

Neighbor Billing and Network Neutrality

JUSTIN (GUS) HURWITZ[†]

ABSTRACT

This Article considers the ongoing legal, political, and technical debates about network neutrality. It argues that the common understanding of this debate is misplaced: all complex networks, including the Internet, are inherently non-neutral. Indeed, this non-neutrality stems from the very architectural features that make the Internet a cost-effective means of communication. These debates are better understood as being about the allocation of the costs of the Internet infrastructure. Under the current rules, there is concern that Internet access providers can exploit market failures to unduly push their costs onto content providers; but the remedy proposed by the content providers—prohibiting differential pricing—creates its own market failures that unduly assign costs to the access providers. Instead of focusing on whether different parties are “neutrally” charged for Internet access, this Article focuses on how best to allocate these costs to maximize the value of the infrastructure. To this end, this Article proposes a model rule, “neighbor billing,” to reduce the effects of the current potential market failures without creating new ones. This rule pushes pricing decisions into the network core, which allows internal market forces to allocate costs as appropriate, without need for external regulation. Though not a perfect rule, this proposal suggests considerations of importance to the ongoing network neutrality debate.

© 2006 Virginia Journal of Law & Technology Association, at <http://www.vjolt.net>. Use paragraph numbers for pinpoint citations.

[†] JD Candidate 2007, University of Chicago Law School. Thanks are due to many for their help and encouragement in bringing this Article to print. Of particular note are Marsha Nagorsky, Doug Lichtman, and Lisa Bernstein, all of the University of Chicago Law School, and Debbie Minehart of the United States Department of Justice Antitrust Division Economic Analysis Group. All errors are, of course, my own.

TABLE OF CONTENTS

I. Introduction.....	2
II. The Network	4
A. Network Architecture and Theory	5
B. Network Neutrality	11
III. The Post-Neutrality World.....	15
A. Things Have Changed.....	15
B. If Not Competition Then the Law?.....	18
IV. Money: Still the Root of All Evil.....	21
A. Building The Infrastructure.....	21
B. Allocating Infrastructure Costs	23
C. Neighbor Billing	26
D. Does it Work?	29
V. Conclusion	32



I. INTRODUCTION

¶1 On November 7, 2005, AT&T CEO Ed Whitacre shook the Internet at its core. In an interview with BusinessWeek Online, he suggested—in none too friendly terms—that AT&T might undertake to charge content providers for access to its network:

Q: How concerned are you about Internet upstarts like Google . . . , MSN, Vonage, and others?

A: How do you think they're going to get to customers? Through a broadband pipe. Cable companies have them. We have them. Now what they would like to do is use my pipes free, but I ain't going to let them do that because we have spent this capital and we have to have a return on it. So there's going to have to be some mechanism for these people who use these pipes to pay for the portion they're using. Why should they be allowed to use my pipes?

The Internet can't be free in that sense, because we and the cable companies have made an investment and for a Google or Yahoo! . . . or Vonage or anybody to expect to use these pipes [for] free is nuts!¹

This statement has come at a volatile time. AT&T has just completed its merger with

1. Patricia O'Connell, *At SBC, It's All About "Scale and Scope,"* BUS. WK. ONLINE, Nov. 7, 2005, available at http://www.businessweek.com/magazine/content/05_45/b3958092.htm?campaign_id=search (reciting an interview with AT&T CEO Ed Whitacre). Duane Ackerman, the CEO of BellSouth, has made similar comments. See Jonathan Krim, *Executive Wants to Charge for Web Speed*, WASHINGTON POST, Dec. 1, 2005, at D05) available at <http://www.washingtonpost.com/wp-dyn/content/article/2005/11/30/AR2005113002109.html>.

SBC² and its merger with BellSouth has received final regulatory approval³—giving it substantial power in the retail, wholesale, and Internet backbone markets. At the same time, Congress is undertaking the largest revision of the telecommunications laws since the Telecommunications Act of 1996—a revision meant to supplant the massive 1996 effort. And the Supreme Court’s summer 2005 decision in *Brand X*⁴ cleared the way for the FCC to reclassify cable and DSL broadband connections as “information services” (as opposed to “telecommunications services”), thereby exempting them from common carrier regulations.⁵

¶2 Concern over Whitacre’s statement has erupted into a legislative battle over so-called “network neutrality.” Network neutrality, a term attributed to Tim Wu, is an amorphous concept.⁶ Generally, a “neutral” network is one that does not favor one application over others.⁷ Whitacre’s comments clearly fit into the network neutrality rubric: if AT&T can charge companies for access to its network, it can favor applications whose providers pay for access over those who do not. More dramatically, if the law does not regulate the metrics that AT&T uses to set the rates that it charges companies, AT&T could decide to price Google off of its network in preference of Yahoo! (or even an AT&T-owned search option).

¶3 This Article asks whether we should let AT&T charge content providers for access to its customers. It argues that we should, but with structural limits on whom AT&T can charge. In specific, we should limit a network’s owner to charging only those who are directly connected to the network. I term this structure “neighbor billing.” This

2. This merger is complete pending completion of the ongoing Tunney Act review. See Stephen Labaton, *Quick Approval of Phone Deals Uncertain*, N.Y. TIMES, July 13, 2006, at C9.

3. The AT&T/BellSouth merger has been approved as this Article nears going to print. See Fed. Comm’n. Comm’n., *FCC Approves Merger of AT&T Inc. and BellSouth Corporation* (Dec. 29, 2006), available at http://hraunfoss.fcc.gov/edocs_public/attachmatch/DOC-269275A1.pdf. Importantly, this merger is conditioned upon adherence to certain network neutrality principles. See MERGER COMMITMENTS 9 (Fed. Comm’n. Comm’n. Dec. 28, 2006), available at http://www.fcc.gov/ATT_FINALMergerCommitments12-28.pdf. For some analysis of these provisions, see Tim Wu, *The AT&T Network Neutrality Agreement* (Dec. 29, 2006), available at The AT&T Network Neutrality Agreement. The agreed-to principles contain important exceptions. They do not apply to the AT&T core network; they do not apply to certain services, including IPTV (see sections III.A and IV.A); and they can be supplanted by any subsequent legislation that addresses network neutrality. For a general overview of these mergers, see Ken Belson, *The Bell Merger: The Overview*, N.Y. TIMES, Mar. 7, 2006, at C1.; Geraldine Fabrikant, *Those Bell Mergers Are Giving Cable Companies Even More to Worry About*, N.Y. TIMES, Mar. 13, 2006, at C1.

4. See *Nat’l Cable & Telecomm. Ass’n v. Brand X Internet Serv.,’s.*, 545 U.S. 967 125 S Ct 2688 (2005).

5. See 47 U.S.C. §§ 153(10), (20), (46) (2000) (defining “information services” and “telecommunications services”).

6. See generally Tim Wu, *Network Neutrality, Broadband Discrimination*, 2 J. TELECOMM. & HIGH TECH. L. 141 (2003). The present case is a demonstration of its amorphous nature: Whitacre’s statement was about favoring sources of data (e.g., Google, if it pays), not the applications using the data (e.g., video or e-mail). Of course, content and source are often proxies for one another. AT&T might charge Google a premium for video content, or it might charge everyone a premium for video content. The distinction is only one of granularity. The point to take away is that net neutrality is an amorphous concept under either perspective.

7. *Id.*, at 145.

simple rule forces network owners to internalize most of the negative externalities that any given pricing scheme might create, while at the same time giving them the flexibility to allocate costs as needed to fund further development of their networks. In particular, this rule operates to limit companies' abilities to exploit market failures by bypassing negotiation bottlenecks.

¶4 We need to do some work before we can unpack this conclusion. This Article starts in Part I with an overview of network theory and architecture and proceeds to a discussion of the basis of network neutrality. The architecture of modern networks generates substantial efficiencies by allowing dynamic allocation of resources. But this dynamic allocation of resources also creates neutrality concerns. We cannot impose neutrality without risking the efficiencies that make the Internet possible.

¶5 Part II looks at how pricing models affect the development and use of the Internet. Allocating costs has been the bane of the Internet business model. Two questions have defined the Internet age: how to make money by giving content away (e.g., Google), and how to facilitate the transfer of money when the content requires it (e.g., EBay or Paypal)? These are both content-level questions. The present discussion operates at an even lower level: how do we pay for the infrastructure over which the content flows? The same dynamic allocation of resources allowed by modern networks that raises neutrality concerns also makes it difficult to properly allocate infrastructure-related costs. I argue first that the structure of the Internet is changing in ways that necessitate more aggressive pricing models to properly match infrastructure costs to the uses of the infrastructure, and, second, that neighbor billing reduces the viability of pricing models that distort the proper allocation of costs for the construction and use of network infrastructures.

¶6 Part III considers how neighbor billing fits into the network neutrality discussion. The present policy debates have been animated by concerns over AT&T's proposed billing models—but they have been framed in the language of network neutrality. This is unsurprising: few people would be visibly affected if AT&T were to start charging content providers for access to its network, so long as the content providers paid up. But once some of the content providers stop paying—or AT&T starts selectively offering premium access—then consumers start to experience the Internet differently: Yahoo! is fast because it has paid, and Google is slow or non-existent because it has not. From the layperson's perspective, this is a neutrality problem, not a billing problem. Part III evaluates whether neighbor billing exacerbates or mitigates the network neutrality concerns raised by the need for more aggressive pricing models.

II. THE NETWORK

¶7 As far as most users are concerned, the Internet looks like Figure 1. Some more sophisticated users might put an Internet Service Provider ("ISP") between themselves and the "Internet." Some less sophisticated users might cross out the "Internet" label and replace it with the name of their ISP—or perhaps even with the name of companies like Google and Yahoo!. The basic point is that most users view the Internet as a black box to which their computers connect to get content off of computers on the other side of the Internet.

¶8 This Part looks inside the black box understanding of the Internet. It starts by discussing the packet switched network theory and inter-network architecture. It then applies these ideas to examine the practical meaning of network neutrality. This discussion is generally historical, setting the stage for Part II to argue that modern circumstances have changed, forcing us to address questions that have previously only been hypothetical.

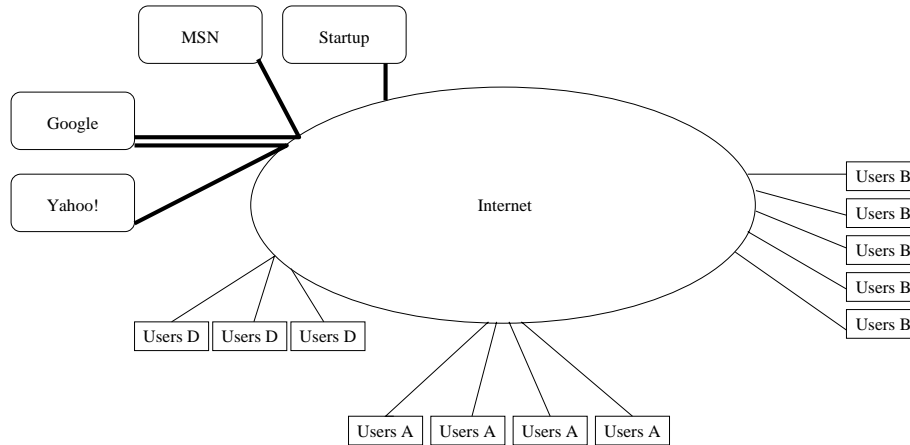


Figure 1: Common view of the Internet

A. Network Architecture and Theory

¶9 The modern network was born in the early 1960s.⁸ It was then that researchers—simultaneously on both sides of the Atlantic—developed the theory of packet switched networks (“PSNs”).⁹ PSNs are the building blocks of the Internet. But before we can discuss them, we need to take a step back and trace the evolution of networks that preceded them.

¶10 Before the advent of the PSN, networks were either static or circuit switched. In a static network, each node has a fixed communications pipe (“circuit”)¹⁰ between each other node with which it can communicate. An example is the “red phone” network used by the US government to coordinate the launch of nuclear missiles: each phone can only call a single destination. When you pick up the phone on one end, it closes a circuit—a

8. For the purposes of this Article, “network” is taken to mean communications network.

9. See generally JOHN NAUGHTON, A BRIEF HISTORY OF THE FUTURE: THE ORIGINS OF THE INTERNET chs. 6, 8 (Phoenix 1999); Paul Baran, *The Beginnings of Packet Switching: Some Underlying Concepts*, IEEE COMM. MAG., July 2002, at 42; Barry M. Leiner et al, *The Past and Future History of the Internet*, COMM. ACM, Feb. 1997, at 102; see also Leonard Kleinrock, *Information Flow in Large Communication Nets*, (May 31, 1961) (unpublished Ph.D. proposal, Mass. Inst. of Tech.), available at <http://www.lk.cs.ucla.edu/LK/Bib/REPORT/PhD/proposal.html>. The FCC’s relevant orders contain substantial developments of the relevant histories of this technology. See e.g., In re Appropriate Framework for Broadband Access to the Internet over Wireline Facilities, 20 F.C.C.R. 14853, ¶¶ 34–37 (Sept. 23, 2005), available at http://hraunfoss.fcc.gov/edocs_public/attachmatch/FCC-05-150A1.pdf.

10. These are called circuits because each connection consists of a loop of wire that can form a complete circuit. The caller on each end modulates a current across the circuit to the remote node.

loop of wire—that causes the phone on the other end to ring. These networks are by definition limited: connecting n people in a static network requires $(n*(n-1)/2)$ connections, which makes them expensive to create, and requires each person to have $n-1$ phones, making them difficult to manage. And each new connection costs more than the last, because connecting person n to the network requires making $n-1$ new connections. Figure 2 demonstrates the complexity of a static network.

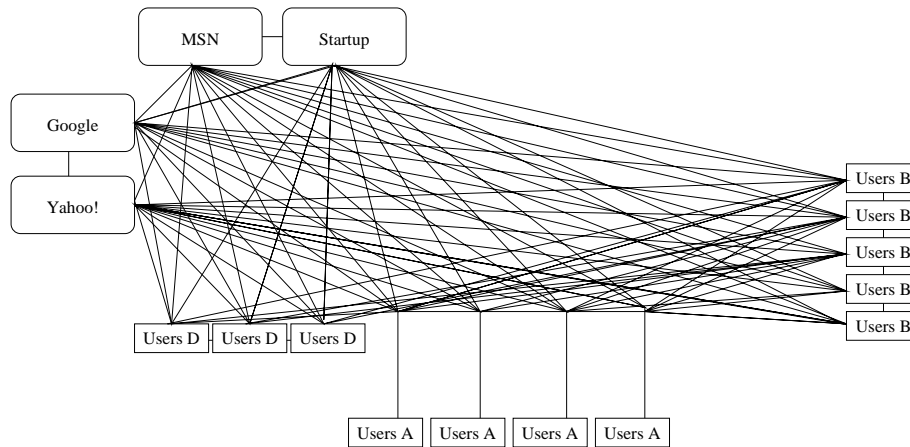


Figure 2: A static N -to- N network

¶ 11 Circuit-switched networks help to reduce the number of connections needed to connect a given number of people. They operate on the same principle as static networks: the creation of an electrical circuit between two nodes. But the theory adds an assumption that phone calls are *blocking*—that any given person will only ever use one phone at any given time. We therefore want each person to have a single circuit that can be switched between remote endpoints. The image to have in mind is that of the old switchboard. In such a network, every person has a phone line that runs to a central switchboard. When one person calls another, a temporary, but dedicated, circuit is established between the phones. To connect n people in a circuit switched network, we need n lines run to a central switchboard, and each person needs only one phone. Adding each new connection costs the same as adding each previous connection—however much it costs to run a single line from phone to the switchboard. We also need $n/2$ cables for connecting the calls. While the cost of a three inch cable is negligible relative to that of the several thousand feet of cable connecting the switchboard to the phones, we do need to keep track of this requirement—in other models the cost structure might be different, limiting the number of simultaneous calls that can be made. Even in the switchboard case, it might be physically impossible to connect every single caller at the same time. Figure 3 shows a circuit switched version of the network from Figure 2.

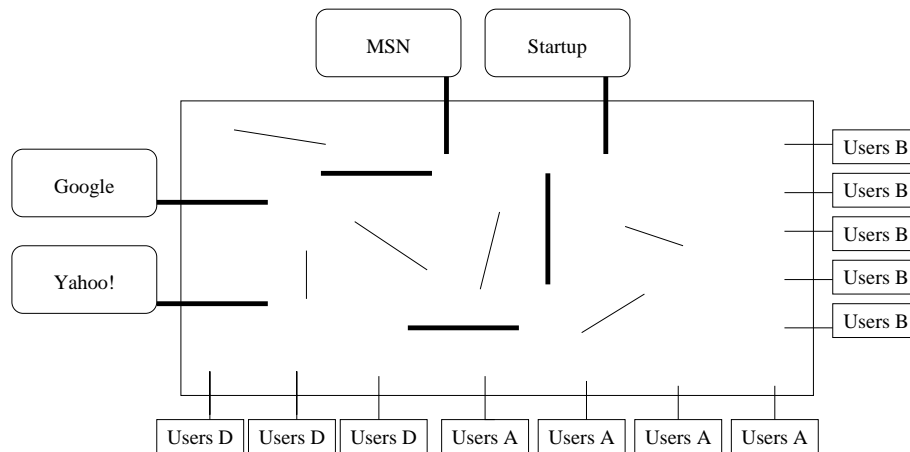


Figure 3: A circuit-switched N -to- N network

¶¹² The circuit-switched network is the building block of the modern telephone system. It is true that we no longer adhere strictly to the physical circuit and switchboard models, but the basic theory still holds.¹¹ The principal difference is that modern telephone calls lines are “multiplexed” into high-capacity trunk lines in a process known as Time Division Multiplexing (“TDM”).¹² A common trunk line is the T-1, which can carry up to twenty four voice lines at a time.¹³

¶¹³ TDM works by converting a voice conversation into digital form and then transmitting that digital representation in higher bandwidth chunks than that at which analog voice data travels. In particular, human voices can be captured by an eight kilohertz sample rate by taking eight thousand digital bytes of data every second. A TDM line, however, can transmit twenty-four times this much digital data per second. To take advantage of this, the phone network runs high-speed data lines between its central office locations—one TDM line for every twenty four voice lines—instead of running many low speed lines.

11. As described immediately below, circuits have largely been replaced by “virtual” circuits created over TDM networks. Switchboards have been replaced by Electronic Switching System Switches, which are capable of switching hundreds of thousands of circuits at a time.

12. See A HISTORY OF ENGINEERING AND SCIENCE IN THE BELL SYSTEM: TRANSMISSION TECHNOLOGY (1925-1975) 527-29 (E. F. O’Neill, ed., AT&T Bell Labs 1985). Explanations of the technologies discussed in this section can be found in HARRY NEWTON, NEWTON’S TELECOM DICTIONARY (20th ed. 2004). For quick reference, Wikipedia provides a reasonable account of most of this material as well. See generally Wikipedia, <http://www.wikipedia.com>.

13. This Article uses “TDM line” as a generic term. All of the examples are based on a specific type of TDM line, the T-1 line. A T-1 line can carry twenty four voice channels, or 1.544 megabits of data per second. There are other capacities of TDM lines. I use “TDM line” in the text both to reduce the number of technical terms, and to emphasize that these lines work through Time Division Multiplexing. The literature on the development of the phone networks generally focuses on the transition to “Pulse Code Modulation” (“PCM”)—which enabled digital to analog transmission. From a technical perspective, this was a more fundamental shift in technologies. But the greater purpose of this transition was that it permitted time division multiplexing, along with improving the quality of signals transmitted over long distances. See O’NEILL, *supra* note 12.

¶ 14 The process of merging multiple voice lines into a single high-capacity data line is called multiplexing. Voice lines are connected to a “TDM switch”—in telecom parlance this is called a “Class 5 switch.” A TDM switch divides the high-capacity line along a temporal axis into twenty-four “timeslots,” each of which is allocated one byte of data every 1/8000th of a second. When a voice line is taken off of the hook, the TDM switch assigns it to an available timeslot. If all twenty-four are in use by other phone calls, it gives the new caller a “no available circuits, please try again later” message. Once the line has been assigned a timeslot, the TDM switch starts converting any sounds coming over it into a digital form. This digital data is put onto the high-speed line in its appropriate outgoing timeslot and sent to a timeslot on another TDM switch. This remote switch “demultiplexes” the incoming timeslot, converts the digital data back to sound, and puts the sound on a remote voice line.

¶ 15 It is not particularly intuitive to think about sound in terms of frames. A more intuitive analogy can be found with sight—the human eyes are commonly said to see at about thirty frames per second. We therefore are unable to see any movements that last for less than 1/30th of a second. Conceptually, if one were to interlace the frames of two film reels and play the combined reel back at double speed you would only see one of the two films—which one you saw would depend upon whether your eyes were synchronized with the even- or odd-numbered frames. A TDM switch is doing this same thing, with twenty four “reels” of sound at eight thousand frames per second.

¶ 16 There are two things to take away in the case of either sound or video. First, TDM divides a high-capacity media along a temporal axis into some number of timeslots, and each of these timeslots is given fixed-duration bursts of access to the high-capacity media at a regular interval. If you know this interval, data multiplexed onto the high-capacity media at one end can be demultiplexed off of it at the other, such that there appears to be a direct connection between the ends and without any hint that there are other streams of data sharing the high-capacity line. Because it appears that a TDM creates a direct circuit between endpoints, a TDM switch is said to create a *virtual* circuit. Second, an inherent limitation to TDM is that it can only carry however many of these virtual channels as it has timeslots to allocate. When a twenty-fifth person tries to make a phone call, he will be told to try again later.

¶ 17 The TDM system was developed by AT&T in the 1960s—at the same time computer researchers were developing the first computer networks.¹⁴ The earliest models for modern computer networks took the ideas of circuit switching and multiplexing to their limits.¹⁵ When I make a phone call, a virtual circuit is established during the entire phone call, no matter whether I am actually saying anything. If I make a phone call and leave the phone off of the hook, I am wasting 1/24th of the TDM line’s capacity. The system is setup like this for two reasons: first, humans usually only make one phone call at a time and, second, silence is part of their conversations. Computer networks were developed for a very different reason: they were developed in the era of mainframes as a

14. See O’NEILL, *supra* note 12, at 527, 560–62; NAUGHTON, *supra* note 9; Frederick T. Andrews, *The Telephone Network of the 1960s*, IEEE COMM. MAG., July 2002, at 49-51.

15. NAUGHTON, *supra* note 9, at 102, 123–28; Leiner, *supra* note 9.

way for a large number of people to access a small number of computers at once without needing to keep a connection open when data is not being sent. Because the connection does not need to be kept open, there is no need to establish a circuit for each session. This in turn allows multiple sessions to be run over a single circuit, reducing the number of circuits needed to allow many people to connect to a single computer. To exploit these differences, a non-circuit-based means of transferring data was created: the Packet-Switched Network.

¶ 18 There is no circuit in a PSN. Data is broken up into packets and sent over the network. A TDM line sends eight thousand frames of data every second, with 1/24th of each frame assigned to a given timeslot. This allows up to twenty-four virtual circuits each to transfer eight thousand bytes of data per second. A PSN can multiplex data onto the same high-capacity line. But instead of assigning a timeslot to each connection and packing multiple timeslots into each frame, a PSN fills entire groups of frames with data from a single sender. Each group of frames—called a packet—contains a chunk of data, akin to the address on an envelope, which tells the other end of the connection how to distribute the data in that frame. A PSN therefore can send eight thousand frames per second and allocate them in any combination between any number of connections.¹⁶ It could send eight thousand frames from a single source to a single destination; eight thousand frames from a single source to eight thousand different destinations; eight thousand frames from different sources to a single destination; or eight thousand frames from different sources to different destinations.

¶ 19 Each packet is like a letter: it is an envelope that contains data, and it has an address written on the outside. When you put the packet on the network, the network delivers it to that address, and the computer at that address figures out what to do with the data. As another example, in a PSN the phone line is like a party line, where many people can speak at the same time. Just because there are many people on the line does not mean that two people cannot have their own conversation: before speaking, each person can say the name of the person to whom they are speaking, followed by his own name, and the other party only listens to sentences that start with his name. It might be slower, and less private, than if the parties had a dedicated line, but a conversation can nonetheless be held.

¶ 20 Based upon this description, PSNs do not sound promising: they are slow and lack privacy. But PSNs have a huge advantage over circuit switched networks because they allow multiple concurrent conversations with a single call. This difference can be stated semantically: with a circuit switched network, node *A* connects to node *B* over the network; with a PSN, nodes *A* and *B* can communicate because they are both connected to the network. But because the connection is to the network and not to each other, the nodes are not limited to speaking to one another. Because it more efficiently allocates resources, the PSN is closer to the static network in terms of communication capacity, but like the circuit switched network in terms of cost.

16. For ease of exposition, this assumes that the PSN in question uses the same line configuration as a TDM line. There is no technical requirement that this be the case. Most networks are configured to use packets much larger than the 24 bytes allocated to a single TDM frame. For instance, the Internet's dominant protocol TCP/IP typically uses a maximum packet size of 1500, or even 9000, bytes of data.

¶21 An immediate objection should be that a PSN is inefficient in terms of capacity use. Thinking of the party line example, any conversation in which every sentence started, “Alice, this is Bill,” would be slow. And this assumes that the names were exchanged with every sentence. What if they were exchanged with every word? This is a valid criticism, but it only demonstrates that there are efficiency tradeoffs between a circuit switched and a packet switched network. If two people know that they only want to talk to one another, they might not benefit from a PSN. Similarly, if the addresses that go on the packets are long relative to the data, or if there are lots of people on the party line, the PSN will be less efficient than a direct call.

¶22 These are all important concerns, and they play a central role in the discussion of network neutrality. But right now it is more important that we say why PSNs are good. And they are good. The theory has been empirically proven sound, as nearly all computer networks are PSNs, many of the modern telephone network backbones are PSNs,¹⁷ and the telecommunications industry writ large is largely converting to a pure PSN-based architecture.¹⁸

¶23 Why this success? As a practical matter, because the overhead of the PSN—the size of the headers relative to the data and the complexity of switching the connection for each packet instead of for each circuit—is reasonably small. But this is a threshold condition: there would be no PSN if this condition was not met, but satisfying this condition alone does not explain why PSNs are successful.

¶24 There are three main reasons for the success of PSNs. First, like a TDM line, they allow for shared connections. A single line can carry many conversations (or data streams) at once. This reduces the need for new resources (no need to run a new line to a new node if that node can be multiplexed onto an existing line). It also allows for better use of existing resources, because a single inefficient user cannot necessarily block more efficient users.

¶25 Second, and unique to PSNs, they allow for more dynamic sharing and allocation of resources. For instance, imagine someone sending data from Florida to New Mexico, and assume that the most direct path for that data is through a network in Texas. Halfway through the transfer someone in Texas needs to use the network to make a local call. In a circuit switched network, he would get a busy signal (or some other signal saying there were not enough free circuits). In a PSN, we can reroute the transfer from Florida to New

17. We should note that a TDM circuit is not a PSN, because each call placed over a TDM circuit establishes a persistent virtual circuit. That is, when the call is placed, the header information is exchanged only once. The call is then transferred in pre-allocated chunks until the parties hang up and the circuit is released. A PSN requires that each chunk of data be a packet with its own set of headers.

18. See, e.g., Petition to Deny of Time Warner Telecom, at 2, In re AT&T Inc. and BellSouth Corporation Application for Approval of Transfer of Control, No. FCC-06-74 (FCC June 5, 2006) (summarizing state of telecommunications market IP convergence), available at http://gullfoss2.fcc.gov/prod/ecfs/retrieve.cgi?native_or_pdf=pdf&id_document=6518358787; Denise Pappalardo, *AT&T outlines migration to IP core*, NETWORKWORLD, Sept. 15, 2003, available at <http://www.networkworld.com/news/2003/0915att.html>; Gillian Law, *Sprint Plans Move to Packet-switch Network: The Company Envisions a 30% Drop in Operating Costs*, COMPUTERWORLD, May 27, 2003, available at <http://www.computerworld.com/mobiletopics/mobile/story/0,10801,81552,00.html>.

Mexico through an Oklahoma-based network—and we can do this without disrupting the connection. Similarly, if the network in Texas were to crash, the users that had been using it could be rerouted.

¶26 This highlights a central difference between PSN-style multiplexing and traditional TDM.¹⁹ You can transfer data over a TDM connection, but this requires allotting timeslots for that purpose—timeslots which remain allocated for the duration of the transfer, regardless of other needs. A common TDM configuration used by businesses might allocate twelve timeslots (768 kbps) for data and twelve for phone lines. This is inherently inefficient: whenever any one of the phone lines is not being used, its timeslot could be transferring data. Conversely, if all twelve phone lines are in use and the business's CEO wants to call home to his wife, he probably would like to slow down the internet connection a little bit to make his call. TDM does not allow for this dynamic allocation, but it is built in to PSN.

¶27 Third, and most unique to PSNs, packetized data not only allows concurrent transfers, but it allows diverse transfers, over diverse media. A PSN can be used to transfer data, voice, or video communications (or anything else that can be digitized) over any media capable of sending digital data (for instance, TDM, cable, or wireless).²⁰ While some other networks can carry different types of data, they do so by splitting the connection into separate logical connections (like the twelve timeslots for voice and twelve for data in the example above). The PSN puts all data—no matter the type—onto the same network.

¶28 At a theoretical level, the basis for these above points is that PSNs can be implemented on top of any other network, and can simultaneously replicate the functioning of any number of other networks, in exchange for the cost of imposing some amount of overhead. But PSNs introduce a new question: when sharing a scarce resource, who gets access to it when there is congestion? PSNs only tell us how to packetize data at point *A* and de-packetize it at point *B*. There is nothing inherent in their design that tells us how to deal with congestion. Historically this has not been a substantial problem because the data sent over PSNs has not been time-sensitive. But the nature of the data being sent over the Internet is changing, requiring us to rethink how we deal with congested networks. This is the genesis of the network neutrality debate.

B. Network Neutrality

¶29 The Internet is not neutral, nor should it be. This section argues that the network

19. Some modern TDM systems have developed a finer granularity than simple circuit-based connections; this trend mirrors the industry's general convergence towards PSN-based networks. *See, e.g., id.*

20. This limitation is very broad, as “digital data” only requires being able to send a minimum of two states—binary data. Common PSN media includes telephone networks (telephone, T-1, and DSL lines), cable networks, satellite and other wireless networks, fiber optic links, ethernet, and even electricity transmission lines. More esoteric PSNs have been created, or hypothesized, as running over barbed wire fences, smoke signals, carrier pigeons, sound waves, and even shockwaves through deep-earth drilling equipment.

physically cannot be neutral, based upon the mechanics of PSNs, and that it should not be, based upon our understanding of why PSNs are good.²¹

¶30 The previous section discussed the mechanics of PSNs. They take digital data, cut it up, send it over an arbitrary network, and glue it together on the other side. The “arbitrary network” in the middle is still a black box. Indeed, most protocols—the rules that govern how data is cut up and glued back together—assume that the network is a black box. But no matter what the network looks like on the inside, it is going to have a few common characteristics: there will be capacity bottlenecks, and it will take time for data to cross it. This first constraint is called bandwidth, and the second is called latency.

¶31 The first and more familiar constraint is bandwidth. As a practical matter, it is impossible to engineer a network between more than two nodes that is not bandwidth constrained. If a network connects more than two nodes it is necessarily the case that at least one of the nodes is connected to at least two others (Figure 4). Given that each node²² and each link has its own maximum speed, if this central node (*B*) can keep up with both other nodes (*A* and *C*), its bandwidth must be at least equal to that of $A+C$; but that means it is faster than both *A* and *C*, so there are bottlenecks between it and both *A* and *C*. And if the central node is not at least as fast as $A+C$, then there is by definition a bottleneck when both *A* and *C* are talking to *B*.

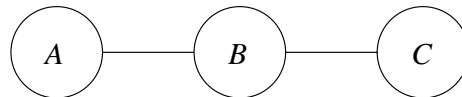


Figure 4: A simple network of more than 2 nodes

¶32 Latency is a less intuitive constraint. There is no question that latency exists. Even a dedicated circuit experiences latency, as anyone who has ever spoken on a cross country or international call might know. The most obvious source of latency is the speed of light. Traveling at 186,000 miles per second, a piece of data sent 3000 miles across the country takes 0.016 seconds to go from mouthpiece to receiver.

¶33 But most latency does not come from this physical propagation delay. Just like cars at an intersection, packets need to wait in line to get onto and off of each section of the road. These intersections are controlled by routers or switches (collectively “routers”). Routers take incoming packets off of one connection, store them in a

21. This section draws from a large body of non-legal literature. Useful overviews of various queuing algorithms, described below, are presented in Chuck Semeria, *Supporting Differentiated Service Classes: Queue Scheduling Disciplines*, Juniper Networks White Paper (Dec. 2001), available at http://www.juniper.net/solutions/literature/white_papers/200020.pdf. See John Nagle, *On Packet Switches With Infinite Storage*, RFC-970 (Dec. 1985), available at <http://www.ietf.org/rfc/rfc0970.txt>, for an example of the earliest literature and CHRISTIAN HUITEMA, *ROUTING IN THE INTERNET* (Prentice Hall 2000)), for more recent treatment. Interested readers should also look at Sally Floyd and Van Jacobson, *Random Early Detection Gateways for Congestion Avoidance*, 1 *IEEE/ACM Transactions on Networking* 397 (1993), available at <http://www.aciri.org/floyd/papers/early.twocolumn.pdf>, to see a modern development of how packetization-based and router-based congestion algorithms relate.

22. “Node” in this context is not limited to end-nodes (e.g., *A* and *B* in Figure 4), but also includes network core elements like switches and routers.

temporary queue, figure out which networks they need to be sent away on, wait for open spots in those networks, and then send them along on their way. As with real cars and real intersections, this is all very easy until cars start backing up. There is an important conceptual difference between intersections and routers. When an intersection gets backed up, the line of cars leading up to it keeps getting longer, extending further and further onto the road. Packets are not like cars: they are bursts of electricity or light. When a packet gets to the router, it can no more wait for a free space on the router than a lightning bolt can stop mid-strike. Packets that do not fit into the router's queue are quietly "dropped," or ignored—it is up to the sending parties to realize that the packet was dropped and to try again.

¶34 Routers use one of any number of "queuing algorithms"²³ to control these backups. The earliest queuing algorithms were simple First-In-First-Out ("FIFO")²⁴ queues.²⁵ A FIFO queue is like a one-lane road at a stop light: cars go through the light in the same order that they arrive. This seems like a fair way to queue packets—and it might be if the router had an infinite amount of queue space.²⁶ But it does not. Assume that the router has a four unit queue size, and that it can send four units per second. Next assume that there is a constant stream of small (one unit) packets coming in at a rate of four packets per second. Finally, assume that there is a second stream of large (three units) packets coming in at one packet per second. The total input is seven units per second, which is greater than the output of four units per second, so the queue is going to fill up. Because packets can only enter the queue when there is enough space for them, once the queue is filled the second stream will never be able to enter the queue because the first stream is keeping it full. Obviously a network built around a FIFO queuing algorithm is not neutral—the small stream gets perfect transmission, and the large one gets nothing.

¶35 We reach a different conclusion if we change the parameters slightly. Assume the same router and the same two streams, but each of the streams sends packets out at random intervals instead of at constant rates. The queue was four units large, and the small packet stream was sending one unit packets—each packet could therefore be in line behind zero, one, two, or three other packets, giving it an average latency of the time that it takes to process of 1.5 packets through the queue. But the second stream's packets only fit in the queue if there are zero or one (single-unit) packets in front of it. On average, therefore, the large packets need to spend one third the amount of time waiting in the router's queue as the small packets.²⁷

¶36 This shows that even the simplest of algorithms simultaneously gives both detrimental and preferential treatment to different types of data based upon exogenous

23. These are also called queuing disciplines.

24. See Nagle, *supra* note 21, at 3 (discussion of FIFO queuing algorithms), RFC 970 for an early discussion of FIFO queuing algorithms (cited in note 21).

25. Indeed, the majority of routers in use today use FIFO queuing—although Internet core routers are more likely to implement more advanced algorithms.

26 See Nagle, *supra* note 21, at 3.

27. These are rough calculations that don't consider the distribution of packets in the queue, or other similar factors. In particular, I don't consider the latency incurred when a large packet gets dropped due to insufficient queue space. But the general point holds. Many non-FIFO algorithms create these types of latency distortions.

circumstances. There are many other queuing algorithms, each of which has its own peculiar effects on different types of data.²⁸ The point to take away is that every transportation medium can be characterized in terms of bandwidth and latency, and managing how data flows between different media requires making tradeoffs between the two.

¶ 37 Wherever rule-based decisions are made there is bias: this is as true with networks as with everything else. Just imagine a hypothetical perfect routing algorithm that has two packets arrive simultaneously but only has space for one of them. Which does it choose? The bias is clear if it chooses one systematically over the other. If it chooses one at random, it is biased against data flows that need low latency (because the packet will need to be resent). Even if it drops both packets it shows biases—latency-sensitive packets, regular-interval packets, and small packets are all disproportionately harmed by such a rule.

¶ 38 Most discussions about network neutrality make exceptions for biases that are incidental to running a network. So long as there is a sound technical reason for configuring the network in a given way it can be called neutral. This, of course, is an exception that begs abuse.

¶ 39 Network neutrality is at best an ideal. Given the choice between two otherwise identical networks, the more neutral of the two is preferable. Here “more neutral” means less prone to distorting traffic. Given a network as in Figure 5, in which nodes *A* and *B* are sending data to nodes *E* and *F*, respectively, we might say that the network is neutral if the connection between nodes *C* and *D* does not distort traffic between *A* and *E* and traffic between *B* and *F*. That is, if *A* is talking to *E* and *B* is talking to *F*, the network is considered neutral if neither *E* nor *F* can tell that any other node is there.

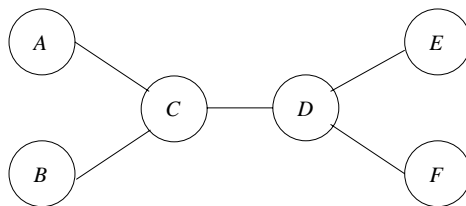


Figure 5: A network with a connection shared between multiple nodes

¶ 40 But this ignores the bandwidth and latency limitations inherent in networks. There is little doubt that we would not call the network neutral if it needlessly preferred traffic from *A* over that from *B*. But “needlessly” is a huge qualifier. If *A* is a surgeon performing surgery over the network, and *B* is a teenager sending pirated Britney Spears videos, there is legitimate need to prefer *A*’s traffic.

¶ 41 This returns us to a core feature of PSNs: they allow for efficient and dynamic allocation of underused resources. Assume that performing remote surgery requires a great deal of bandwidth, but it is a facility that is only used once or twice a month. A

28. See Semeria, *supra* note 21 (discussion and overview of several common algorithms). See generally Nagle, *supra* note 21; Huitema, *supra* note 21; Floyd & Van Jacobson, *supra* note 21.

hospital could purchase a dedicated circuit—but this is very expensive, particularly given how rarely it will be used. A PSN allows the hospital to buy high speed network access, and to share that access when it is not being used. Perhaps the hospital's general internet traffic could share this connection, reducing the hospital's general internet costs. Or the hospital's ISP could share unused bandwidth with other customers—the teenager sending Britney Spears videos—which in turn reduces the price to the hospital.

¶42 PSNs are useful because they allow the sharing of resources in this way. Without packetization, separate networks are required for each type of application. With packetization, each node has access to each other node's unused capacity, reducing the size of the infrastructure that needs to be built.

¶43 It is necessary to remember, however, who built the network. When a doctor needs to perform telesurgery, the PSN needs to be able to give him sole access to his bandwidth. The hospital is not subsidizing the teenager's internet access; it is allowing the ISP to sell its unused capacity to the teenager, in exchange for reduced costs—the teenager is subsidizing the hospital. An effective PSN needs to offer some level of minimum service guarantees. Otherwise the hospital won't be willing to share its bandwidth.

¶44 Historically, the Internet has not offered these guarantees—there has been no need, because it has historically been used to transfer exclusively non-time sensitive data. There is no harm in delaying the transfer of an e-mail or database, or even a web page, by a couple of seconds. But users will not tolerate even a brief pause in a streaming video or a telephone call. Those few institutions needing minimum service guarantees, like hospitals experimenting with telesurgery, have built their own private networks.

¶45 But the Internet is changing. Telephone companies used to have separate circuit-switched voice and packet-switched data networks. Over the past several years they have converged these into unified PSNs for both voice and data. They are now starting to offer packetized television programming. Cable companies are following this same trend: having set aside some of their cable bandwidth for packetized data, they are now moving their television programming to PSNs and are offering packetized telephone service. Perhaps the greatest change is coming from the business world. Businesses that used to buy thousands of phone and data lines are now consolidating all of these facilities onto single high speed data lines. All of these new services require some minimum level of guaranteed service. When Ed Whitacre, CEO of AT&T, picks up his phone to talk to Duane Ackerman, CEO of BellSouth, he does not want to be disconnected because some teenager is downloading a Britney Spears video—and he's not going to let his networks converge onto a single PSN if he cannot prevent this.

III. THE POST-NEUTRALITY WORLD

A. Things Have Changed

¶46 Ed is right. PSNs are great for maximizing the value of underutilized resources.

But many networks have excess capacity that cannot be shared so long as the Internet cannot “unshare” this capacity when the people paying for it need it. If we want to reap all of the benefits that PSNs have to offer, we need a way to guarantee that these networks only share their *excess* capacity.

¶47 Of course, Ed Whitacre is not talking about ensuring the minimum levels of service on other people’s networks. AT&T has one of the largest network infrastructures in the world. And they are in the process of deploying their IPTV service—television over the Internet—to run over their network. Most Americans’ high-speed Internet connections top out in the range of 1.5 to 4.5 mbps. Stable IPTV service requires speeds of between 6 and 18 mbps—and AT&T (like Verizon and BellSouth, AT&T’s main competitors in this area) is in the process of a multi-billion dollar upgrade of its network to support these speeds.²⁹ This vision for the PSN-based future simultaneously moves more users to PSN networks and increases the bandwidth allocated to each by at least an order of magnitude. Unsurprisingly, then, this upgrade requires improving the speed of core network facilities, in addition to the bandwidth at the edges of the network.

¶48 The question facing AT&T is how to pay for these upgrades.³⁰ On its own this question is innocuous. AT&T should be able to sell its excess capacity; and it should be able to preserve minimum guaranteed levels of service to ensure that only its excess capacity is sold. As discussed in Part I of this Article, network neutrality should not preclude service providers from offering minimum service guarantees.

¶49 The problem is that the same technologies that allow AT&T to provide these guarantees so that we can more efficiently allocate resources also allow AT&T to discriminatorily allocate the same resources. This sort of problem is not unique to networks—it is a problem with any toll or tax system. The same taxes that can be used to ensure a stable supply of food to hospitals, or that toll roads do not become overcrowded, can also be used to keep the king’s tables full while the peasants starve, or his cars out of traffic while the peasants sit in gridlock. This type of discrimination is orders of magnitude more powerful than that caused by routers’ different queuing algorithms. Rather than being a (potentially misused) side effect, the very purpose of this discrimination is to distort traffic.

¶50 Prior to 2005 the extent to which companies could distort the flow of traffic over their networks was limited. The major Internet backbone providers were all

29. These numbers can be found in any number of sources covering the industry. See e.g., Stephen Hardy, *FTH Conference Illustrates Applications' Progress; Fiber to the Home*, LIGHTWAVE, Nov. 1, 2006, at 1(3); Michael Finneran, *Three Architectures for The Telco Triple-Play: The RBOCs Must Modernize Their Infrastructure to Deliver Voice, Video and Data*, BUS. COMM’N. REV., Mar. 1 2005, at 32.

30. Consumers also want to answer this question. We want AT&T to make these upgrades to its network, because they create a great deal of value. AT&T, however, will not spend money on these upgrades unless it knows that it will be able to recover its costs. See Adam Thierer, *Are “Dumb Pipe” Mandates Smart Public Policy? Vertical Integration, Net Neutrality, And the Network Layers Model*, 3 J. TELECOM. & HIGH TECH. L. 275, 288 (2005) (“If a Net Neutrality/dumb pipe mandate is put in place, carriers might struggle to find ways to recoup their significant fixed costs of doing business and be discouraged from further innovating.”).

telecommunications companies; and their data connections were classified by the FCC as “telecommunications services.” Telecommunications services are subject to common carrier requirements: a set of rules requiring that providers offer service at equal prices, terms, and conditions to all new and existing customers.³¹ In June of 2005, the Supreme Court decided *National Cable & Telecommunications Ass’n v. Brand-X Internet Services*, 545 U.S. 967 (2005). This decision upheld the FCC’s previous reclassification of cable-based Internet services as “information services”—a classification that does not impose common carrier requirements.³² This decision paved the way for the FCC to similarly reclassify traditional telecommunications systems as information services—a change that was made in September 2005.³³

¶ 51 The FCC’s decision recognized that the transition from circuit-switched to packet-switched networks has led to a convergence of network facilities.³⁴ The FCC based its decision to eliminate the common carrier requirements specifically in their effects on telecommunications companies’ abilities to deploy new services (like IPTV):

[T]he inability to customize broadband service offerings inherent in the [common carrier] nondiscriminatory access requirement impedes deployment of innovative wireline broadband services taking into account technological advances and consumer demand. Thus, continuing to impose such requirements would only perpetuate wireline broadband Internet access providers’ inability to make better use of the latest integrated broadband equipment and would deprive consumers of more efficient and

31. *See* In the Matter of Policy and Rules Concerning Rates for Competitive Common Carrier Services and Facilities Authorizations Therefore, 77 F.C.C.2d 308, 312, 315-320 (Notice of Inquiry and Proposed Rulemaking Nov. 2, 1979).

32. *National Cable & Telecommunications Ass’n v. Brand X Internet Services*, 545 U.S. 967 (2005).

33. *See* Appropriate Framework for Broadband Access to the Internet over Wireline Facilities, 20 F.C.C.R. 14853, 14913 (Proposed Order Sept. 23, 2005) (This order generally exempted high speed internet services from classification under Title II of the Telecommunication Act—the elimination of common carrier requirements was one aspect of this change.).

34. In its earliest form, packet switching technology had limited uses Transmission speeds, of course, were extremely slow.

Digital technology and its applications have come a long way since the introduction of packet switching

. . . Packet-based technology is now deployed throughout wireline networks and is used in many circumstances

Wireline networks are now using digital, packet-based technology to deliver a wider range of services. Many of these services are IP-based....

. . . [T]he technology used to build networks, and the purposes for which they are built, are fundamentally changing, and will likely continue to do so for the foreseeable future. A wide variety of IP-based services can be provided regardless of the nature of the broadband platform used to connect the consumer and the ISP. Network platforms therefore will be multi-purpose in nature and more application-based, rather than existing for a single, unitary, technologically specific purpose. More generally, the erosion of barriers between various networks and the limitations inherent in those barriers will lead to greater capacity for innovation to offer new services and products.

. . . [W]ireline broadband Internet access service providers are no longer subject to the requirement [to offer] transmission ... service . . . on a common carrier basis.

Id.

innovative enhanced services. Similarly, a continued obligation to provide any new broadband transmission capability to all ISPs indiscriminately ... would reduce incentives to develop innovative wireline broadband capabilities.³⁵

¶ 52 Whitacre's famous statement that "what they would like to do is use my pipes free, but I ain't going to let them do that" was made only a few weeks after the FCC released these revised rules. Surely these events are related. And just as surely, the FCC was aware that these worms were inside when it opened the can.

B. If Not Competition Then the Law?

¶ 53 The FCC decided to exempt high speed Internet services from common carrier requirements on the grounds that there was a competitive marketplace that would regulate any abuses. They highlight intermodal sources of competition—cable, wireless, satellite, and broadband-over-powerline are all mentioned as competitors to traditional telecommunications service providers³⁶—and intramodal competition—between incumbent local phone companies and the competing exchange services created by the 1996 Telecommunications Act, for instance³⁷—to support the proposition that "[a]s any provider increases its market share or upgrades its broadband Internet access service, other providers are likely to mount competitive challenges, which likely will lead to wider deployment of broadband Internet access service, more choices, and better terms."³⁸ Their thinking is clear: if competition will improve quality and terms of service, we do not need common carrier regulations to regulate minimum terms.

¶ 54 But there is an apparent contradiction in the Commission's reasoning. The Commission squarely states that consumer demand for Internet service is driven by applications and content.³⁹ Yet the basis for abrogating common carrier requirements was competition between access providers, not content providers. The access providers therefore do not compete in terms of the quality of their access—but on the quality of the

35. *Id.* at 14905.

36. *Id.* at 1480.

37. We anticipate that, as the availability of cable modem and DSL broadband Internet access services grows with the modernization of network infrastructure and increased service deployment, more households will have the option of choosing between the cable and DSL broadband options. Increased intermodal and intramodal competition will continue to encourage these two broadband providers to deploy broadband Internet access services throughout their respective service areas. In addition, the threat of competition from other forms of broadband Internet access, whether satellite, fixed or mobile wireless, or a yet-to-be-realized alternative, will further stimulate deployment of broadband infrastructure, including more advanced infrastructure such as fiber to the home.

Id. at 14884.

38. *Id.* at 14485.

39. As the Internet and related applications mature and continue to evolve, the demand for broadband Internet access services will likely grow. The presence of more content available through the Internet and the enhanced means of presenting the content, together with growth in broadband-related applications, such as streaming video, will lead more subscribers to seek broadband Internet access service.

Id.

content provided. An analogous argument in a related area of the law is advertiser-sponsored television: the incentive is for television stations to show the programs that draw the largest audiences, not to show the best programs. In the Internet access context, this leaves open the possibility that competition will most reward the access providers that provide optimized access to the top ten or so Internet destinations, but that do so at the expense of all other sites. Indeed, the danger might well be even more pronounced where access providers also become content providers—here companies like AT&T have an incentive to give their own services preferential treatment over those of their competitors and to cross-subsidize any resulting losses.

¶ 55 These are the common concerns that are raised about network neutrality. While they have only recently exploded onto the front pages of newspapers and into the halls of Congress, they are not entirely new concerns. The FCC is acutely aware of them. In its 2005 Order, the FCC expressly recognized the concern:

Some commenters request that we impose certain content-related requirements on wireline broadband Internet access service providers that would prohibit them from blocking or otherwise denying access to any lawful Internet content, applications, or services a consumer wishes to access. While we agree that actively interfering with consumer access to any lawful Internet information, products, or services would be inconsistent with the statutory goals of encouraging broadband deployment and preserving and promoting the open and interconnected nature of the public Internet, we do not find sufficient evidence in the record before us that such interference by facilities-based wireline broadband Internet access service providers or others is currently occurring. Nonetheless, we articulate principles recognizing the importance of consumer choice and competition in regard to accessing and using the Internet: the *Internet Policy Statement* that we adopt today adopts such principles. We intend to incorporate these principles into our ongoing policymaking activities. Should we see evidence that providers of telecommunications for Internet access or IP-enabled services are violating these principles, we will not hesitate to take action to address that conduct.⁴⁰

It remains to be seen whether the FCC will step in should AT&T or other Internet providers begin to engage in discriminatory practices. Many commentators are skeptical that it will. At the same time, its September 23, 2005 policy statement⁴¹ is a poignant reminder that the Commission does retain the power to regulate Information Services—even to the extent that it can impose the same common carrier requirements that apply to Telecommunications Services on them. At a minimum, this suggests that there is no need for Congress to preemptively impose neutrality regulations. This is, in fact, the view that is apparently winning out in Congress. On June 8, the House of Representatives passed

40. *Id.* at 14904.

41. Appropriate Framework for Broadband Access to the Internet over Wireline Facilities, 20 FCC.C.C.R. 14986, 14987-88 (Policy Statement Sept. 23, 2005).

the Communications Opportunity, Promotion, and Enhancement Act of 2006⁴² (“COPE Act”). This Act—eventually defeated in the Senate—contains no specific network neutrality language, although it explicitly preserves the FCC’s ability to enforce the September 23, 2005 policy statement through and increases the fines that the FCC can assess to \$500,000 per violation.⁴³

¶56 But there are even better reasons for Congress—and for the FCC—not to try to regulate these issues: it’s hard to do right,⁴⁴ and there’s no need to even try. As argued earlier, networks are inherently non-neutral. Until Congress develops an expertise in building networks, it will be ill-equipped to determine whether a given network architecture was implemented for technical or discriminatory reasons.⁴⁵ What’s more, non-neutral elements are often buried deep inside of a network. Given the complexity of modern networks—common infrastructure routers can route terabytes of data between dozens of networks every second, and a large network peering point can have a dozen such routers—the government is unlikely to be able to exercise any effective oversight authority.

¶57 Where we are able to identify non-neutral effects, we then run into the difficult task of separating legitimate non-neutrality—guaranteeing minimum qualities of service in exchange for access to excess capacity—from non-legitimate cases. And finally, where we do find successful exercises of illegitimate non-neutral power the antitrust laws should kick in on any number of theories. Given that there is intermodal competition at the access and content levels, that a non-neutral strategy would work is itself strong evidence that it is in abrogation of competition.⁴⁶

¶58 Regulation is hard to do right. Fortunately in this case we do not need to regulate neutrality practices. Network neutrality concerns are symptomatic of a problem with how we fund the Internet. Part III of this Article turns from talking about network neutrality to talking about its underlying causes and how we can resolve them.

42. Communications Opportunity, Promotion, and Enhancement Act of 2006, H.R. 5252, 109th Cong. (2006). This legislation did not pass the Senate. A bill specifically targeting network neutrality was introduced in the Senate in the first days of the 2007 session. Internet Freedom Preservation Act, S. 215, 110th Cong. (2007). See Stephen Labaton, *Congress To Take Up Net's Future*, N.Y. TIMES, Jan. 10, 2007.

43. *Id.* “The Commission shall have the authority to enforce the Commission’s broadband policy statement and the principles incorporated therein.” This bill creates a new section in the Telecommunications Act, 47 U.S.C. § 715.

44. See Christopher S. Yoo, *Beyond Network Neutrality*, 19 HARV. J.L. & TECH. 1, 11 (2005) (“In addition, the regulatory tools needed to implement the regime of interconnection, standardization, rate regulation, and nondiscrimination implicit in network neutrality have long been criticized as difficult to implement and unlikely to be effective in industries like broadband . . .”).

45. Consider, for example, *Verizon Communications Inc. v. Law Offices of Curtis V. Trinko, L.L.P.*, 540 U.S. 398, 408 (2004) (“Enforced sharing . . . requires antitrust courts to act as central planners . . . a role for which they are ill suited.”).

46. My earlier suggestion that a successful competitive campaign might be built upon non-neutrality might be thrown back at me here. But that discussion argued that competition for consumers would benefit the non-neutral provider. This leaves open the possibility that content competitors have an action against the non-neutral party.

IV. MONEY: STILL THE ROOT OF ALL EVIL

¶ 59 The FCC might be right—competition might be enough to keep companies like AT&T from discriminatorily flexing their network muscles. But even if it is not, the Commission is wise to stay out of the network neutrality debate. Network neutrality is a great rallying cry—it evokes visceral feelings of freedom and liberty, but the potential for network non-neutrality is only a symptom of a more fundamental problem: who pays for the Internet. This part looks at the current model used for paying for the Internet, and proposes a simple new rule, “neighbor billing,” to reduce higher-level network neutrality concerns.

A. Building The Infrastructure

¶ 60 AT&T and other companies are spending billions of dollars to upgrade their networks to support high-bandwidth next generation services like IPTV.⁴⁷ If this infrastructure were to be used exclusively for television there would be no question about who pays for it: consumers and advertisers, just like traditional cable and broadcast television. Of course, it is ultimately the consumers who foot the bill in these models because advertisers pass their advertising costs on to them.

¶ 61 But AT&T’s new infrastructure is not just for television. To support typical IPTV service, which needs about ten megabits per second of network capacity, AT&T is upgrading its DSL infrastructure to guarantee twelve to fifteen megabits per second of capacity to most of its customers. When the service is in use—when people are watching television—this would leave a few free megabits of capacity for other uses, about the same amount available from current cable or DSL internet access. But when the service is not in use, the benefits of PSNs kick in: the customer has ten megabits per second of excess capacity at his disposal, which can be used for downloading Britney Spears videos. Suddenly AT&T’s shiny new high-speed IPTV infrastructure is being used to access other services. Looking to the traditional television market, if Comcast and ABC can charge advertisers for access to their networks, why should we not let AT&T charge Google and YouTube in the same way?

¶ 62 From a very long-run economics perspective there is no reason not to let AT&T charge whomever it wants however much it wants for access to its network. It might take more time, and different people might get rich along the way, but these costs will ultimately be passed along to the consumers. Either AT&T bills consumer directly, or it bills Google. If it bills Google, then Google charges its advertisers a bit more and the advertisers pass those costs on to the consumers.

¶ 63 But the time-frame to achieve this long-run equilibrium is too long to rely on the market to get these prices right There are several reasons why letting AT&T bill Google

47. See, e.g., Petition to Deny of Time Warner Telecom, *supra*, note 18, at 8; Spencer E. Ante, *Lightspeed's Slow Start: Questions dog the new high-speed AT&T network that offers phone, Net, and TV*, BUS. WEEK, Feb. 12, 2007, at 60 (“AT&T says it will pump \$4.6 billion into building enough fiber-optic cable and supporting technology to reach 19 million homes by the end of 2008.”).

does not get the pricing right. First, the last-mile market is still not adequately competitive. Many consumers still have only one choice for their high-speed internet access, and recent consolidation in the telecommunications marketplace threatens what little competition there is. But, ironically, the concern here is not that AT&T is a monopoly—the concern is actually quite the opposite. Where AT&T is a monopoly, it can charge whatever it wants, but where it faces competition, even potential competition, it needs to keep its prices low. In a competitive market, if AT&T can push its costs on to Google it is going to because this lets it reduce the price that it charges its customers—and that it can offer to its competitors' customers.

¶ 64 Why can't other ISPs respond in kind by charging Google for access to their customers and then lowering their prices to compete with AT&T? Some can, and some can't. AT&T, assuming that its pending merger with BellSouth is approved, will control access to nearly 23% of the residential and small-business high-speed Internet access market.⁴⁸ Google had revenues last year of just over \$6 billion,⁴⁹ which amounts to roughly \$3.75 million per day from AT&T customers. Most other access providers do not have a comparable bargaining position. If a customer on one of AT&T's small competitors' networks cannot access Google, he will think it is the fault of his service provider, look at AT&T's Google-lowered price, and become an AT&T customer. The small competitors cannot afford to cut off access to Google and Google cannot afford to have access cut off to the large service providers.

¶ 65 Not all of AT&T's competitors are small. Verizon, Comcast, and Time Warner each account for more than ten percent of the market, and Cox, with about six percent of the market still pushes about \$1 million per day into Google's bank account. We face a second, different problem from those companies that do have leverage against Google: how much are they going to charge? Their demands are tied to Google's revenue, not to their costs, which are already sunk. They will have every incentive to demand rents equal to however much revenue they send to Google, and little incentive to accept no for an answer.

¶ 66 A third reason—related to this last point that AT&T and others won't get the pricing right if we let them bill Google directly—is that competitive pressures will push the access providers to charge Google. If AT&T passes its costs directly on to its consumers, those consumers can directly attribute the increased costs to AT&T. They will know that costs are going up, so their demand will go down. But if AT&T can funnel these same costs through Google and then through advertisers, the consumers will be less able to attribute the increased costs to AT&T, even if it's the same group of consumers

48. These and the following numbers come from AT&T's Public Interest filing in support of its merger with BellSouth. Description of Transaction, Public Interest Showing, and Related Documentation, at 103, In re AT&T Inc. and BellSouth Corporation Application for Approval of Transfer of Control, No. FCC-06-74 (FCC Mar. 31, 2006), available at http://gullfoss2.fcc.gov/prod/ecfs/retrieve.cgi?native_or_pdf=pdf&id_document=6518332550.

49. Google Inc., 2005 Annual Report (Form 10-K), at 41 (Jan. 16, 2006), available at <http://www.sec.gov/Archives/edgar/data/1288776/000119312506056598/d10k.htm>. As this article prepares to go to print, Google's revenues for 2006 are reported in excess of \$10.5 billion. Google Inc., *Google Announces Fourth Quarter And Fiscal Year 2006 Results*, Jan. 31, 2007, available at <http://investor.google.com/releases/2006Q4.html>.

footing the bill (in reality the costs will probably be distributed over a wider population). And to the extent that there is decreased demand, it will be for the advertisers' products, not for AT&T.⁵⁰ AT&T's incentive is therefore to palm off as much of its cost as it can—doing so is like getting free money.

¶ 67 A final problem is that not every content provider is Google. If we allow AT&T and others to charge content providers directly, we subject the market to a number of distortions. The most commonly identified distortion is increased barriers to entry. If a startup is subject to rents from every major ISP, it will be subject to substantial fees and transaction costs.⁵¹ There are other effects that this billing model can have on content providers. First, as has been a theme throughout this Article, any rule distorts norms. If we let AT&T set its own rates, it will become a gatekeeper to the internet. It can charge preferred companies small fees and other companies substantial fees. In the alternative, if we require it to charge a set rate, be it fixed or proportional to some metric, and whomever sets that rate will distort the market. We might be able to prevent AT&T from controlling the distortion, but there will still be distortion. A last effect on content providers is that because these distortions and increased barriers to entry are likely to reduce their numbers, there will be fewer content providers around to fight against AT&T. Recall from the FCC's 2005 Order that one of its bases for reclassifying high-speed Internet connections as information services was that it believed that competition was moving from the access arena to the content arena.⁵² If we let AT&T adopt a pricing model that hampers this competition, we need to reconsider the wisdom of this reclassification.

B. Allocating Infrastructure Costs

¶ 68 These market failures can be more formally understood in the language of multi-sided markets. The analysis of multi-sided markets is an emerging area of economics literature⁵³ and it is one well suited to describing the Internet infrastructure's cost structure.

¶ 69 Generally speaking, a multi-sided market is one in which some platform is used to facilitate interaction between parties that cannot otherwise interact. The parties are charged for access to the platform. These markets differ from traditional markets in two significant ways. First, the demand for the platform is incidental to demand for the interaction it facilitates—for instance, users connect to the Internet to interact with content providers, not to interact with the Internet backbone. Demand for the platform is

50. For the pedantic in the audience, the reduced advertiser revenue will cut the advertisers' profits and their employees' salaries, until it ultimately comes back to AT&T. But, by that point, the increased cost has been spread among so many intermediaries that it will be written off as overhead or transaction costs.

51. Though by how much this is different from the status quo is less clear—all companies face startup costs. That the current model has low startup costs does not mean that these startup costs will be low.

52. See *supra* notes 33, 34 and accompanying text.

53. See Jean-Charles Rochet & Jean Tirole, *Defining Two-Sided Markets*, RAND J. ECON. (forthcoming 2006), (copy as presented at The Economics of Two-Sided Markets, conference (Jan. 23–24, 2004)), on file with author. This discussion will refer to two-sided markets as a species of multi-sided markets generally.

based in the value that each side of the market places on interaction with the other side. Second, the platform can charge differentiated prices to each side of the market to maximize the value of the platform. This goes beyond simple price discrimination, where different classes of customers are charged differently for similar products. Multi-sided markets can generate platform demand by differentiated pricing for products offered on different sides of the platform⁵⁴—for instance, by subsidizing content providers, the platform can create demand for new users to connect to the Internet.⁵⁵ As identified by Jean-Charles Rochet and Jean Tirole, two-sided markets importantly presuppose that the Coase Theorem⁵⁶ does not apply—if this condition fails, the parties on each side of the platform can negotiate around the platforms’ differentiated pricing scheme.

¶ 70 Common examples of two-sided markets include advertising fora, such as newspapers, dating clubs, and payment card networks. Newspapers are platforms that bring advertisers and readers together. Readers have no (or limited) interest in reading ads; and advertisers have little interest in subsidizing the publication of news. Dating clubs are similarly a platform for bringing two differentiated groups together—and again the different services are offered to draw each group to the platform. In the case of clubs, the draw for women (generally) is reduced-price drinks, and for men it is the women.

¶ 71 The literature has focused a great deal in recent years on the example of payment card networks, due to recent antitrust litigation.⁵⁷ Payment card networks create a platform that brings cardholders and merchants together. These networks illustrate another important element of two-sided markets—single- vs. multi-homing. A side of the market is said to multi-home if it connects to the network through multiple platforms; and a single-homing side is one that single sources. For instance, many consumers have only one credit card that they use to access payment card networks—they are single-homed. But merchants often accept a variety of credit cards—they are multi-homed.

¶ 72 The most peculiar aspect of these markets is how pricing is used to maximize use of the platform. Very often one side of the market is charged less for access than—or even subsidized by—the other side. Newspapers cost more to print than their 35 cent price⁵⁸ and clubs lose money giving drinks away to women. But in both cases any losses are recouped due to the increased demand from the higher-value side of the market. These differentiated pricing models make understanding the efficiency of a two-sided market platform’s allocation of costs difficult relative to that of a traditional one-sided market.

54. *Id.* § 4.

55. Many two sided markets are characterized by a chicken-and-egg problem, which subsidies can be used to overcome. For instance, by subsidizing new content providers the platform can attract more users; or by subsidizing new users the platform can attract more content providers. The preferred direction of the subsidies might depend upon the specifics of the market.

56. *Supra* note 53, § 4; Ronald Coase, *The Problem of Social Cost*, 3 J.L. & ECON. 1 (1960).

57. *See* *United States v. Visa U.S.A., Inc.*, 163 F. Supp. 2d 322 (S.D.N.Y. 2001); *United States v. First Data Corp.*, 287 F. Supp. 2d 69 (D.D.C. 2003).

58. *See* Jean-Charles Rochet & Jean Tirole, *Platform Competition in Two-Sided Markets*, 1 J. EUR. ECON. ASSOC. 990, § 7.3 (June 2003). *See also* Ulrich Kaiser & Julian Wright, *Price Structure in Two-Sided Markets: Evidence from the Magazine Industry*, 24 INT’L J. INDUS. ORG. 1, 23 (2006).

¶73 The Internet is a two-sided market.⁵⁹ It is even multiple two-sided markets, created by intermediaries.⁶⁰ Internet backbone providers create a platform for access providers, such as AT&T's consumer Internet service and content providers like Google. The access providers in turn offer a platform to customers for accessing content providers; and the content providers often provide a platform for advertisers to access the content providers' customers. Access providers' customers are typically single-homed, viewing their single ISP as a conduit to a single Internet. Content providers are typically multi-homed, sometimes directly—with multiple connections to the Internet—and sometimes indirectly—always viewing customers as coming from separate parts of the Internet.

¶74 The broader literature identifies several factors that apply to the Internet's market structure. A platform's profit-maximizing price is not necessarily the social-welfare maximizing price;⁶¹ but neither is either side's own profit-maximizing price.⁶² The relative importance of participants on one side of the platform to those on the other side effects the platform's pricing.⁶³ Generally, this benefits social welfare by reducing the platform's rent-seeking potential—even if this benefit cannot maximize social welfare. But this does not hold if one side's demand tips the other to monopoly.⁶⁴ And similarly, platforms connecting single- and multi-homed sides of a market can create competitive bottlenecks, which allow the platform to extract supra-competitive rents from the multi-homed side of the market.⁶⁵

¶75 Additional literature has focused directly on pricing structures in the Internet market. There is an entire sub-literature describing the inter-provider pricing models used by Internet access and backbone providers—so-called peering and transit agreements.⁶⁶ This literature does not specifically address questions of pricing between access and content sides of the market. However, it does draw some relevant conclusions. Perhaps most important, this literature suggests that each side of the market has no incentive to deviate from the optimal pricing absent asymmetric potentials for either side of the market to assess direct access fees against the other.⁶⁷ This conclusion ties these markets

59. See Rochet & Tirole *supra* note 53, § 4.2.

60. *Id.* § 6; see also Julian Wright, *The Determinants of Optimal Interchange Fees in Payment Systems*, 52 J. INDUS. ECON. 1, 25 (2004) (“[S]ocially optimal interchange fees may be higher or lower than the profit maximizing interchange fee because of an asymmetry in inframarginal effects.”).

61. Rochet & Tirole, *supra* note 53, §§ 4.1, 6.1; see also Wright, *supra* note 60, at 25 (“Privately optimal interchange fees may be too high.”).

62. This applies either in the case that the platform were to try to maximize the profits for one of the market's sides, or if one of the sides is able to extract a rent from the other side. *Id.* § 5 (“[I]f [one side] charges an important markup to [the other side], the platform ought to reduce [its charge to the other side] so as to limit double marginalization on that side, and increase [its charge to the first side].”).

63. Mark Armstrong, *Competition in Two-Sided Markets*, RAND J. ECON. (forthcoming 2006) (§ 4).

64. *Id.*

65. *Id.* § 5. see also Mark Armstrong & Julian Wright, *Two-Sided Markets, Competitive Bottlenecks and Exclusive Contracts*, ECON. THEORY (forthcoming 2006) (on file with author).

66. See, e.g., Iain Little & Julian Wright, *Peering and Settlement in the Internet: An Economic Analysis*, 18 J. REG. ECON. 151 (2000).

67. Jean-Jacques Laffont et al., *Internet Interconnection and the Off-Net-Cost Pricing Principle*, 34 RAND J. ECON. 370, 386 (2003) (“[T]he access charge, which mainly affects how the cost of traffic is

into the two-sided market literature, to the extent that such asymmetries exist. Unsurprisingly then, we also find that market power on one side of the market leads a platform to seek supra-competitive access fees from the other side.⁶⁸ But even without such market power, access fees alone are insufficient to achieve the socially optimal prices.⁶⁹ At the same time, access fees are a necessary component of socially optimal prices.⁷⁰ Of particular interest to the present discussion, the literature develops this last point in terms of socially optimal investment in net infrastructure.

¶76 There are three general themes to take away from the economics literature: asymmetry matters, size matters, and information matters. Applying these ideas to the multi-level, multi-sided markets seen in the Internet, we see that if customers are single-homed and content providers are multi-homed, the larger access providers will try to extract supra-competitive rents from the content providers. The content providers are powerless to oppose these rents, unless they are able to pass them directly back to those same access providers' customers. When this happens, social welfare is harmed. But, this harm can be mitigated if the upper-level platform, the backbone providers, can counterbalance this distortion—indeed, the upper-level platforms have an incentive to do just this.

¶77 Tying these ideas to reality, we can say the following: We know that generally Internet users do single-home and content providers do multi-home.⁷¹ Many content providers cannot efficiently pass access fees directly back to customers—particularly in advertising-based content markets. Therefore, allowing access providers to directly assess fees against content providers will harm social welfare, unless those fees are passed through Internet backbone providers.

C. Neighbor Billing

¶78 How then should we allocate the costs of paying for the Internet? This question has a lot of money hanging in the balance. As the network neutrality debate has gone from academic and technical circles to the FCC, to the press, and, finally, to Congress, battle lines have been drawn. We've seen unusual coalitions form between groups like MoveOn.org and the Christian Coalition. Ominously named Political Action Committees ("PACs") with difficult-to-discover affiliations and names like "Save The Internet"⁷² and "Hands Off The Internet"⁷³ have been spreading propaganda faster that you can download a Britney Spears video from YouTube. Ominous PACs are a sure sign that

divided between senders and receivers, may have no impact on the consumers' final demand and thus on traffic volume.").

68. *Id.* at 388 ("In contrast, if they have market power, backbone operators' interests are in general no longer aligned with social welfare.").

69. *Id.* at 371 ("[T]he access charge cannot by itself induce all the price differentiation that would be required for an efficient allocation in the Internet.").

70. Roberto Roson, *Incentives for the Expansion of Network Capacity in a "Peering" Free Access Settlement*, 5 NETNOMICS 149 (2003).

71. That is, most users only use a single ISP to access the Internet from a single location; but most content providers have multiple connections to the Internet.

72. Save the Internet, <http://www.savetheinternet.com> (last visited July 3, 2006).

73. Hands Off the Internet, <http://handsoff.org> (last visited July 3, 2006).

there is lots of money at stake.

¶79 Sadly, it is only after the question has received legislative attention that non-technical commentators have begun to recognize network neutrality as a question of how we pay for the Internet.⁷⁴ Even now, for most people the debate remains largely about “freedom” and “who controls the Internet.”⁷⁵ For some, it is about political power—some of the proposed legislation would have changed the political allocation of antitrust enforcement power in Washington—and, for others, it is seen as part of the consolidation of power by the large telecommunications companies—the recent mergers of MCI/Verizon and AT&T/SBC, and the pending AT&T/BellSouth merger.

¶80 Regardless the focus of the contemporary political debate, we can ask how best to resolve the underlying debate about allocating costs. The previous two sections argued that allowing the access providers to directly bill content providers distorts the market. The alternative—prohibiting such billing—is equally untenable. As a threshold matter, everyone on the Internet is a potential content provider. Access providers need to be able to charge *someone* for access, so we need a way to differentiate between content users and content providers—or we need a rule that does not rely on this distinction. We also do not want to create a rule that encourages inefficient bypasses.⁷⁶ Google *should* be able to buy faster access to AT&T’s users, if for no reason other than that it could buy a faster connection between AT&T and its servers. If Google could spend \$250,000 per month to install a direct connection to AT&T that gives it twice as much bandwidth as Yahoo!, and AT&T would be willing to allocate Google similar amounts of bandwidth over its existing network, without requiring Google to buy a dedicated data line, for \$150,000 per month, we want to allow the cheaper equivalent solution. Certainly we are not going to go the other route and tell AT&T that it cannot sell Google more bandwidth than it sells to Google’s competitors.

¶81 Most importantly, we want a rule that allocates costs better than the current system. The previous discussion suggested that allowing AT&T to directly bill Google allows AT&T to shift a disproportionate portion of its infrastructure costs to Google. The current model is not necessarily any better. For instance, imagine that a small percentage of AT&T’s customers use AT&T’s new IPTV-oriented infrastructure extensively to trade

74. The first discussion of network neutrality in these terms that I have seen was published on June 26, 2006 in the Washington Post:

Put another way, if net neutrality passes, the AT&Ts of the world will be forced to pay for all of their equipment upgrades themselves and could not subsidize that effort by imposing premium fees for premium services. If net neutrality fails, they will be able to recoup more of those costs than they can now from the likes of Google Inc., Microsoft Corp. and other major users of the World Wide Web.

At its heart, then, the battle is commercial -- over who pays how much for improvements to the Internet that we all use and sometimes love.

Jeffrey H. Birnbaum, *No Neutral Ground in This Internet Battle*, WASH. POST, June 26, 2006, at D1. Some academics have recognized that network neutrality speaks to allocation issues for longer periods. *See, e.g., Yoo, Beyond Network Neutrality, supra* note 44.

75. *See* Wu, *supra* note 6, at 145.

76. Yoo focuses on this point. *See* Yoo, *Beyond Network Neutrality, supra* note 44, at 8 (“As a result, competition policy should focus on identifying the link that is the most concentrated and the most protected by entry barriers and design regulations to increase its competitiveness.”).

videos on YouTube. Internet customers have shown a distaste for non-unlimited Internet access,⁷⁷ which leaves AT&T with two options. Lest it risk pushing its customers to other ISPs, AT&T can either increase its prices across the board to compensate for a small number of users' excess usage, or pass those costs on to YouTube. Intuitively, we want these costs to be passed on to YouTube or its customers, and if they cannot be passed on to its customers, they should go to YouTube.

¶ 82 Generally, a good billing rule will allow parties that incur costs to pass them on to the parties that create them, but will limit how these costs are passed on, such that they cannot bypass normal competitive restraints. Without these restraints, as seen above, parties can use the billing rule to extract undue rents.

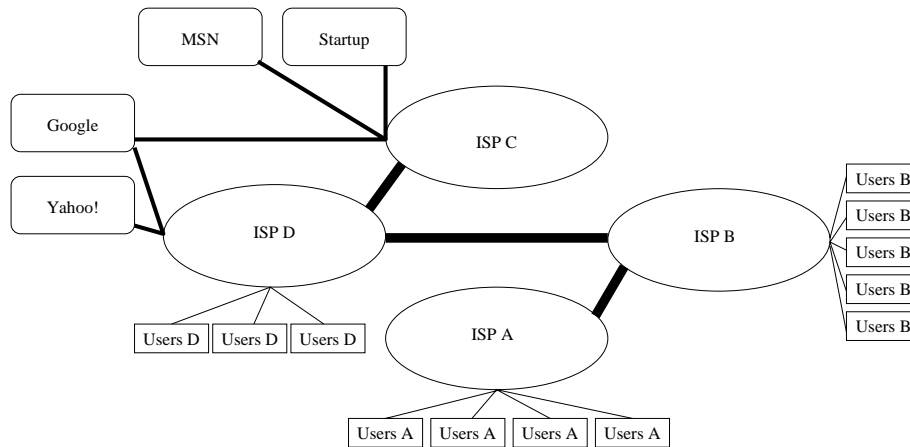


Figure 6: A hypothetical schematic view of the Internet backbone

¶ 83 I propose a simple billing rule to meet these ends: “neighbor billing.”⁷⁸ Concisely stated, neighbor billing creates a privity requirement—it requires an access provider to bill only companies that are directly connected to its network. Consider Figure 6, a schematic diagram of how the networks that make up the Internet might connect. Under a neighbor billing rule, ISP D would be able to bill its users for any bandwidth they use, as well as Google and Yahoo!. ISP D would not be able to directly bill MSN or “Startup.” It could, however, bill ISP C for any bandwidth that MSN and Startup use. Moreover, if MSN wants preferential treatment from ISP D, it can enter into a contract to secure this

77. A trend that some analysts suggest is changing. See Thierer, *supra* note 30, at 299 (“Everyone wants to charge different customers differently for different services. Everyone wants guarantees. Everyone wants to escape simple and flat pricing. Forget it.”) (quoting GEORGE GILDER, *TELECOM: HOW INFINITE BANDWIDTH WILL REVOLUTIONIZE OUR WORLD* 206 (Free Press 2000)).

78. This idea mirrors the concerns of Yoo. See Yoo, *supra* note 44 at 8. Simply, we don’t want to let companies bypass bottlenecks. Neighbor billing mirrors another argument made by Yoo and Daniel Spulber insofar as it doesn’t create new pricing rules. See Daniel F. Spulber & Christopher S. Yoo, *Network Regulation: The Many Faces of Access*, 1 J. COMPETITION L. & ECON. 635 (2005). They expressed concerns that the myriad pricing rules created by the Telecommunications Act of 1996 created opportunities for regulatory arbitrage—a concern that had been borne out in how that act was implemented (it created two price regimes for identical services (UNE-P and Resale-based telecommunications offerings), leaving no doubt that at least one regime was wrong and therefore could be manipulated). *Id.* The neighbor billing proposal rather reduces arbitrage opportunities, by reducing the number of partners to which the existing range of pricing rules can be applied.

treatment—just as it could if it were to buy a data connection directly to ISP *D*. But this agreement either would need to be negotiated through ISP *C*,⁷⁹ or would require the creation of additional infrastructure between MSN and ISP *D*. In the language of the previous section, ISP *C* is the “upper-level platform.”

¶ 84 The essential element in this model is that it forces known parties to internalize their relevant costs. ISP *B* cannot illegitimately subsidize its users by billing Google arbitrary amounts. If it were to try, it would need to send ISP *D* the bill. ISP *D* has more leverage than any of its individual customers when negotiating with ISP *B*. It also faces more competitive markets on both sides of its connection. Most content providers do not make their money directly off of consumers, making it relatively easier for them to pass off price increases without losing demand—though they will lose revenue (it is a price increase, after all). ISP *D*, on the other hand, would be raising its rates on its direct consumer, and therefore risking defection. At the same time, ISPs *B* and *D* compete in the same market, making it more difficult for cross-subsidization strategies to work: if ISP *B* tries to pass its costs off on to ISP *D*, ISP *D* can turn around and try the same strategy back on ISP *B*.

¶ 85 This is not a perfect solution—competition is not perfect in any of these markets. And it has the unfortunate characteristic that it potentially increases the number of transactions: ISP *B* bills ISP *D* who in turn bills Google instead of ISP *B* billing Google directly. But this is a market characterized by asymmetric market failures. Neighbor billing limits access to these structural defects and forces the network core to internalize many of these costs that could otherwise be illegitimately passed on to consumers. In doing so, each element of the core will compete for the minimum amount of costs that it must pass on to its customers, because those customers—unlike consumers at large—have some modicum of competitive alternative. As such, neighbor billing should do a better job of allocating costs than either of the currently proposed models.

¶ 86 Above all else, the rule is administratively simple yet offers an improvement over the status quo. This is achieved by channeling the administratively difficult questions—is a given user a content provider or consumer; is a given network structure discriminatorily or technically based; are costs being allocated fairly—into voluntary agreements between parties that have access to the information and experience necessary to make informed decisions. The law is left to examine easy-to-understand and easy-to-prove elements, allowing it to take a hands-off approach for the harder questions.

D. Does it Work?

¶ 87 This section addresses a number of foreseeable criticisms of neighbor billing. It is not a perfect solution. But those problems that it does not resolve are already left unresolved under the status quo, and it shouldn't make anything worse. This is not meant to be a revolutionary change—and as evolutionary changes go, the goal cannot be to

79. Note that, if MSN negotiated an agreement directly with ISP *D*, this could impose additional costs on ISP *C* as MSN's usage of its network, as the route to ISP *D*'s network, increased. ISP *C* properly has a role to play in this transaction.

resolve every outstanding problem at once, but rather to offer incremental improvements over the status quo.

¶88 An initial concern is that most “neighbor” relations are governed by preexisting contracts. Either the neighbors are an access provider’s customers, or they are networks to which the access provider connects. In networking parlance, connecting networks are called “peers,” and they connect at “peering points.” In the case of existing content-consuming customers, access providers are already actively considering how to revise their contract terms as they upgrade their networks. At a minimum, they can limit users to pre-upgrade speeds, and deny them any excess capacity created by these upgrades, though it remains to be seen how such a decision would affect customer demand.⁸⁰ But remember, the very reason that access providers are considering alternative billing models is to defer costs from these users. This is not the case with an access provider’s content-producing customers. When ISP *B* sends ISP *D* a bill for Google’s access, ISP *D* needs some way to pass those costs on to Google, assuming that the costs are legitimate. Here again, at a threshold level, the access provider can use denial of access to upgraded facilities as leverage in renegotiations. The access provider will presumably only be willing to pass on reasonable rate increases, such that the customer, if it wants access to the upgraded facilities, will have no incentive to switch providers and therefore will agree to pay the passed-on rates.

¶89 The more fundamental existing-contracts problem rests at the peering level. Peering contracts are traditionally “settlement free” or “bill and keep” contracts. Under these contracts, each peer agrees to bill its own customers and not to require other peers to share revenue from their customers. The assumption is that traffic is as likely to flow from ISP *B*’s network to ISP *D* as from ISP *D*’s network to ISP *B*. As such, the total distribution of costs is proportional to the distribution of users on each ISP.⁸¹ A settlement-free model properly allocates costs, but avoids transaction costs. Unfortunately, this proportional distribution does not hold where the distribution of different types of users across access providers is not uniform, or where different providers have different cost structures. These are the circumstances we see with AT&T and Google: AT&T has a disproportionate number of users on its network, and a disproportionately high cost structure for serving them. Basic contract law tells us that AT&T is bound by the terms of its existing peering contracts in these cases—it took a gamble and lost. Empirically, however, these contracts are likely to be renegotiated. In late 2005 a dispute over the terms of a peering contract broke out between Level 3 and Cogent, two access providers.⁸² Level 3 had become a much larger provider than Cogent, and had come to serve a different distribution of customers. Ultimately, after a brief war that involved Level 3 blocking Cogent’s users from its network, the contract was renegotiated. More generally, the FCC has shown a willingness to step in when the

80. See Thierer, *supra* note 30, at 299.

81. Some simple math shows that, so long as users are randomly distributed across the networks and the networks have similar costs, the sizes of the networks don’t change this equilibrium.

82. Ben Charny, *Cogent-Level 3 Peering Spat Ends—for Now*, E-WEEK, Oct. 7, 2005, available at <http://www.eweek.com/article2/0,1895,1868765,00.asp>.

balance that makes settlement-free contracts work is upset.⁸³

¶90 Settlement-free peering contracts do not prohibit ISPs from directly billing one another's customers; they only prohibit peers from billing each other. In effect, neighbor billing reverses this configuration. But neighbor billing does not necessarily eliminate the use of peering contracts. Rather, it forces the peers to internalize the costs of negotiations, and draws affected parties to these centralized negotiations. Consider, for instance, what happens if ISP *B* from Figure 6 agrees to prioritize Google's traffic: Yahoo! will go to ISP *D* and negotiate to have *its* traffic prioritized over Google's traffic that is heading to ISP *B*. Suddenly both ISPs are distorting traffic, but are doing so in a way that has the net effect of canceling the distortions out—no one wins. Now consider the alternative: Google buys a connection directly from ISP *B*. In this case Yahoo! cannot “undo” Google's advantage, other than buying more capacity from ISP *B*. But buying this sort of capacity will allow ISP *B* to increase its overall capacity: here Google and Yahoo! are both paying for improved performance, and both receive it. This later model *can* be achieved without requiring Google and Yahoo! to buy dedicated connections to ISP *B*—that's the power of packet-switching.

¶91 Neighbor billing lets us reach this “best of both worlds” result. The problem seen above stems from the fact that Google and Yahoo! each negotiate with a separate access provider. A neighbor billing rule channels these negotiations through a single access provider (ISP *D*). The access provider will not be willing to tell either party that it is degrading that party's performance relative to the other party, for fear of losing a customer. As such, any service premiums will be for excess capacity, not for preventing reduced minimum capacity. Of course, in the long term, this result achieves under either model: when Google learns that ISP *D* is reversing the effects of Google's contract with ISP *B*, it will be unhappy and ISP *D* will face a potential lost customer. But ISP *D* has no way of learning of Google's contract with ISP *B* until Google learns of Yahoo!'s contract, at which point the complexity of escaping this contractual morass can probably only result in litigation. Neighbor billing is a rule that consolidates such information before contracts are signed. The threat of future potential lost competition aligns ISP *D*'s interests with those of both Google and Yahoo!.

¶92 More generally, the technologies to implement service guarantees necessarily need to be implemented in the core of the network. You simply cannot implement a guaranteed a level of service at the edges of the network that will provide the required level of service as data transits away from the edges and through the core. Yet current billing proposals attempt to create service-level guarantees at the edges. To the extent that billing should reflect costs, this is an approach doomed to fail. Neighbor billing takes the opposite course, pushing the billing decisions to the network core. It is only there that billing and service level guarantees can be matched up and allocated to reflect the costs of running the network.

83. *See, e.g.*, CC Docket No. 01-92, In the Matter of Developing a Unified Intercarrier Compensation Regime, 20 FCC Rcd. 4685 (2005); *AT & T Corp. v. Iowa Utilities Bd.* 525 U.S. 366, 378 (1999) (“The FCC has rulemaking authority to carry out the provisions of [the Telecommunications Act], which include [sections governing settlement regimes].”).

¶93 There are other problems that neighbor billing does not directly address. For instance, it does not speak directly to an access provider's ability to prioritize its internal traffic or to leverage its power against smaller access providers. To the extent that these are real concerns, however, they are generally subsumed by antitrust law. And the rule at least marginally increases the power of smaller access providers, because it limits the larger players' ability to bypass the smaller ISPs. If AT&T wants to offer a customer of Mom-n-Pop's Internet premium access on the AT&T network, it will need to go through Mom-n-Pop's network. At worst, Mom-n-Pop's cannot be made any worse off; at best, Mom-n-Pop's can bundle its own premium access on top of that offered by AT&T, allowing it to recover at least any additional costs that AT&T's premium service creates. This likely would result in more traffic from the targeted customer on Mom-n-Pop's network. More generally, the market is in a clear transition phase away from small access providers, and to providers that can offer "triple play" features: voice, television, and data all over a single Internet connection. If we were not facing this transition, the FCC would have never reclassified high-speed Internet as an information service, Ed Whitacre would have never made his famous statement, and network neutrality would still be a topic relegated to (very good) papers written by Tim Wu.

V. CONCLUSION

¶94 Network neutrality has become a very contentious issue over the past year. The stage for the war that is now playing out in the press and in Congress was set in September of 2005 when the FCC abrogated the traditional common carrier rules that applied to high-speed Internet connections. This paved the way for AT&T and other network access providers to suggest an intention to charge content providers for access to users.

¶95 While the debate is heated and passion-filled, it really is just about money: who pays for the Internet. There are no great big freedom-of-speech crushing bogeymen here. Ed Whitacre has nothing against liberty and freedom—he just wants to be paid his due by the people who profit off of the network he's paying to build.

¶96 Indeed, network neutrality cannot be about who controls how content is transferred over the Internet. The Internet is inherently non-neutral. Packet switching networks, by their very nature, create the possibility of congestion on the network. This is a fair tradeoff, given the substantial benefits that these networks create. But where there is congestion, rule-based decisions about how to allocate scarce resources will always distort the "fair" or "neutral" ordering of access. In this spirit, it is essential that any rule adopted preserve the incentive of network owners to converge their networks into a packetized infrastructure. This means that they need to be able to secure minimum quality of service guarantees. This is a small price to pay for the excess capacity and potential for innovation that these networks create.

¶97 Once we understand that the Internet is inherently not neutral, and that we do not want it to be, only then can we focus on the real question: how do we properly allocate the costs of building these new networks. The existing system is riddled with market failures, the largest of which is that access providers can push costs onto content

providers with more ease than the content providers can push them back. As a result, the current pricing model does not correctly allocate costs. This has not historically been a substantial problem, but as access providers like AT&T upgrade their networks to offer next-generation services like IPTV and voice-over-internet protocol it has become, and will increasingly be, one. We want AT&T to spend billions of dollars to create these new networks, which means that we want AT&T to be able to secure funding to build them. The concern is, however, that AT&T will be able to push too much of its costs off on to other companies by manipulating existing market failures.

¶98 This Article proposes a simple rule—“neighbor billing”—as a way to reduce these problems. It is not a perfect rule; but it does internalize many of the most dangerous externalities. Most important, it limits the ability of one access provider to push its costs onto another access provider’s customers. In effect, this makes each access provider the agent of its customers, sandwiching its interests between those of the content providers on one end and its competing access providers on the other. Even if neighbor billing is impractical as a billing rule, it highlights elements that are largely absent from the present discussions about net neutrality.

¶99 This rule aside, the key points to take away from this Article are three: the Internet is inherently not neutral; we do not want it to be neutral; and the net neutrality debate is really about how we pay for the next-generation Internet.