ISP Surplus	CSP Surplus	Total Welfare	ISP Investment	CSP Innovation	NN OK?
	# ISPs	# CSPs									
ARWHX	1	Many	Z	Yes	+⋆	+⋆	+⋆	+⋆	?	?	Yes
B	1	2	D,Z	No	−	−⋆	−⋆	−⋆	−	?	No
BKV	2	Many	D,Z	Yes	⋆	⋆	⋆	−	−	−	No
C	1	Many	Z	No	+	−	+	+	−	+	Yes
CBG	1	2	D,Z	Yes	−⋆	−	+	−⋆	+	?	No
CJK1	1	Many	D,Z	Yes	⋆	⋆	⋆	−⋆	+⋆	−⋆	Yes
CJK2	1/2	Many	D,Z	No	−⋆	−	⋆	+⋆	?	?	Yes
CK	1	2	D,Z	Yes	−	⋆	+⋆	⋆	+⋆	⋆	?
DR	1/2	2	Z	No	+	−	+	+	?	?	Yes
EH	1	Many	D,Z	Yes	+	−	+	+	−	?	Yes
ET1	1/2	Many	Z	No	⋆	⋆	+	⋆	?	?	?
ET2	1	2	D,Z	No	⋆	−	⋆	⋆	−	?	?
G	1	2	D,Z	No	0	−	+	0	0	+	Yes
GBCY	1	2	D,Z	Yes	−	−	+	−	?	?	No
GCB1	1	2	D,Z	Yes	−	−	+⋆	−	−	+⋆	No
GCB2	1	2	D,Z	Yes	⋆	−	+⋆	−⋆	?	?	No
HCCR	1/Many	Many	Z	Yes	−⋆	−	⋆	−⋆	?	?	No
HK1	1/2	Many	D,Z	No	−⋆	−⋆	⋆	−⋆	?	?	No
KKV	2	2	D,Z	No	+⋆	−	+	+⋆	?	?	?
JH	1	Many	D,Z	Yes	−	−	+	−	−	−	No
KW	1	Many	D,Z	Yes	0	−	+	−	−	−⋆	No
MB	1	Many	Z	No	⋆	−⋆	+⋆	⋆	?	?	?
MSW	Many	Many	Z	No	⋆	⋆	⋆	⋆	⋆	⋆	?
NOSW	2	Many	D,Z	Yes	−⋆	⋆	−⋆	−⋆	−	−⋆	No
PS	1	Many	D,Z	Yes	−⋆	−⋆	⋆	−⋆	?	?	No
RV	1	Many	D,Z	Yes	−⋆	−⋆	−	−⋆	−⋆	⋆	No
S	1	2	D,Z	No	+	−	⋆	+	−	−	Yes
WS	1	1	D,Z	Yes	⋆	−	+⋆	⋆	?	⋆	?
Aggregate Prediction					−⋆	−	+	−⋆	−	⋆	No?

In general, the literature on price discrimination finds its effects somewhat ambiguous but with a tendency to be welfare-enhancing (Kulick and Weisman 2010). The literature on network neutrality, which can be treated as a form of price discrimination, continues along a similar path. Indeed, economists are clearly cautious about network neutrality, and the aggregate sentiment toward neutrality can be condensed as ambiguous-to-negative. A noteworthy remark is that the few models that define neutrality solely as the zero-price rule give ambiguous-to-positive total-welfare predictions; still, the sample size is too small for clear conclusions. Models generally agree with the intuition that ISPs are the gainers from deregulation, whereas CSPs are the losers, end users' position being more ambiguous. We can conclude that if the regulator is a total-welfare maximizer, *economic theory is inclined against neutrality regulation*. If the regulator is biased toward consumers, neutrality regulation still has no theoretical backing. The predicted ambiguous-to-negative dynamic effects of neutrality on ISP investment and CSP innovation can be taken to complement the ambiguous-to-negative static effects.

Based on the models, then, the effect of neutrality regulation on end-user surplus is ambiguous but more likely to be negative than positive. In many models, the negative effect results from higher subscription fees under neutrality due to the ISP not being able to allocate revenue extraction onto the supply side. The dynamic models would together indicate higher ISP investment under non-neutrality, which contributes to higher QoS and, consequently, end-user surplus. On the other hand, some models predict that termination pricing has also a negative impact channel on end-user surplus through reduced CSP participation and a lower amount of content available to end users. Moreover, some models predict that differential pricing allows the ISP to extract more of end users' surplus, as discussed earlier. In the aggregate, the negative end-user-surplus effect channels of neutrality regulation outweigh the positive ones.

Prima facie, the welfare effects of neutrality regulation on the ISP seem clear-cut. Under non-neutrality, the ISP would have the option to charge customers on both sides of the market, force CSPs to internalize the negative congestion externality, and practice differential pricing – clearly the value of this option cannot be negative? As said, most models confirm that the ISP stands to gain from deregulation. Under special circumstances, non-neutrality could conceivably shape the market in such a way that the total ISP revenue decreases, even if the ISP's profit extraction becomes more efficient. If the cost reductions from network management are not sufficient to cover this revenue reduction, the ISP's profit can decrease. In particular, the ISP can be harmed by deregulation if CSP participation in the market decreases substantially.

CSP surplus is predicted to be higher under neutrality by the majority of the models. The logic here is quite straightforward, as termination fees are directly subtracted from CSPs' profits. The possibility of total CSP surplus being higher under non-neutrality mainly emerges when CSPs are heterogeneous: CSPs purchasing a fast lane may benefit more from the preferential delivery than those stuck with the slow lane lose. Additionally, higher ISP investment and reduced overall congestion can contribute to higher total CSP surplus. In general, the idea that smaller CSPs suffer more from non-neutrality is corroborated by the models.

All in all, based on this meta-analysis, the net total welfare effect of neutrality regulation appears to be negative, and in light of economic theory created so far, strong network neutrality should not be the basis for future regulation. However, as I hope to have shown, the neutrality question is infinitely more nuanced than a table of plus and minus signs. Indeed, this formal analysis concerns strong forms of neutrality, i.e. the zero-price rule and the prohibition of data discrimination, neglecting any intermediate degrees of regulation. It does *not* imply that all degrees of regulation are detrimental to economic welfare. In the next subsection, I evaluate some regulatory regimes in the middle of the spectrum.

5.3 Policy Implications

I now briefly discuss the practical policy implications of the analysis so far. Network non-neutrality, while definitely not necessarily detrimental to the market, has to be regarded as a potential issue that the regulator needs to recognize and assess. Policy decisions should ultimately be based on cost-benefit analysis. To set up the policy-analysis mindset, we ask the following questions when considering the regulatory options (Joskow 2007):

1) What is the magnitude of the problem?
2) What policy instruments are available?
3) How would regulation increase market performance?
4) What are the side effects and costs of regulation?

While I have analyzed these factors in the preceding text, the optimal degree of ISP regulation admittedly remains obscure. The optimum policy criterion itself is not easy to define. Even the first, supposedly trivial issue of determining the reference point for policy changes becomes difficult, as the *status quo* is not clearly either neutral or non-neutral. *Pareto efficiency*, where no-one can be better off without someone being worse off, is quite unlikely to be achieved by *any* policy change, and still leaves distributional issues (who are the gainers and losers?) out of the equation. Strong neutrality clearly appears Pareto inefficient. *Kaldor-Hicks efficiency* is a more pragmatic criterion, requiring that the parties better off could in theory compensate those worse off, and that Pareto efficiency could be achieved with some set of transfers (Hicks 1939; Kaldor 1939). If, under neutrality, the increase in CSP or consumer surplus is larger than the decrease in ISP surplus, CSPs and end-users would be able to compensate ISPs, and the society would be better off. In effect, an increase in total welfare corresponds to a Kaldor-Hicks improvement. Distributional efficiency and any transfer costs are still left out of the equation. Policymakers have traditionally sympathized more with consumers than corporations. In any case, even if a consensus among economists emerges against all odds, an ideal solution will not probably be found amid all the political forces. Despite this, we can try to make some sense of the policy implications.

Under neutrality, the ISP recovers its network costs from end users in the retail market.[53] Under supply-side non-neutrality, the ISP recovers its costs from both CSPs and end users, even though the there is only one data stream flowing through the

[53] Network Neutrality in Europe. European Parliamentary Research Service briefing, 2014.

system. Whether this can be thought of as charging twice for the same thing depends on the sum of the fees charged to each side. Were the sum to remain *fixed*, equal to retail prices today, splitting the cost between the CSP and the end user would be merely reallocation, sound ethically tenable, and not reduce total welfare *unless* it had negative dynamic consequences. With discriminatory termination fees, CSPs internalize the negative congestion externality. Assuming a fixed total price in the two-sided market and no dynamic effects, usage-based pricing in the retail or wholesale market or both can be seen as equivalent scenarios with respect to total welfare. Now the choice of the specific regime would be reduced to a distributional issue, a 50-50 split between CSPs and end users being the simplest one. A total price cap is intended to prevent an excess inflation of the total revenue extracted by the ISP.

The reality in its dynamics is more complex, however. The incentives and prices will change in response to regulation. The majority of economists appear to cast a wary eye toward enforcing a strict form neutrality, i.e. the NN regimes or the strictest extensions of the NRT regime, even though the zero-price rule is tempting in its simplicity, an obvious merit for any policy. At the other extreme, total deregulation is likewise tempting in its simplicity (Vogelsang 2013). Were the decision only between total regulation and deregulation, the rational social planner would be inclined to choose to deregulate. As far as the efficiency of the market outcome goes, however, not all middle-road regimes can be ruled out as suboptimal. Hybrid regimes can come at the cost of complexity and, consequently, costs of implementation. Still, weak forms of neutrality, at the least, are worth considering. While certain government approaches such as placing additional taxes on ISPs or CSPs, or even taxing Internet users for each gigabyte transmitted[54] are, to my belief, not warranted in any case, policy instruments potentially applicable to the broadband market are diverse. I cannot comment on the many of these, but restricting the context to neutrality regulation, I next assess some middle-road regulatory proposals.

[54] The Hungarian government recently planned to tax Internet users 150 forints (circa €0.50) per gigabyte, but this was not followed through.

The no-exclusivity rule would be equally simple and a better policy tool than the zero-price rule (Kourandi et al. 2014) to prevent fragmentation. However, the no-exclusivity rule allows the ISP to discriminate CSPs through differential termination pricing even when it cannot directly exclude them. Intuitively, then, the no-exclusivity rule would appear to work best in combination with uniform termination pricing. Without network regulation, the no-exclusivity rule does not eliminate the possibility of tacit collusion between ISPs and CSPs where they implicitly coordinate on content priorities – the payment from the CSP to the ISP would be made in the form of direct interconnect deals, for example.

Any cost-based regulatory approach to the broadband market poses significant challenges. One proposal is neutrality regulation in the form of regulated termination fees. The price regulation model could follow those in place in the electricity distribution markets. The regulator enforces a price cap such that termination fee $f \leq MC + a$, i.e. only margin a would be allowed over the marginal cost of bandwidth. Then, the termination fee is higher for data-intensive CSPs to reflect the ISP's capacity provisioning costs. However, high sunk costs in the broadband market may render cost-plus regulation infeasible (Hausman 1999; Sidak 2006). At the fundamental level, Owen and Rosston (2003) conclude: "[If] we assign property rights in access to users rather than suppliers, resulting in an efficient price of access (zero), there will be no long run supply of Internet services. [...] While the benefits of the Internet can be made available to any *particular* user at zero cost, they cannot be made available to *all* users at zero cost." Pricing at marginal cost may be statically efficient but falls short on dynamic efficiency.

The most popular pricing rule in the telecommunications literature has been the *efficient component pricing rule* (ECPR), where the operator charges an interconnection fee equal to the incremental capacity provisioning costs plus an "opportunity cost" (Vogelsang 2003). The opportunity cost emerges from the possibility that the customer's products and services displace the operator's own, so the operator effectively facilitates competition against itself. Under special conditions, the ECPR coincides with Ramsey pricing. The ECPR and its derivatives, however, have been criticized as either inefficient in most scenarios or too complicated when efficient (Vogelsang 2003).

In addition to the theoretical obstacles with cost-plus regulation, immediate pragmatic difficulties arise with calculating costs in sufficiently general terms (Section 3.3). Moreover, the trickiness of cost accounting in telecommunications makes the optimal tuning of cost-plus regulation challenging. As the OECD report (referred to in Figure 5) finds, "A large number of different cost allocations (and therefore prices) are consistent with the regulated firm recovering its costs." And even if these challenges are overcome, cost-plus regulation has heavy information requirements, and is expensive to implement and monitor. As is known, cost-based price regulation has also other problems, such as weak cost-reduction incentives (Section 3.1) and the tendency to evolve into convoluted regulatory schemes.

Price regulation sometimes has perverse effects on the prices. As Armstrong (2014) recognizes, a binding price ceiling in some cases causes the average price to rise. Moreover, the two-sided nature of the broadband market complicates things. Regulating prices on one side may put upward pressure on prices on the other side (Vogelsang 2013). This phenomenon is called the *waterbed effect* or the "seesaw principle." If termination fees are capped, subscription fees may rise in the retail market, unless those are regulated, as well, and would result in inefficiencies (Bourreau and Lestage 2013; Jullien and Sand-Zantman 2014). Inderst and Peitz (2012) find an analogous effect with access-price regulation. A simulation by Shapira (2004) would point to the opposite, predicting a higher consumer surplus higher in a price regulated, single product-line telecommunications industry than in free competition, which has the tendency to evolve into a natural monopoly. The same conclusion would apply to a regulation scheme where innovations of a monopoly firm are transferred (at a cost) to competing firms.

In terms of implementation, a much simpler regulatory proposal than cost-plus regulation would be to enforce a uniform termination fee across CSPs, leaving the price otherwise unregulated. Uniform termination pricing would mitigate the problem of the ISP being able to discriminate CSPs, although Waterman (2012) notes that, analogously to the waterbed effect, a non-discrimination rule in the wholesale market might simply shift the ISP's discriminatory behavior downstream into the retail market unless the retail side is regulated, as well. On the other hand, termination pricing need not have dubious motives behind it, as it can be justified as the first-best solution in mitigating the negative externality of congestion – prioritization, throttling, and data caps being second-best alternatives (Crocioni 2011). Uniform termination fees as such would not address the problem of inefficiency from the negative externality data-intensive CSPs create. Hence, uniform termination pricing would probably work best in combination with network management. Continuing on the same line of reasoning, neutrality in the form of the zero-price rule is inefficient because it turns the network into a common resource, and the tragedy of the commons is not resolved as long as CSPs do not internalize the externality they create (Peitz and Schuett 2013).

In principle, a straightforward form of regulation is to outlaw throttling and blocking. This sort of congestion-based regulation (as proposed in e.g. Frieden 2014) would be resource-consuming due to the information and reporting requirements; the regulator is required to monitor a multitude of technical parameters to ensure that prioritization always has a technical justification. As hinted in Section 2.2, drawing the line between network management and throttling is exceedingly difficult.

A minimum-QoS standard, already in place in some EU countries such as Finland, is a middle-road regime that would prevent Internet fragmentation while leaving room for ISPs to offer differentiated broadband services, optimize their networks, and charge CSPs for congestion. Current regulation does not explicitly address whether or not the end user's connection can be QoS- or access-tiered, but only specifies the required average QoS; the interpretation of the law seems somewhat unclear at the moment. To be effective, minimum-QoS regulation should cover both sides of the market so that each CSPs is provided a slow lane and each end user is provided a neutral slow lane at a minimum. (A guaranteed full-access slow lane on the supply side implies a neutral slow lane on the demand side.) A minimum-QoS standard would not suppress innovation at the core to a significant degree, although setting the standard too high (as potentially pursued by industry incumbents with high-performance infrastructure in place) would further suppress ISP market entry (Brennan 2011). Setting the standard at an optimal level is the main challenge in implementing the policy.

Based on the observations so far, we can conclude that not only is strong neutrality regulation cautioned against by economists but also middle-road regimes appear to cause difficulties. Cost-based ISP regulation may be inefficient, low-set price caps on either side of the market may have negative side effects, and congestion-based regulation may have severe implementation obstacles. Minimum-QoS regulation might provide one of the less bad compromises between effective regulation and costs of implementation. Given that regulatory intervention does not seem to lead to an optimal market design, is it possible for a market-driven equilibrium solution to emerge as an alternative? As has been established, this is unlikely due to the conflicting interests of the stakeholders. An opposing argument is presented by Knieps and Stocker (2014) who predict that weaker market-driven network management (NMZ and NMT regimes) built on the DiffServ architecture would be the deregulated outcome. This equilibrium would stem from the fact that "only a price and quality differentiation strategy based on the opportunity costs of traffic capacity usage can be stable." Further, the authors argue that product differentiation based on DiffServ is economically efficient. It remains to be seen how DiffServ and other architectural developments of the Internet pan out on a large scale.

While the ISP's revenue scales with the long-run demand for connectivity, its capacity provisioning costs grow with the peak load on the network, as the ISP must overprovision and cannot let the network become too congested at any point (Joe-Wong et al. 2013). This mismatch is a problem that can be alleviated by either throttling or more efficient pricing. Flat-rate pricing is suboptimal, leaving the timescale mismatch: The ISP's revenue depends on fixed monthly fees, but capacity provisioning for peak-hour demand dominates its cost. With flat pricing, lighter users in effect subsidize the bandwidth consumption of heavier users, who have little incentive to economize it (Liu 2004; Lyons 2013).

ISPs have not so far partaken in pricing practices where the price of connectivity is based on network congestion or interconnection costs at a given moment, analogously to congestion pricing used with e.g. electricity distribution and toll roads. Instead, some ISPs' answer to increased traffic volumes has been to start experimenting with monthly data caps and overage charges. This is close but not equivalent to pure usage-based pricing, which would target precisely the negative congestion externality, and therefore the reintroduction of data caps is not surprising. Somewhat surprising, though, is why full-blown usage-based pricing has yet to resurface. The reason might ultimately boil down to the stickiness of flat-fee traditions and marketing, predictability, and transparency considerations. Crucially, while usage-based pricing mitigates the congestion externality, it is still suboptimal: It does not address the timescale mismatch problem. Recall that the ISP's costs are determined by peak consumption. In plain usage-based pricing, the unit price is independent of the congestion level at a given moment. Yet another pricing method would be needed to address this problem.

A key market-driven possibility that might provide a compromise between economic efficiency and the abovementioned marketing concerns is *dynamic pricing*. Dynamic pricing, as opposed to discriminatory pricing, could give an out-of-the-box way to circumvent the neutrality question altogether. Perhaps the most elegant novel pricing method for the broadband market is *time-dependent usage pricing* (TDP), where the unit price depends on the time of the day (or day of the week etc.). TDP would induce heavier users to economize their bandwidth consumption and thus help reduce demand volatility and spread traffic more evenly over the day (Joe-Wong et al. 2013). Admittedly, while congestion correlates with the time of the day, TDP is strictly speaking a fundamentally "static" pricing model, as it is not directly based on the actual degree of congestion at a given moment (Liu 2004). On the other hand, TDP could be adequately convenient for the end user due to its predictability compared to real-time congestion-based pricing. At any rate, TDP would align the ISP's and end users' interests and enable at least partial internalization of the negative network externality, as those who exert load on the network at the prime time would be the ones to compensate the ISP. Using simulations, Gupta et al. (2011) find congestion-based pricing generally socially beneficial compared to flat-rate pricing.

It might be enough to use TDP in the retail market without network or price regulation. TDP already "punishes" data-intensive CSPs' users and hence indirectly the CSPs according to the bandwidth their users demand at the most congested hours. Thus termination pricing would not be warranted. If worried about transparency or excess price fluctuations, the regulator might decide to limit the fluctuation ratio as it sees fit – a straightforward and easily enforceable rule. Dynamic pricing would have win-win potential from the neutrality-debate perspective, as it would not discriminate users as harshly based on their traffic profiles as flat-rate pricing does, and could reduce peak loads. (As noted earlier, while perhaps fair-sounding, flat pricing is actually discriminatory in an economic sense when lighter users are subjected to higher common subscription fees due to congestion costs caused by heavier users.)

Therefore, dynamic pricing might perhaps be a realistically simple and effective market-driven remedy to the neutrality problem. The appropriate policy action would then be to initiate discussion to encourage ISPs to devise dynamic rather than discriminatory pricing schemes. Cyclic pricing could be tolerated by the public, provided it is practiced in a sufficiently transparent manner. The skeptical reader might ponder whether ISPs would end up using TDP more for price-discrimination purposes than congestion avoidance. As Lyons (2013) emphasizes, while usage-based (or dynamic) pricing could be used in an anticompetitive manner, the specific pricing scheme in itself is not the main question as much as the degree of market power of ISPs, which can prove to be a problem in the neutral and non-neutral worlds alike.

Qualitative reviews on the neutrality question tend to continue along the line of thought that allowing market-driven solutions and monitoring for potential abuses is the most flexible regulation model (e.g. Becker et al. 2010; Lyons 2013; Marsden 2007; Yoo 2006a, 2006b). In other words, it is quite widely maintained that *ex-ante* neutrality regulation is not warranted, and an *ex-post* competition-policy approach remains a better regulatory regime. In the competition-policy context, the definition of the relevant market at both the demand and the supply side becomes an essential issue. Yoo (2006a) provides a key insight to this end: "[CSPs] care about the total number of users they can reach. So long as their total potential customer base is sufficiently large [...], the fact that a particular network owner may refuse carriage in any particular locality is of no consequence." In other words, an ISP's power over a CSP with respect to a subset of end users does not always restrict the CSP's activity to a significant degree in relative terms. In this light, neutrality regulation seems less relevant in the case of a small ISP, at least from the CSP-innovation perspective. Further, relieving small ISPs of regulation, even if larger ISPs were more strictly regulated, would pave way for network diversity and competition. Indeed, competition policy generally targets the large players.

In conclusion, Table 3 summarizes some of the economic arguments for and against network neutrality.

Table 3. Economic arguments for and against network neutrality.

Network Neutrality	Network Non-Neutrality
Higher theoretical CSP surplus (very likely)	Higher theoretical ISP surplus (very likely)
Higher theoretical consumer surplus (unlikely)	Higher theoretical consumer surplus (likely)
Higher theoretical total welfare (unlikely)	Higher theoretical total welfare (likely)
High ISP market power: More tools to abuse a dominant position under non-neutrality	Neutrality does not rid ISPs of market power. *Ex-post* competition policy can be used to address abuse
Internet infrastructure exhibits natural-monopoly characteristics	More incentives for ISPs to invest in infrastructure
Vertical integration can increase ISPs' market power and reduce competition among CSPs	Vertical integration is often economically efficient
Price discrimination can be welfare-reducing	Price discrimination is often welfare-enhancing
Market-driven solutions, e.g. dynamic or usage-based pricing, can emerge allowing ISPs to mitigate the congestion externality	Economic efficiency from internalizing the congestion externality
CSPs, especially small CSPs, likely to be harmed by non-neutrality. Discourages innovation at the edge	Neutrality discourages innovation at the core, hindering the evolution of the Internet and preventing efficient network operation
Termination pricing can hinder competition among CSPs and create barriers to entry	Price regulation can cause a waterbed effect

Can limit ISP's market power and increase competition in the market	Can foster network diversity and increase competition in the market
Strong neutrality straightforward to implement as an *ex-ante* rule	Complexity of the market makes effective regulation difficult. Reporting and monitoring can be costly for ISPs and regulators
Neutrality protects the diffusion of positive externalities on the Internet. Threat of fragmentation under non-neutrality	Title II-type regulation can lead to potentially inefficient government intervention and taxation

6. Conclusions

In this paper, I have analyzed the economics and regulation of the broadband market through the lens of the so-called network-neutrality debate. I have discussed the characteristics, developments, and models of the market. Further, I have evaluated the qualitative economic effects of network neutrality and alternative regulatory regimes.

To sum up, my analysis suggests that while the broadband industry exhibits characteristics comparable to those of more regulated industries, such as the electricity distribution market, the complexity of the Internet ecosystem differentiates the industry from many others. This complexity does not stem merely from the physical layouts of the networks but from the business relations between market participants, and makes effective regulation difficult. Models from economic theory give mixed results on the welfare effects of network neutrality. There is no economic evidence that network neutrality generally increases total welfare. In fact, it turns out that from a well-rounded economic perspective, strong network neutrality appears in most cases as detrimental to both consumer surplus and total welfare. In certain scenarios, however, models predict that neutrality to a degree increases static and dynamic efficiency. The results depend crucially on model specifications and parameters, which differ significantly across the literature. So far, there is no consensus among economists on the optimal level of ISP regulation. Market-driven solutions such as dynamic pricing might provide a way to circumvent the neutrality question.

At this point, I should make a couple of concluding remarks about the limitations in my analysis. First, not all of the subject matter has been previously covered in the literature as far as I am aware, and a certain amount of exploration, anecdotal evidence, and common sense is involved. Second, I have discarded the possibility of regulating the membership fees CSPs can charge end users. This would be questionable and have many ramifications. Third, I have not considered policies where the subscription fee is regulated but termination fee deregulated, as these sound unrealistic under current legislative developments that revolve around data discrimination and termination pricing rather than purely retail price regulation. Were these included, we would have an additional column in Table 1. Fourth, I have not considered simply nationalizing the principal data routes or municipalizing residential access ISPs, which has also been proposed.[55] (Municipal Internet provision alongside private ISPs could, however, be economically sustainable, as concluded in Hauge et al. [2008].) Finally, I have not ventured far into the realm of possible ISP regulation policies unrelated to network neutrality. Following Claffy and Clark (2014), different levels of "structural separation" or "open interfacing" (conceptually related to unbundling) at different layers of the Internet ecosystem provide particularly interesting possibilities for future exploration.

The neutrality debate is not merely about technical pedantry but has ramifications for our information society at large. The recent and upcoming regulatory decisions will reverberate through the entire ecosystem, affecting both the random surfer and the society. They are, however, unlikely to be enough to settle the accounts; the debate is likely go on for a long time. In the end, I believe regulation ought to be looked at more from a pragmatic cost-benefit analysis perspective than as a purely ethical principle. The complexity of the broadband market makes not only regulation but also modeling highly difficult. Nevertheless, the astonishing diversity of Internet content and services calls for models that try to capture more of this heterogeneity. Fortunately, if the trend of the last year or two continues, we are going to see more research about network neutrality in the near future. More generally, there is still a lot to study about the economics of the broadband market. In particular, the interconnection market is something most models have abstracted away. What I would find meaningful are still

[55] See e.g. www.nytimes.com. Accessed 27 Dec 2014.

more complete models of the broadband market that see the forest for the trees, for these lay the foundation for successful policy analysis and decision-making. Now is the time to contribute to what the future of the Internet will look like.

References

Afèche, P. (2006). Economics of Data Communications. *Handbooks in Information Systems* 1, 53-135, Elsevier.

Altman, E., Hanawal, M., and Sundaresan, R. (2014). Regulation of Off-Network Pricing in a Nonneutral Network. *ACM Transactions on Internet Technology* 14(2).

Altman, E., Legout, A., Xu, Y. (2011). Network Non-Neutrality Debate: An Economic Analysis. *Lecture Notes in Computer Science* 6641, 68-81, Springer.

Altman, E., Rojas, J., Wong, S., Hanawal, M., Xu, Y. (2012). Net Neutrality and Quality of Service. *Game Theory for Networks*, 137-152, Springer.

Amelio, A. and Jullien, B. (2012). Tying and Freebies in Two-Sided Markets. *International Journal of Industrial Organization* 30(5), 436-446.

Anderson, S. (2003). Market Provision of Broadcasting: A Welfare Analysis. *Review of Economic Studies* 72(4), 947-972.

Armstrong, M. (2002). The Theory of Access Pricing and Interconnection. *Handbook of Telecommunications Economics* 1, 297-386, North-Holland.

Armstrong, M. (2006). Competition in Two-Sided Markets. *RAND Journal of Economics* 37(3), 668-691.

Armstrong, M. (2014). Search and Ripoff Externalities. Economics Series Working Papers 715, University of Oxford.

Baranes, E. (2014). The Interplay between Network Investment and Content Quality: Implications to Net Neutrality on the Internet. *Information Economics and Policy* 28, 57-69.

Becker, G., Carlton, D., Sider, H. (2010). Net Neutrality and Consumer Welfare. *Journal of Competition Law & Economics* 6(3), 497-519.

Beckert, W. (2005). Estimation of Heterogeneous Preferences, with an Application to Demand for Internet Services. *Review of Economics and Statistics* 87(3), 495-502.

Besen, S. and Israel, M. (2013). The Evolution of Internet Interconnection from Hierarchy to "Mesh": Implications for Government Regulation. *Information Economics and Policy* 25(4), 235-245.

Bourreau, M., Kourandi, F., Valletti, T. (2014). Net Neutrality with Competing Internet Platforms. *Journal of Industrial Economics*, forthcoming.

Bourreau, M., Lestage, R. (2013). Net Neutrality and Asymmetric Platform Competition. Télécom ParisTech working paper.

Brennan, T. (2011). Net Neutrality or Minimum Standards: Network Effects vs. Market Power Justifications. *Network Neutrality and Open Access*, 61-78, Nomos Publishers.

Caesar, M. and Rexford, J. (2005). BGP Routing Policies in ISP Networks. *Network, IEEE* 19(6), 5-11.

Caillaud, B. and Jullien, B. (2003). Chicken & Egg: Competition among Intermediation Service Providers. *RAND Journal of Economics* 34(2), 309-328.

Calzada, J. and Martínez-Santos, F. (2014). Broadband Prices in the European Union: Competition and Commercial Strategies. *Information Economics and Policy* 27, 24-38.

Cambini, C. and Jiang, Y. (2009). Broadband Investment and Regulation: A Literature Review. *Telecommunications Policy* 33(10), 55-574.

Cañón, C. (2009). Regulation Effects on Investment Decisions in Two-Sided Market Industries: The Net Neutrality Debate. Toulouse School of Economics working paper.

Cheng, H., Bandyopadhyay, S., Guo, H. (2011). The Debate on Net Neutrality: A Policy Perspective. *Information Systems Research* 22(1), 60-82.

Choi, J. (2010). Tying in Two-Sided Markets with Multi-Homing. *Journal of Industrial Economics* 53(3), 607-626.

Choi, J. and Kim, B. (2010). Net Neutrality and Investment Incentives. *RAND Journal of Economics* 41(3), 446-471.

Choi, J., Jeon, D., Kim, B. (2014a). Asymmetric Neutrality Regulation and Innovation at the Edges: Fixed vs. Mobile Networks. CESifo Working Paper Series 4974.

Choi, J., Jeon, D., Kim, B. (2014b). Net Neutrality, Business Models, and Internet Interconnection. *American Economic Journal: Microeconomics*, forthcoming.

Claffy, K. and Clark, D. (2014). Platform Models for Sustainable Internet Regulation. *Journal of Information Policy* 4, 463-488.

Crocioni, P. (2011). Net Neutrality in Europe: Desperately Seeking a Market Failure. *Telecommunications Policy* 35(1), 1-11.

D'Annunzio, A. and Russo, A. (2013). Network Neutrality, Access to Content and Online Advertising. KOF Working Papers 344, ETH Zurich.

Economides, N. and Hermalin, B. (2012). The Economics of Network Neutrality. *RAND Journal of Economics* 43(4), 602-629.

Economides, N. and Tåg, J. (2012). Network Neutrality on the Internet: A Two-Sided Market Analysis. *Information Economics and Policy* 24(2), 91-104.

Economides, N. and Tåg, J. (2013). Network Neutrality and Network Management Regulation: Quality of Service, Price Discrimination, and Exclusive Contracts. *Research Handbook on Governance of the Internet*, 21-142.

Eisenmann, T., Parker, G., van Alstyne, M. (2006). Strategies for Two-Sided Markets. *Harvard Business Review* 84(10), 92-101.

Faulhaber, G. (2011). Economics of Net Neutrality: A Review. *Communications & Convergence Review* 3(1), 53-64.

Frieden, R. (2014). New Models and Conflicts in the Interconnection and Delivery of Internet-Mediated Content. *Telecommunications Policy* 38(11), 970-978.

Ganley, P. and Allgrove, B. (2006). Net Neutrality: A User's Guide. *Computer Law & Security Report* 22(6), 454-463.

Gans, J. (2014). Weak versus Strong Net Neutrality. NBER working paper.

Götz, G. (2013). Competition, Regulation, and Broadband Access to the Internet. *Telecommunications Policy* 37(11), 1095-1109.

Greenstein, S. and McDevitt, R. (2011). The Broadband Bonus: Estimating Broadband Internet's Economic Value. *Telecommunications Policy* 35(7), 617-632.

Guo, H., Bandyopadhyay, S., Cheng, H., Yang, Y. (2010). Net Neutrality and Vertical Integration of Content and Broadband Services. *Journal of Management Information Systems* 27(2), 243-275.

Guo, H., Cheng, H., Bandyopadhyay, S. (2012). Net Neutrality, Broadband Market Coverage, and Innovation at the Edge. *Decision Sciences* 43(1), 141-172.

Guo, H., Cheng, H., Bandyopadhyay, S. (2013). Broadband Network Management and the Net Neutrality Debate. *Production and Operations Management* 22(5), 1287-1298.

Gupta, A., Jukic, B., Stahl, D., Whinston, A. (2011). An Analysis of Incentives for Network Infrastructure Investment Under Different Pricing Strategies. *Information Systems Research* 22(2), 215-232.

Hagiu, A. (2006). Pricing and Commitment by Two-Sided Platforms. *RAND Journal of Economics* 37(3), 720-737.

Hanawal, M. and Altman, E. (2013). Network Non-Neutrality through Preferential Signaling. arXiv:1303.4199.

Hande, P., Chiang. M., Calderbank, R., Rangan, S. (2009). Network Pricing and Rate Allocation with Content Provider Participation. INFOCOM 2009, IEEE, 990-998.

Hauge, J., Jamison, M., Gentry, R. (2008). Bureaucrats as Entrepreneurs: Do Municipal Telecommunications Providers Hinder Private Entrepreneurs? *Information Economics and Policy* 20(1), 89-102.

Hausman, J. (1999). The Effect of Sunk Costs in Telecommunications Regulation. *Topics in Regulatory Economics and Policy* 34, 191-204, Springer.

Hazlett, T. and Wright, J. (2012). The Law and Economics of Network Neutrality. *Indiana Law Review* 45(3), 767-840.

Hemphill, S. (2008). Network Neutrality and the False Promise of Zero-Price Regulation. *Yale Journal on Regulation* 25(2), 135-180.

Hermalin, B. and Katz, M. (2007). The Economics of Product-Line Restrictions with an Application to the Network Neutrality Debate. *Information Economics and Policy* 19(2), 215-248.

Hicks, J. (1939). The Foundations of Welfare Economics. *Economic Journal* 49(196), 696-712.

Hill, J. (2012). Internet Fragmentation: Highlighting the Major Technical, Governance and Diplomatic Challenges for U.S. Policy Makers. Belfer Center for Science and International Affairs thesis, Harvard University.

Hogendorn, C. (2007). Broadband Internet: Net Neutrality versus Open Access. *International Economics and Economic Policy* 4(2), 185-208.

Hotelling, H. (1929). Stability in Competition. *Economic Journal* 39(153), 41-57.

Inderst, R. and Peitz, M. (2012). Network Investment, Access and Competition. *Telecommunications Policy* 36(5), 407-418.

Jahn, E. and Prüfer, J. (2008). Interconnection and Competition among Asymmetric Networks in the Internet Backbone Market. *Information Economics and Policy* 20(3), 243-256.

Jamison, M. (2012). Methods for Increasing Competition in Telecommunications Markets. University of Florida working paper.

Jamison, M. and Hauge, J. (2008). Getting What You Pay For: Analyzing the Net Neutrality Debate. University of Florida working paper.

Joe-Wong, C., Ha, S., Sen, S., Chiang, M. (2013). Optimized Dynamic Pricing for Broadband Access. Princeton University working paper.

Johansson, K., Zander J., Furuskär, A. (2007). Modelling the Cost of Heterogeneous Wireless Access Networks. *International Journal of Mobile Network Design and Innovation* 2(1), 58-66.

Joskow, P. (2007). Regulation of Natural Monopoly. *Handbook of Law and Economics* 2, 1227-1348, Elsevier.

Jullien, B. and Sand-Zantman, W. (2013). Pricing Internet Traffic: Exclusion, Signalling, and Screening. IDEI Working Papers 735.

Kaldor, N. (1939). Welfare Propositions in Economics and Interpersonal Comparisons of Utility. *Economic Journal* 49(195), 549-552.

Kleinrock, L. (1967). Optimum Bribing for Queue Position. *Operations Research* 15(2), 304-318.

Knieps, G. and Stocker, V. (2014). Market Driven Network Neutrality and the Fallacy of a Two-Tiered Internet Traffic Regulation. Institut für Verkehrswissenschaft und Regionalpolitik Discussion Paper 149, University of Freiburg.

Kotakorpi, K. (2006). Access Price Regulation, Investment and Entry in Telecommunications. *International Journal of Industrial Organization* 24(5), 1013-1020.

Kourandi, F., Krämer, J., Valletti, T. (2014). Net Neutrality, Exclusivity Contracts and Internet Fragmentation. *Information Systems Research*, forthcoming.

Krämer, J. and Vogelsang, I. (2014). Co-Investments and Tacit Collusion in Regulated Network Industries: Experimental Evidence. International Telecommunications Society working paper.

Krämer, J. and Wiewiorra, L. (2012). Network Neutrality and Congestion Sensitive Content Providers: Implications for Content Variety, Broadband Investment, and Regulation. *Information Systems Research* 23(4), 1303-1321.

Krämer, J., Wiewiorra, L., Weinhardt, C. (2013). Net Neutrality: A Progress Report. *Telecommunications Policy* 37(9), 794-813.

Krishnan, S. and Sitaraman R. (2012). Video Stream Quality Impacts Viewer Behavior: Inferring Causality Using Quasi-Experimental Designs. *Proceedings of the 2012 ACM Conference on Internet Measurement Conference*, 211-224.

Kulick, R. and Weisman, D. (2010). Price Discrimination, Two-Sided Markets and Net Neutrality Regulation. *Tulane Journal of Technology & Intellectual Property* 13, 82-101.

Laffont, J. and Tirole, J. (2000). Competition in Telecommunications. MIT Press.

Leal, M. (2014). The EU Approach to Net Neutrality: Network Operators and Over-the-Top Players, Friends or Foes? *Computer Law & Security Review* 30(5), 506-520.

Lee, D. and Kim, Y. (2014). Empirical Evidence of Network Neutrality – The Incentives for Discrimination. *Information Economics and Policy* 29, 1-9.

Lee, R. and Wu, T. (2009). Subsidizing Creativity through Network Design: Zero-Pricing and Net Neutrality. *Journal of Economic Perspectives* 23(3), 61-76.

Liu, W. (2003). Congestion Based Pricing. Helsinki University of Technology working paper.

Lotfi, M., Kesidis, G., Sarkar, S. (2014). Network Non-Neutrality on the Internet: Content Provision under a Subscription Revenue Model. University of Pennsylvania working paper.

Lyons, D. (2013). Internet Policy's Next Frontier: Usage-Based Broadband Pricing. Legal Studies Research Paper Series 302, Boston University.

Maister, D. (1985). The Psychology of Waiting Lines. *The Service Encounter*, 113-123, Lexington Books.

Marsden, C. (2007). Net Neutrality and Consumer Access to Content. *SCRIPTed* 4(4).

Maxwell, W. and Brenner, D. (2012). Confronting the FCC Net Neutrality Order with European Regulatory Principles. *Journal of Regulation* 2.

Mialon, S. and Banerjee, S. (2014). Regulation of Internet Access. Emory University working paper.

Musacchio, J., Schwartz, G., Walrand, J. (2009). A Two-Sided Market Analysis of Provider Investment Incentives with an Application to the Net-Neutrality Issue. *Review of Network Economics* 8(1), 1-18.

Njoroge, P., Ozdaglar, A., Stier, N., Weintraub, G. (2009). Competition, Market Coverage, and Quality Choice in Interconnected Platforms. NetEcon '09 discussion paper, Stanford.

Njoroge, P., Ozdaglar, A., Stier-Moses, N., Weintraub, G. (2013). Investment in Two-Sided Markets and the Net Neutrality Debate. *Review of Network Economics* 12(4), 355-402.

Nurski, L. (2014). Net Neutrality and Online Innovation: An Empirical Study of the UK. University of Leuven working paper.

Owen, B. and Rosston, G. (2003). Local Broadband Access: Primum Non Nocere or Primum Processi? A Property Rights Approach. Stanford Institute for Economic Policy Research Discussion paper 02-37.

Parker, G. and van Alstyne, M. (2005). Two-sided Network Effects: A Theory of Information Product Design. *Management Science* 51(10), 1494-1504.

Peitz, M. and Schuett, F. (2013). Net Neutrality and Inflation of Traffic. University of Mannheim and Tilburg University working paper.

Ramsey, F. (1927). A Contribution to the Theory of Taxation. *Economic Journal* 37(145), 47-61.

Reggiani, C. and Valletti, T. (2012). Net Neutrality and Innovation at the Core and at the Edge. The School of Economics Discussion Paper Series 1202, University of Manchester.

Rochet, J. and Tirole, J. (2003). Platform Competition in Two-Sided Markets. *Journal of the European Economic Association* 1(4), 990-1029.

Rochet, J. and Tirole, J. (2006). Two-Sided Markets: A Progress Report. *RAND Journal of Economics* 37(3), 645-667.

Rosston, G. and Topper, M. (2010). An Antitrust Analysis of the Case for Wireless Network Neutrality. *Information Economics and Policy* 22(1), 103-119.

Rosston, G., Savage, S., Waldman, D. (2010). Household Demand for Broadband Internet Service. Report to Federal Communications Commission.

Rysman, M. (2009). The Economics of Two-Sided Markets. *Journal of Economic Perspectives* 23(3), 125-143.

Shapira, G. (2004). System Dynamics Simulation of the Telecom Industry. MIT MSc thesis.

Shin, D. (2014). A Comparative Analysis of Net Neutrality: Insights Gained by Juxtaposing the U.S. and Korea. *Telecommunications Policy*, forthcoming.

Shleifer, A. (1985). A Theory of Yardstick Competition. *RAND Journal of Economics* 16(3), 319-327.

Shrimali, G. (2008). Surplus Extraction by Network Providers: Implications for Net Neutrality and Innovation. *Telecommunications Policy* 32(8), 545-558.

Sidak, G. (2006). A Consumer-Welfare Approach to Network Neutrality Regulation of the Internet. *Journal of Competition Law and Economics* 2(3), 349-474.

Tangerås, T. (2014). Network Competition with Income Effects. *RAND Journal of Economics* 45(3), 645-673.

van Schewick, B. (2007). Towards an Economic Framework for Network Neutrality Regulation. *Journal on Telecommunications and High Technology Law* 5(2), 329-391.

Vogelsang, I. (2003). Price Regulation of Access to Telecommunications Networks. *Journal of Economic Literature* 41(3), 830-862.

Vogelsang, I. (2006). Access and Interconnection Pricing Issues in Telecommunications. CESifo DICE Report 4(3), 3-9.

Vogelsang, I. (2013). The Endgame of Telecommunications Policy? A Survey. CESifo Working Paper Series 4545.

Vogelsang, I. (2014). Will the U.S. and EU Telecommunications Policies Converge? A Survey. CESifo Working Paper Series 4843.

Wallsten, S. and Hausladen, S. (2009). Net Neutrality, Unbundling, and their Effects on International Investment in Next-Generation Networks. *Review of Network Economics* 8(1), 90-112.

Wang, T. and Sun H. (2012). Net Neutrality, Boon or Bane for Free File Sharing Service? Toulouse School of Economics working paper.

Waterman, D. (2012). Network Neutrality and "Discrimination Shifting". Indiana University working paper.

Waterman, D. and Choi, S. (2012). Non-Discrimination Rules for ISPs and Vertical Integration: Lessons from Cable Television. *Telecommunications Policy* 35(11), 970-983.

Weyl, G. (2010). A Price Theory of Multi-Sided Platforms. *American Economic Review* 100(4), 1642-1672.

Wu, T. (2003). Network Neutrality, Broadband Discrimination. *Journal of Telecommunications and High Technology Law* 2, 141-179.

Yoo, C. (2006a). Network Neutrality and the Economics of Congestion. Vanderbilt Law and Economics Research Paper 05-28.

Yoo, C. (2006b). Network Neutrality and Competition Policy: A Complex Relationship. *Net Neutrality or Net Neutering: Should Broadband Internet Services be Regulated?*, 25-71, Springer.

Zhu, K. (2007). Bringing Neutrality to Net Neutrality. *Berkeley Technology Law Journal* 22(1), 61-645.

Appendix: Acronyms

AC = Average cost
BEREC = Body of European Regulators for Electronic Communications
CDN = Content delivery network, content distribution network
CNP = Content network platform
CSP = Content/service provider
DDR = Data discrimination with price regulation
DDZ = Data discrimination with the zero-price rule
DiffServ = Differentiated services
DPI = Deep packet inspection
DSL = Digital subscriber line; xDSL: Umbrella for DSL technologies
ECPR = Efficient component pricing rule
FCC = Federal Communications Commission
FICORA = Finnish Communications Regulatory Authority
FIFO = First-in-first-out
FTTH = Fiber-to-the-home; FTTx = Fiber-to-the-x: Umbrella for fiber local-loop technologies
GB = Gigabyte; Gbps = Gigabits per second
GSM = Global System for Mobile Communications
HTTP = Hypertext Transfer Protocol
IP = Internet Protocol
IRU = Indefeasible right to use
ISP = Internet service provider
IXP = Internet exchange point
LTE-A = LTE Advanced; LTE = Long-Term Evolution
MB = Megabyte; Mbit = Megabit; Mbps = Megabits per second
MC = Marginal cost
NMR = Network management with price regulation
NMT = Network management with two-sided pricing
NMZ = Network management with the zero-price rule
NN = Network neutrality
NNN = Network non-neutrality
NR = Network regulation
NRT = Network regulation with two-sided pricing
P2P = Peer-to-peer
PoP = Point of presence
PPP = Purchasing power parity
QoS = Quality of service
SLA = Service-level agreement
SPE = Subgame perfect Nash equilibrium
TC = Total cost
TCP = Transmission Control Protocol
TDP = Time-dependent usage pricing
TSP = Telecommunications service provider
VoIP = Voice-over-IP
WWW = World Wide Web