Federal Communications Law Journal

Volume 63 | Issue 2 Article 6

3-2011

Wireless Efficiency Versus Net Neutrality

Charles L. Jackson
George Washington University

Wireless Efficiency Versus Net Neutrality

Charles L. Jackson*

* Dr. Charles L. Jackson is an electrical engineer who has worked extensively in communications and wireless. He has been both a digital designer and a system programmer. He works as a consultant and as an adjunct professor at The George Washington University, where he has taught graduate courses on computer security, networking and the Internet, mobile communications, and wireless networks. Dr. Jackson consults on technology issues—primarily wireless and telecommunications. Dr. Jackson served three terms on the FCC's Technological Advisory Council. He previously worked at both the FCC and the House Commerce Committee. He holds two U.S. patents. Dr. Jackson received his Ph.D. from MIT.

I. INTRODUCTION

Almost all systems in the world have limited capacity. Nature makes the capacity of systems variable, despite the best efforts of their designers and operators; they are best modeled as a random quantity. Consider the capacity of the airways between Washington, D.C., and New York. Although there is an upper limit set by the capacity of the airports at each end, weather often reduces capacity well below that upper limit. The supply of electricity also fluctuates. Generators and transmission lines fail; river flows and winds vary. The capacity of some geostationary communications satellites comes in physical units called transponders, which can fail unexpectedly. The electrical power industry and the satellite industry have developed a variety of priority mechanisms to deal with such fluctuations.

Wireless networks and the Internet face similar limits. Equipment failures and fluctuating demand can result in situations in which users try to transmit more traffic than the network can carry. As described, one response to such overload in electricity and satellite communications is to give preferential treatment to one type of use or class of customers in order to match demand with capacity. There are currently a variety of policy proposals for wireless and Internet communications, referred to under the broad term *network neutrality*, that propose to prohibit or limit such preferential treatment when traffic overloads occur. This Article reviews congestion and interconnection issues in the Internet and wireless networks, and points out a number of ways in which such limits on preferential treatment could harm consumers.

This Article first reviews congestion and congestion control in the Internet; second, the Article turns to wireless networks and shows that in addition to congestion issues, priority routing in wireless can make available capacity that would otherwise go unused.

Policies that facilitate the wider availability and adoption of broadband access to the Internet promote a wide variety of public interest objectives, including jobs, safety of life, and quality of life. Conversely,

restrictive regulations tie the hands of network engineers and managers, and prevent continued innovation that would make broadband networks less robust, less useful, and less secure. In addition, such regulations deny consumers certain services that may be effectively precluded in the absence of particular forms of network management. The successful operation of a broadband network requires considerable attention by network operators to many significant background details, such as protecting against security threats, controlling congestion, and making sure that delay-sensitive applications like VoIP and interactive games perform well. Allowing providers the flexibility to employ the tools and practices that most effectively address these concerns benefits all broadband consumers.

II. CONGESTION IN THE INTERNET

Congestion has long been a real problem for the Internet. Priority routing can, among other things, be an effective tool for controlling and minimizing the harms of congestion. Giving one class of traffic priority over another can substantially reduce the harms from congestion by enabling latency-sensitive applications that would fail in the absence of network management. Moreover, in the wireless world, giving some traffic priority over others permits expanding capacity without imposing significant costs.

This Article discusses congestion control in the Internet as it has been practiced in the past and as it is practiced today. It also describes recent incidents of system collapse and how blocking low-priority traffic was a key factor in recovering from such collapses. The Article concludes that congestion controls within the network—congestion controls that do not treat each packet equally—offer substantial benefits for consumer welfare and public safety. In this context, the Article describes how certain tools, technologies, and congestion control techniques—including packet inspection technologies—though criticized by some,[1] can provide highly effective defenses against network attacks, in particular against denial-of-service attacks.

As this discussion will show, imposing any form of a rule that prohibits any differential treatment or handling of different packets would create substantial efficiency losses by prohibiting the use of technologies that expand capacity, protect against congestion, and enable services or applications that would otherwise not function effectively. Such a rule would also make broadband networks less robust and less secure than they

1. *See, e.g.*, M. Chris Riley & Ben Scott, *Deep Packet Inspection: The End of the Internet as We Know It?*, FREE PRESS (Mar. 2009), http://www.freepress.net/files/Deep_Packet_Inspection_The_End_of_the_Internet_As_We_Know_It.pdf.

would otherwise be.

A. Controlling Internet Congestion

Congestion in the Internet is not merely a theoretical concern—it has long presented a real-world challenge for network engineers. A famous paper by Van Jacobson and Michael Karels describes several congestion collapses of the Internet.[2] The development of effective congestion control mechanisms was a key step in developing the modern Internet. Unfortunately, the primary congestion control mechanisms in today's Internet depend on the honor system for their effective operation. Incompetent or malicious programmers may subvert the honor system and set the stage for congestion failures. Happenstance, malicious acts, or equipment failure may also lead to congestion failures. Congestion is not just a problem of the 1980s, as evidenced by more recent system collapses.

The early Internet suffered a series of congestion collapses in the mid-1980s.[3] The collapses arose from a simple cause—users were transmitting more data on some paths than the paths could handle. Router queues would fill up, and subsequently arriving packets would be discarded. User machines would retransmit the lost packets, and congestion would continue. The Internet congestion was like the Beltway in Prince George's County after a Washington Redskins home game—except for the retransmissions.[4]

1. Internet Congestion Control on the Honor System

In 1993, researcher Van Jacobson of Lawrence Berkeley Laboratory described the congestion problem and the solution that he and his coworkers developed:

> "If too many people try to communicate at once," explains Jacobson, "the network can't deal with that and rejects the packets, sending them back. When a workstation retransmits immediately, this aggravates the situation. What we did was write polite protocols that require a slight wait before a packet is retransmitted. *Everybody has to use these polite protocols or the Internet doesn't work for anybody.*"[5]

---

2. Van Jacobson & Michael J. Karels, *Congestion Avoidance and Control*, 18 ACM SIGCOMM COMPUTER COMM. REV. 158 (1988).

3. Jacobson and Karels state, "In October of '86, the Internet had the first of what became a series of 'congestion collapses'. During this period, the data throughput from LBL to UC Berkeley (sites separated by 400 yards and two IMP hops) dropped from 32 Kbps to 40 bps. [We] were fascinated by this sudden factor-of-thousand drop in bandwidth and embarked on an investigation of why things had gotten so bad." *Id.* at 158.

4. Redskins fans stuck in a traffic jam are not magically cloned in the parking lot to start out again and add even more to the congestion.

5. Jeffery Kahn, *Building and Rescuing the Information Superhighway*, SCI. BEAT (Summer 1993), http://www.lbl.gov/Science-Articles/Archive/information-

Substantial thought and research went into developing congestion control mechanisms that have been embedded in TCP implementations. Although these methods are complex and subtle, the basic idea is simple: if a server or user terminal senses that the network seems to be losing packets, the server or user terminal should cut back sharply the rate at which it is transmitting data. Putting congestion control in the user devices at the edge of the network made sense for many reasons, and over the next few years, TCP implementations included congestion control features and such congestion failures became far rarer and more localized.[6]

It is, however, widely recognized that the fundamental problem still remains. There is finite capacity at every point in a network Consider automobiles arriving at an intersection of a north-south and an east-west hightway. If heavy traffic from the north, east, and west all tries to go south, the southbound road will be unable to carry the traffic and a traffic jam will ensue. Similarly, if the flow of packets arriving at a point in the Internet exceeds the traffic that can flow away from that point, some packets must be discarded. Furthermore, today's Internet congestion control works mostly on the honor system. Windows, Linux, and the Apple operating systems all come with TCP congestion control built in, but users can install software that violates (or at least abuses) the honor system.[7]

Claiming that congestion control on the Internet works on the honor system is not merely a metaphor—it is a statement of fact. Users' systems must act altruistically, sacrificing their network service for the greater good, in order for these congestion control approaches to be effective. The Internet standards body, the Internet Engineering Task Force (IETF), in its May 2009 publication, made this point:

> In the current Internet architecture, *congestion control depends on parties acting against their own interests.* It is not in a receiver's interest to honestly return feedback about congestion on the path, effectively requesting a slower transfer. It is not in the sender's interest to reduce its rate in response to congestion if it can rely on others to do so. Additionally, networks may have strategic reasons to make other networks appear congested.[8]

---

superhighway.html (emphasis added).

6. The reasons that deploying congestion control at the edges was appropriate included the facts that deploying changes to user and server software can be easier than changing routers, that user and server computers have more computing capacity available for managing such congestion, and that a key part of congestion control is a change in the behavior of devices connected to the network.

7. *See generally* George Ou, *Fixing the Unfairness of TCP Congestion Control*, ZDNET.COM (Mar. 24, 2008), http://www.zdnet.com/blog/ou/fixing-the-unfairness-of-tcp-congestion-control/1078. For example, the BitTorrent file-sharing software uploads and downloads files using multiple, simultaneous connections. If a BitTorrent client opens three connections, it can grab three times as much capacity as a traditional file download.

8. Open Research Issues in Internet Congestion Control 26 (Michael Welzl & Dimitri

A recent textbook made much the same point: "it is possible for an ill-behaved source (flow) to capture an arbitrarily large fraction of the network capacity. . . . Such an application is able to flood the Internet's routers with its own packets, thereby causing other applications' packets to be discarded."[9]

Despite the success of TCP congestion control mechanisms developed in the 1980s and 1990s, researchers have remained concerned about the threat of congestion caused by software that violates the honor code. In 1998, for example, a group of prominent computer scientists authored RFC[10] 2309, titled *Recommendations on Queue Management and Congestion Avoidance in the Internet*, setting forth some of their concerns.[11] The fifteen authors of this RFC include many of the best-known researchers on congestion control in the Internet. The authors repeatedly express concern about "the potential for future congestion collapse of the Internet" and describe scenarios in which "the Internet is chronically congested."[12] In particular, they address congestion from applications which "can grab an unfair share of the network bandwidth."[13] As the authors recognized, software with the capability to do exactly that was available a decade ago. Such software is far more widespread today.[14]

In the web-services context, persistent connections are TCP connections that are kept alive over time in order to speed web-server response by avoiding connection setup delays. Persistent connections speed up web downloading, but they can impose higher traffic bursts than newly established connections. If a user kept a large number of persistent connections open to a web server, he could download multiple files quickly—but at the risk of creating congestion problems on the route between the web server and the user's computer. Consequently, Internet standards recommend that web browsers have no more than two persistent connections to a single website.[15] However, not all web browsers follow

---

Papadimitriou eds., May 2009) (working draft expired Nov. 16, 2009), http://tools.ietf.org/html/draft-irtf-iccrg-welzl-congestion-control-open-research-04 (emphasis added).

9. LARRY L. PETERSON & BRUCE S. DAVIE, COMPUTER NETWORKS: A SYSTEMS APPROACH 470 (4th ed. 2007).

10. Requests for comments (RFCs) are the standardization documents for the Internet and are published by the IETF. *Requests for Comments*, INTERNET ENGINEERING TASK FORCE, http://www.ietf.org/rfc.html (last visited Feb. 21, 2011).

11. B. Braden et al., *Recommendations on Queue Management and Congestion Avoidance in the Internet*, IETF RFC 2309 (rel. Apr. 1998), http://datatracker.ietf.org/doc/rfc2309.

12. *Id.* at 9.

13. *Id.*

14. BitTorrent file-sharing software is one example of software that violates the honor system.

15. RFC 2914 states:

this recommendation. The extensively used Firefox web browser, for example, allows the user to edit some of the network settings. Figure 1 shows the control panel of an add-in that simplifies that editing process with the number of persistent connections per server set to sixteen and the maximum connections per server set to sixty-four. These settings improve performance, but they clearly violate the honor system and have the potential to hinder the overall performance of the network and to degrade the service of other users, especially if widely used.

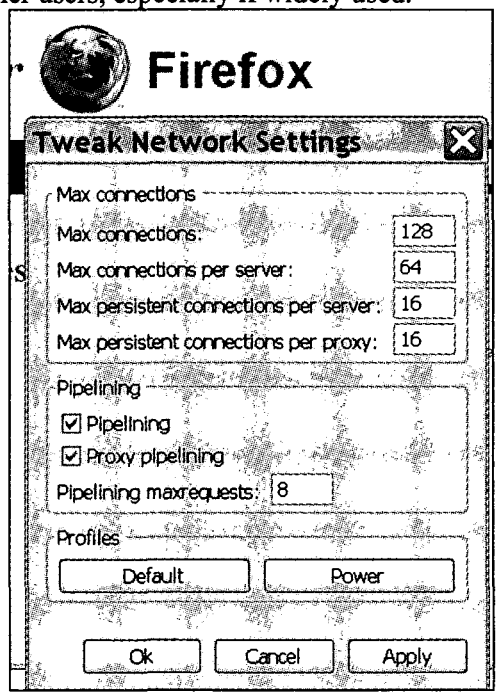


Figure 1. Firefox network control panel showing a maximum of 16 persistent connections rather than the RFC 2616 maximum of 2.[16]

The Internet community is well aware of the congestion risk created

---

The specific issue of a browser opening multiple connections to the same destination has been addressed by RFC 2616, which states in Section 8.1.4 that "Clients that use persistent connections SHOULD limit the number of simultaneous connections that they maintain to a given server. A single-user client SHOULD NOT maintain more than 2 connections with any server or proxy."

S. Floyd, AT&T Ctr. for Internet Research at ICSI, *Congestion Control Principles*, IETF RFC 2914, at 5 (rel. Sept. 2000), http://www.rfc-editor.org/rfc/pdfrfc/rfc2914.txt.pdf.

16. Figure 1 shows the Author's Firefox browser configured to maintain sixteen connections to a server or proxy—that is eight times more than the number in the standard. This setup is illustrative. I run my browser with the default settings, not these greedy settings. Of course, the default setting is six—triple the recommended number.

by nonconforming applications such as the Firefox browser. For example, an Agilent white paper states:

> **Mischievous Applications** - In spite of efforts to modify TCP or queue management to improve fairness, achieve better link utilization, and so on, an important consideration is that applications themselves are evolving to exploit the nature of networks and take an unfair share of bandwidth. For example, the open-source browser Firefox opens multiple TCP connections in [an] attempt to manipulate the network. More widespread and problematic are peer-to-peer applications such as BitTorrent that make multiple small requests over different TCP connections, ultimately defeating the principle of fairness that TCP and queue management researchers seek to uphold. Properly managing such mischievous applications requires going beyond dealing with individual flows or connections.[17]

Sophisticated users and developers of applications are also well aware of both the potential individual benefits and collective harms of violating the congestion-control honor code. For instance, a blog entry describing how to improve Firefox performance included the qualifier: "Bear in mind however that the more connections you are tying up, the less that will be available to others wishing to connect to the same server - so don't set this excessively high just because you can."[18]

Web browsers are not the only software that may violate the honor code of the Internet and contribute disproportionately to network congestion and increased delay. Some peer-to-peer software also does. The Agilent white paper notes that BitTorrent can open dozens of TCP connections to download a file—thus greatly speeding downloading, but risking congestion and possibly taking an unfair share of network resources.[19] Agilent's reference to taking an unfair share of network

---

17. AGILENT TECHS., *TCP and Queue Management*, at 6 (2008), http://cp.literature.agilent.com/litweb/pdf/5989-7873EN.pdf.

18. *About FireFox's Connection*, PINGUY'S WEBSITE, http://pinguy.infogami.com/blog/3915 (last visited Feb. 21, 2011). Other blogs also suggest tuning Firefox to increase performance, but do not explain the negative consequences for others. *See* Sandip Dedhia, *21 About:Config Hacks(Tweaks) for Firefox 3*, BLOGSDNA (June 22, 2008), http://www.blogsdna.com/372/21-aboutconfig-hackstweaks-for-firefox-3.htm; Serdar Yegulalp, *Hacking Firefox: The Secrets of About:Config*, COMPUTERWORLD (May 29, 2007, 12:00 PM), http://www.computerworld.com/action/article.do?command=viewArticleBasic&taxonomyName=Networking+and+Internet&articleId=9020880&taxonomyId=16&pageNumber=5; Damien Oh, *28 Coolest Firefox About:Config Tricks*, MAKETECHEASIER (Aug. 21, 2008), http://maketecheasier.com/28-coolest-firefox-aboutconfig-tricks/2008/08/21. The help page for the Opera browser states, "It is recommended to keep the default setting of 16 [maximum connections to a server], but you can try changing the maximum number of connections to a single server if you are experiencing problems with browsing speed." *Advanced Preferences: Network*, OPERA HELP, http://help.opera.com/Windows/10.63/en/network.html (last visited Feb. 21, 2011).

19. BitTorrent opens multiple TCP connections that together are less responsive to congestion than a single TCP connection. See the discussion of BitTorrent, *infra* notes 20–22 and accompanying text.

resources reflects the fact that if two users are sharing a communications link—one using a web browser to view a video feed from Hulu.com and the other using BitTorrent to download a movie—the BitTorrent user might receive fifty times as much of the link's capacity than would the viewer of the video. This unfair sharing would not create a problem if the link had one hundred times more capacity than needed to view the video stream. But, if the link had only ten times as much capacity as needed to view the video stream, the Hulu.com user would get about one-fifth of a video channel and the BitTorrent user would get about 9.8 video channels of capacity.[20] The Hulu.com user would get to watch the clip, but he or she would either have to wait half an hour to watch a six-minute clip with interruptions or have to accept pauses in viewing while the programming trickled into the buffer. Applications such as BitTorrent can also fill network buffers and thereby delay other applications and other users.

BitTorrent does not dispute this latter fact. About two years ago, a BitTorrent position paper explained:

> When a user starts a typical implementation of BitTorrent today, multiple uploading TCP connections entirely saturate the uplink and fill the buffer in the bottleneck device, typically cable or DSL modem. This imposes an additional delay on all traffic, equal to the size of this buffer divided by the uplink bitrate. In typical home usage cases, this additional delay can range from a second to four seconds or so. An increase in RTT of this magnitude not only starves out other TCP connections, *it quickly makes real-time communication, such as VoIP and games, entirely impossible.*[21]

BitTorrent is aware of the problems created by its protocol and is working to develop, deploy, and standardize a protocol that can coexist more peacefully with VoIP and interactive gaming.[22] Even if BitTorrent does fix its protocol to be more friendly to other applications, ISPs will always have to deal with new software and new problems. Denying ISPs tools to deal with disruptive or unfair software will harm consumers.

One of the factors that permits the public Internet to work is that most software follows the honor system for congestion control. However, if ISPs lack the ability both to manage traffic that is not obeying the honor system

---

20. On January 27, 2011, I used packet capture tools to verify that Hulu.com uses a single TCP connection to transfer a video clip.

21. Stanislav Shalunov, *Users Want P2P, We Make It Work*, HACKING STARTUPS (May 28, 2008), http://shlang.com/talks/20080528-BitTorrent-position-IETF-P2P.pdf (emphasis added).

22. *See 2010-06-03 Charter*, LEDBAT STATUS PAGES, http://tools.ietf.org/wg/ledbat/charters (last visited Feb. 21, 2011) (setting forth the current charter of the Low Extra Delay Background Transport (LEDBAT) Working Group of the IETF's Transport Area). When the group first came into being it was cochaired by a BitTorrent employee, and BitTorrent has contributed in other ways to the working group's operation.

and to use approaches that make their networks "smarter," then they may be unable in the future to keep their networks running—at least at a level that satisfies consumers' expectations and needs—if widespread violations of the honor system proliferate.

2. More Recent System Collapses

Concern about congestion collapse in today's Internet is not theoretical. On December 26, 2006, a large earthquake took down twelve of the eighteen cables between Taiwan and the Philippines. Internet service in much of Asia was seriously impaired. Bob Briscoe reported that an ISP in Singapore, SingNet, restored service before the cables were repaired by blocking video downloads and gaming traffic.[23] That is, by the simple expedient of giving e-mail, VoIP, and normal web browsing priority over video downloads and gaming, SingNet was able to restore Internet service to most users.

In this case, network overload was precipitated by a massive hardware failure. But network overload can arise from many other factors. Flawed hardware can create overloads as can malicious or faulty software. Automated access to Network Time Protocol (NTP) servers has been the source of several localized network overloads. The NTP provides the Internet's equivalent of a clock on the wall. Any computer on the Internet can query an NTP server and find out the current time. Operating systems and network hardware often have NTP clients built in. These built-in clients permit the equipment to set the time automatically without any operator intervention. For example, once a week, the time-of-day clock on my computer asks the NTP server at time.windows.com to provide the correct time.

There have been several incidents in which such NTP client software went awry and overloaded some facilities. Perhaps the most well known occurred in May 2003, when the University of Wisconsin NTP server was flooded with hundreds of megabits per second of NTP traffic.[24] The cause of this traffic was a router manufactured by NETGEAR that was hard coded to query the university's NTP server. That code in the router queried

---

23. Bob Briscoe, Toby Moncaster & Louise Burness, We Don't Have to Do Fairness Ourselves (Nov. 12, 2007) (unpublished working paper), http://www.bobbriscoe.net/projects/2020comms/accountability/draft-briscoe-tsvwg-relax-fairness-00.html. Cable failures in the Mediterranean in January 2008 also precipitated Internet failures. *See* Tomasz Bilski, *Disaster's Impact on Internet Performance–Case Study*, 39 COMM. COMPUTER & INFO. SCI. 210, 213–14 (2009), http://www.springerlink.com/content/r4278513t4424254/fulltext.pdf.

24. *See, e.g.*, Dave Plonka, *Flawed Routers Flood University of Wisconsin Internet Time Server* (Aug. 21, 2003), http://pages.cs.wisc.edu/~plonka/netgear-sntp/; *University of Wisconsin - Madison and NETGEAR Joint Statement on NTP*, NETGEAR (Dec. 10, 2009), http://kb.netgear.com/app/answers/detail/a_id/1112.

the NTP server once per second until it received an answer. If the NETGEAR router was located behind a firewall that blocked incoming UDP packets, then the router would send one query per second continuously. Dave Plonka reported that NETGEAR had manufactured about 700,000 of the affected products.[25] If all of these were operating in the defective mode, they would send about 426 megabits per second of traffic towards the University of Wisconsin.[26]

Perhaps a greater threat is posed by widely used software that automatically downloads and installs software updates. Microsoft Windows has such an automatic update feature. Consider a hypothetical but plausible scenario. Assume that Microsoft included some faulty code in an update to Windows in May and that the faulty code had the property that beginning on August 1, it would query the time server once a second. Buy August 1, there would be many tens or hundreds of computers running Windows with that update installed. At midnight on July 31, there would be a sudden flood of queries to the time server—a flood that would grow as midnight rolled across the globe. If we assume, conservatively, that only ten million Windows machines would have installed the software update and would be connected to the Internet, they would generate a flow of about six gigabytes per second toward the time.windows.com time server.[27] This sudden flow might disrupt parts of the network.[28] And, if many more copies of the software had been installed before the error surfaced, say it was installed on one hundred million machines, then the disruption might be widespread.

Brett Glass operates a wireless ISP named Lariat in Laramie, Wyoming.[29] In May 2009, his network was brought to its knees by his

25. Plonka, *supra* note 24.

26. NETGEAR was not the only firm to make such defective equipment. *See* Richard Clayton, *When Firmware Attacks! (DDoS by D-Link)*, LIGHT BLUE TOUCHPAPER (Apr. 7, 2006, 5:12 PM), http://www.lightbluetouchpaper.org/2006/04/07/when-firmware-attacks-ddos-by-d-link/.

27. Microsoft has its own large network that is interconnected with that of many ISPs at various locations. Consequently, the attack I describe might cause problems only on Microsoft's internal network rather than on the public Internet. I chose Microsoft Windows to illustrate this threat because most people are aware of how pervasive Windows is in the computing environment. However, many other software packages automatically download and install updates and thus impose similar risks.

28. It may seem unreasonable to posit such a programming error. However, the list of programming errors that caused massive losses is extensive. For example, CNN reported that in 2007, a flight of U.S. Air Force F-22s lost its navigation and communication systems as it flew across the International Date Line. *See Transcripts: This Week at War*, CNN.COM (Feb. 24, 2007, 7:00 PM), http://transcripts.cnn.com/TRANSCRIPTS/0702/24/tww.01.html. Navigation and communications systems support safety of life and are critical to the mission of these fighters, so one would expect that the software in these systems is subject to substantial testing and quality verification. Yet this critical software failed as the aircraft passed across the International Date Line. *Id.*

29. *See* David Farber, *[IP] An Unusual Denial of Service Attack*, INTERESTING-PEOPLE

customers' Windows machines.[30] The customer machines were all automatically downloading a large security update to Windows.[31] Glass restored normal service by managing the traffic triggered by the Microsoft update in order to ensure that it did not overwhelm the network.[32]

In addition to incompetent software, there is also the threat of malicious code. Botnets—networks of user computers that have been infected with software that permits operators of the network to use those computers—are often used to create distributed denial-of-service attacks.[33] In April 2007, there was what appeared to be an attack on the Internet in Estonia resulting in substantial disruption of Internet service there.[34]

More recently, on July 4, 2009, a wave of denial-of-service attacks hit federal government computer facilities and a few commercial computers in the United States.[35] Some computers in South Korea were also attacked.[36] The web server for the Department of Transportation appears to have been out of service for two days.[37] One can also imagine malicious code being embedded in widely used software and being used in a similar fashion to flood networks.

As the above discussion illustrates, the threat of a congestion failure on the Internet is real. Congestion failures of various magnitudes occur in parts of the Internet today, as the Estonia, SingNet, Lariat, and recent attacks of U.S. government computers all demonstrate. Congestion failure

---

MESSAGE (May 4, 2009, 11:56 AM), http://www.interesting-people.org/archives/interesting-people/200905/msg00021.html.

30. *Id.*

31. *Id.*

32. *Id.* Notice that Glass restored service by throttling legitimate Internet traffic. *Id.* The Windows security update was valuable and having user machines automatically download and install such updates is a sound practice that benefits others as well as those whose machines receive the updated software. However, having them all download it at the same time over Lariat's relatively small middle-mile connection to the larger Internet did not serve efficiency. *Id.*

33. The term "botnet" is derived from ro*bot net*work. *See Botnet*, WIKIPEDIA, http://en.wikipedia.org/wiki/Botnets (last visited Feb. 21, 2011). In 2007, Google's Vint Cerf estimated that one-sixth to one-quarter of the computers on the Internet had been subverted by botnet operators. *See* Tim Weber, *Criminals 'May Overwhelm the Web,'* BBC NEWS (Jan. 25, 2007, 2:18 PM), http://news.bbc.co.uk/2/hi/business/6298641.stm.

34. *See* Joshua Davis, *Hackers Take Down the Most Wired Country in Europe*, WIRED (Aug. 21, 2007), http://www.wired.com/politics/security/magazine/15-09/ff_estonia.

35. Lolita C. Baldor, *Federal Web Sites Knocked Out by Cyber Attack*, ASSOCIATED PRESS, July 8, 2009. Several articles indicated that the attacks were triggered by the government of North Korea. *See, e.g.*, Choe Sang-Hun & John Markoff, *Cyberattacks Jam Government and Commercial Web Sites in U.S. and South Korea*, N.Y. TIMES, July 8, 2009; Ellen Nakashima, Brian Krebs & Blaine Harden, *U.S., South Korea Targeted in Swarm of Internet Attacks*, WASH. POST, July 9, 2009, at A11.

36. Baldor, *supra* note 35.

37. *Id.*

can be caused by hardware failures, software that fails to follow the honor system, incompetently designed hardware and software, and malicious actors.

A well-accepted and essential tool in fighting these failures is the ability of ISPs to differentiate among different types of traffic, including directly managing the threat caused by particular harmful traffic. If SingNet had been unable to block file-sharing applications, it would have taken days or weeks before basic Internet services were functioning properly again. If Brett Glass had been unable to address the Microsoft downloads that were causing the problems, the users on his network would have had to endure poor service. A technology called *deep packet inspection* is one of the tools that ISPs can use to identify and manage the traffic that is disrupting network performance. Priority routing, tools such as deep packet inspection, and ISPs that are permitted to be flexible and agile are important factors that are well accepted by network engineers for their role in averting and resolving congestion failures.

3. Use of Established Congestion-Avoidance Technologies

The concept of priority traffic is not new to the twenty-first century. Networking researchers experimented with voice-over-packet networks as early as the mid-1970s.[38] It was immediately clear to these researchers that it would make sense in many situations to give voice priority over applications such as file transfer. And, from the very first days of TCP/IP, the Internet community adopted standards supporting such priority routing. To date, multiple Internet standards have been established that can be used to provide priority routing of packets. These include type of service, DiffServ, IntServ/RSVP, and MPLS.[39] For a variety of reasons, the first

38. I clearly recall attending a demonstration of voice over the ARPANET in the 1970s done by, as I recall, Bob Kahn and others. The voice did not sound very good.

39. Type of service was an option in the original IP standard, RFC 760, which had a 3-bit field for priority. INFO. SCI. INST., DoD STANDARD INTERNET PROTOCOL RFC 760 (Jan. 1980), http://www.rfc-editor.org/rfc/pdfrfc/rfc760.txt.pdf. This was modified slightly by RFC 791. INFO. SCI. INST., DARPA INTERNET PROGRAM PROTOCOL SPECIFICATION RFC 791 (Sept. 1981) [hereinafter RFC 791], http://www.rfc-editor.org/rfc/pdfrfc/rfc791.txt.pdf. Later RFCs provided substantial modifications to the priority mechanism, creating a new approach to priority that was called differentiated services of DiffServ. *See, e.g.*, P. Almquist, *Type of Service in the Internet Protocol Suite*, IETF RFC 1349 (rel. July 1992), http://www.rfc-editor.org/rfc/pdfrfc/rfc1349.txt.pdf; K. Nichols et. al., *Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers*, IETF RFC 2474 (rel. Dec. 1998), http://www.rfc-editor.org/rfc/pdfrfc/rfc2474.txt.pdf; D. Grossman, *New Terminology and Clarifications for Diffserv*, IETF RFC 3260 (rel. Apr. 2002), http://www.rfc-editor.org/rfc/pdfrfc/rfc3260.txt.pdf. RFC 2205 defined the Resource ReSerVation Protocol (RSVP). R. Braden et al., *Resource ReSerVation Protocol (RSVP)—Version 1 Functional Specification*, IETF RFC 2205 (rel. Sept. 1997), http://www.rfc-editor.org/rfc/pdfrfc/rfc2205.txt.pdf. RSVP permits the reservation of resources, such as bandwidth and queue capacity in routers, along the path between two computers on the

three of these approaches have not been extensively adopted in the Internet. However, the fourth approach, MPLS, is widely used. For example, Level 3 operates a converged MPLS core network. Level 3's public Internet and private virtual network traffic travels on the same core network, with private network traffic being given assured performance levels.[40] Any rule that requires all packets to be treated the same would probably outlaw the use of long-established approaches like DiffServ, IntServ, and RSVP. It might also threaten the efficient and beneficial separation of traffic into various priority classes on MPLS networks—a common and efficient practice benefitting consumers today.

Technology does not stand still. There are multiple research efforts to find better ways to provide priority service or assured quality of service over the Internet. A December 2008 presentation by Tim Gibson of the Defense Advanced Research Projects Agency (DARPA) described the performance of a new router developed by HP and Anagran with funding from DARPA.[41] Energy efficiency was improved by a factor of four, and throughput under conditions unfavorable to TCP was improved by a factor of forty.[42] Intimately tied to the efficiency gains of the new router are priority mechanisms that give some flows priority over others or can completely exclude flows that would overload the network. The IETF's NSIS working group is also working on improved quality of service over the Internet.[43]

4. Security

Adoption of the proposals mandating undifferentiated treatment of

---

Internet. RSVP permits reserving capacity for a communications process, such as VoIP connection, before the process begins. Such a reservation assures that the communication process will not suffer from congestion when it is active. MPLS, described in RFC 3031, can be regarded as a cross between ATM and TCP/IP—a hybrid that has advantages over either of its parents. E. Rosen et al., *Multiple Label Switching Architecture*, IETF RFC 3031 (rel. Jan. 2001), http://www.rfc-editor.org/rfc/pdfrfc/rfc3031.txt.pdf. MPLS permits network operators to employ a wide range of quality-of-service and traffic engineering techniques. RFC 4094 offers a survey of some of these quality-of-service technologies. J. Manner & X. Fu, *Analysis of Existing Quality-of-Service Signaling Protocols*, IETF RFC 4094 (rel. May 2005), http://www.rfc-editor.org/rfc/pdfrfc/rfc4094.txt.pdf.

40. *See Level 3 IP VPN Service*, LEVEL 3 COMMUN., http://www.level3.com/downloads/IP_VPN_ebrochure.pdf (last visited Feb. 21, 2011).

41. *See* Tim Gibson, *Building Authenticated and Responsive Networks that Are Faster and More Efficient*, DARPA (Dec. 18, 2008). A more detailed description of this research is given in Jack Brassil et al., *The CHART System: A High-Performance, Fair Transport Architecture Based on Explicit-Rate Signaling*, HP LABS, http://www.hpl.hp.com/news/2009/jan-mar/pdf/brassil_osr_crc_21.pdf (last visited Feb. 21, 2011).

42. *See* Brassil et al., *supra* note 41, § 7.

43. *Next Steps in Signaling (NSIS) – Charter*, INTERNET ENGINEERING TASK FORCE, http://datatracker.ietf.org/wg/nsis/charter (last visited Feb. 21, 2011).

packets could also make broadband networks and services less secure and less able to defend against a variety of threats.[44] The same tools that can limit inadvertent causes of congestion can be used to prevent and address malicious congestion.

Packet inspection or deep packet inspection provides one potentially significant tool for increasing security. Cisco sells a pair of products—the Traffic Anomaly Detector and the Anomaly Guard Module—that are designed to detect distributed denial-of-service attacks and to mitigate their harms.[45] Cisco described the functioning of the system:

> When the [Cisco] Traffic Anomaly Detector XT identifies a potential attack . . . it alerts the Guard XT to begin diverting traffic destined for the targeted devices-and only that traffic-for inspection. All other traffic continues to flow freely, reducing the impact on overall business operations while increasing the number of devices or zones a single Guard XT can protect.
>
> Diverted traffic is rerouted through the Cisco Guard XT, which is typically deployed off the critical path at any point in the network . . . . The diverted traffic is then scrutinized to identify and separate "bad" flows from legitimate transactions. Attack packets are identified and removed, while legitimate traffic is forwarded to its original destination, ensuring that real users and real transactions always get through, guaranteeing maximum availability.[46]

Some denial-of-service traffic could be detected by deep packet inspection, but not by inspection of just the headers. The ability to inspect packets also would provide an effective tool to detect and divert spam and e-mails that carry computer viruses and other malware. Packet inspection could also detect some malware that is attempting to propagate itself over the Internet.

The threat from malware is real. The National Science Foundation and the U.S. Army funded an analysis of the Conficker virus by SRI International.[47] SRI made clear the magnitude of the threat:

> Perhaps the most obvious frightening aspect of Conficker C is its clear potential to do harm. Among the long history of malware epidemics, very few can claim sustained worldwide infiltration of multiple millions of infected drones. Perhaps in the best case, Conficker may be used as a sustained and profitable platform for massive Internet fraud

---

44. Many of the various proposals for network neutrality have language that appears to exempt security practices. However, if a policy reduces the incentive to invest in equipment that both controls congestion and can also be used to provide security capabilities, networks will have less investment in security capabilities. Also, the definition of security is unclear.

45. *Cisco Traffic Anomaly Detector XT 5600*, CISCO, http://www.cisco.com/en/US/prod/collateral/vpndevc/ps5879/ps6264/ps5887/product_data_sheet0900aecd800fa552.html (last visited Feb. 21, 2011).

46. *Id.*

47. PHILLIP PORRAS ET AL., SRI INT'L, *Conficker C Analysis*, *in* AN ANALYSIS OF CONFICKER'S LOGIC AND RENDEZVOUS POINTS (2009), *available at* http://mtc.sri.com/Conficker/addendumC/index.html.

> and theft. *In the worst case, Conficker could be turned into a powerful offensive weapon for performing concerted information warfare attacks that could disrupt not just countries, but the Internet itself.*[48]

Blocking some packets—those that are harmful to users or to broadband networks—serves security. A test of my Comcast cable modem service reveals that Comcast blocks incoming traffic to TCP ports 135, 139, and 445. Each of these ports is commonly used for a service on the local network—not on the larger Internet.[49] The U.S. Computer Emergency Response Team (US-CERT), an activity of the Department of Homeland Security, recommends blocking traffic to and from these ports in order to protect against various attacks.[50] Many home computer users lack the knowledge and skills to do such blocking. Consequently, consumers benefit both from Comcast's decision to block traffic to these ports and also from Comcast's ability to block traffic to any other port should that port become a security vulnerability. Many ISPs block TCP access to port 25, as compromised user machines send e-mail spam using connections to port 25.[51]

B. Impacts of Eliminating ISPs' Congestion Control Tools

ISPs engage in a wide range of activities that reduce congestion or limit its negative effects. A requirement that all packets be treated the same, whether they are background file sharing or VoIP, would result in the failure of VoIP services at times of system overload. Choosing to treat all packets the same is an implicit favoring of delay-insensitive applications over delay-sensitive applications. The natural consequence of such a policy would be to create strong incentives for users of delay-sensitive

---

48. *Id.* (emphasis added); *see also* John Markoff, *Computer Experts Unite to Hunt Worm*, N.Y. TIMES, Mar. 19, 2009, at A17.

49. The services are RPC, NetBIOS, and SMB.

50. Several CERT Vulnerability Notes recommend blocking some or all of these ports. *See, e.g., Microsoft Server Service RPC Stack Buffer Overflow Vulnerability*, US-CERT VU #827627, http://www.kb.cert.org/vuls/id/827267 (last visited Feb. 21, 2011).

51. In May 2005, the report issued by Industry Canada's Task Force on Spam recommended practices for ISPs to fight spam. TASK FORCE ON SPAM, STOPPING SPAM: CREATING A STRONGER, SAFER INTERNET (2005), http://www.ic.gc.ca/eic/site/ecic-ceac.nsf/vwapj/stopping_spam_May2005.pdf/$file/stopping_spam_May2005.pdf. These best practices included blocking port 25. The report explained,

> Port 25 has been widely abused by spammers running zombie networks (or "botnets"). By monitoring and limiting the use of port 25, ISPs and other network operators can close off a major avenue for spamming. Canadian ISPs that have already implemented port 25 blocking have seen very significant declines in the amounts of spam originating on their networks.

John Levine, TASK FORCE ON SPAM, COMPANION DOCUMENT TO BEST PRACTICES FOR INTERNET SERVICE PROVIDERS AND OTHER NETWORK OPERATORS 4 (2005), http://www.ic.gc.ca/eic/site/ecic-ceac.nsf/vwapj/Companion_Document.pdf/$file/Companion_Document.pdf.

applications, such as voice or video conferencing, to keep their traffic on separate networks (as is the case with most voice communications today) or to move that traffic to separate networks when scale permits.

III. WIRELESS NETWORKS AND NETWORK NEUTRALITY

Wireless networks provide a particularly interesting example of the benefits of priority routing. Wireless priority routing permits use of capacity that would otherwise lie idle. The phrase "wireless network neutrality" has also been associated with criticism of handset subsidies and the bundling of handsets with wireless service. Regulators, competition policy authorities, professed competitors, and class action plaintiffs have all attacked both the joint provision of wireless service and handsets[52] and the use of various locks that tie a handset to a specific service provider.[53] The arguments raised against these practices are the usual objections to the tying or bundling of a monopoly product with a competitive product.[54] Many of the discussions of such tying focus on purely economic issues—such as consumer preferences for time payments for equipment purchases.[55] However, such discussions have failed to examine all dimensions of this issue.

Below, the Article first discusses priority routing and congestion control in wireless; it then turns to handset issues.

A. Priority Routing Expands Capacity

Modern wireless voice networks transmit signals to and from user handsets over radio channels that carry many conversations simultaneously. The quality of the radio signal received by each user can change quickly—received signal strength can change by a factor of ten within as little as a hundredth of a second. If the radio signal received by User A becomes weaker—say, because he or she has just stepped away from the window in a building—the base station in the wireless system must increase the power it uses to transmit to User A, or the telephone call will be lost. Most of the time, another user's radio channel—say, User B's channel—improves at the same time. When such an improvement occurs the power used to transmit to User B can be lowered. Most of the time these increases and

52. This discussion uses the term *handset* rather than the more clunky phrase *user terminal.* But the system efficiency concerns discussed here apply equally well to all types of terminals.

53. *See, e.g.*, Tim Wu, *Wireless Carterfone*, 1 INT'L J. COMM. 389, 400 (2007).

54. Such concerns are raised even when the argument that the wireless service is a monopoly is clearly laughable.

55. *See, e.g.*, Barry Nalebuff, DEP'T OF TRADE AND INDUSTRY (UNITED KINGDOM), BUNDLING, TYING, AND PORTFOLIO EFFECTS, 2003, ECONOMICS PAPER NO. 1 (2003), http://www.dti.gov.uk/files/file14774.pdf.

decreases cancel and total power from the base stays even.

However, sometimes the increases and decreases do not cancel out and many users need extra power. If a user needs more power on the downlink but the power cannot be increased, the call will be lost. Wireless systems protect against the threat of such failures by keeping some power in reserve—they restrict the number of calls served on a single radio link so that there will be such a power reserve. Consequently, on those occasions when substantially more than the average power is needed, the system can draw on the reserve and avoid dropping any calls.

At times when the reserve power is not needed for voice service, the reserve power can be put to effective use for data services, thus making better use of the finite capacity available in the system. To keep the voice service working acceptably, this data service must necessarily be lower priority than the voice service. At times, the voice service would demand all the downlink power and the data service would have to be suspended for as long as several hundred milliseconds. Nevertheless, a data service with substantial capacity—about fifty percent of the throughput on the voice channels in some circumstances—can be created this way if the system is able to schedule voice packets for transmission ahead of packets for the data service.

This is not a hypothetical analysis. Multiple studies have shown this to be the case for both cdma2000 and WCDMA.[56] Mehmet Yavuz and his coworkers at Qualcomm report:

> DO-Rev A can provide VoIP capacity comparable to circuit-switched cellular CDMA technologies (e.g., IS-2000) and *simultaneously* carry significant amount of other types of traffic such as non-delay sensitive applications and downlink multicast.[57]

Ozcan Ozturk and his coauthors, also at Qualcomm, state:

> Simulations also show that a significant amount of [best effort] traffic can still be served on the downlink at the VoIP capacity operating point.[58]

Imposing a rule on wireless systems that prohibits any differential treatment of packets would present a system operator with a choice between (1) running the system but restricting traffic to the level consistent with high-quality voice, or (2) running the system with more traffic but

---

56. *See, e.g.*, Mehmet Yavuz et al., *VoIP over cdma2000 1xEV-DO Revision A*, IEEE COMM. MAG., Feb. 2006, at 88; Yile Guo & Hemant Chaskar, *Class-Based Quality of Service over Air Interfaces in 4G Mobile Networks*, IEEE COMM. MAG., Mar. 2002, at 132; Ozcan Ozturk et al., Qualcomm, Inc., *Performance of VoIP Services over 3GPP WCDMA Networks*, *in* IEEE 19TH INTERNATIONAL SYMPOSIUM ON PERSONAL, INDOOR AND MOBILE RADIO COMMUNICATIONS 1 (2008), http://latam.qualcomm.com/common/documents/articles/VoIP_WCDMA_Networks.pdf.

57. Yavuz et al., *supra* note 56, at 88.

58. Ozturk et al., *supra* note 56, at 5.

delivering a service with delay and jitter that would make voice service unacceptable. If the operator chooses to offer voice—the all-time most popular service—then the traffic capacity offered by the reserve power would be wasted.

The heart of this issue in wireless arises from the fact that the capacity of the wireless link varies randomly over times that are short compared with a phone call, but that can be long compared with the duration of a single word. Humans find it hard to deal with telephone services in which occasional words are missing—there is a big difference in meaning between "Don't call me after 11:00 p.m." and "Call me after 11:00 p.m." Because people cannot tolerate such dropouts, the wireless system must have enough reserve power to cope with the variations in the radio channel. Similarly, people dislike phone service that often drops calls. In contrast, an e-mail transfer that sometimes is blocked from accessing the radio channel for a second or two works just fine for most people. Consistent with widely accepted practices throughout the industry, priority routing is the tool that lets these differing demands of voice and data customers be satisfied. In this case, priority routing is clearly not a zero-sum game. Priority routing permits use of resources that would otherwise sit idle. Prohibiting ISPs from offering priority services handicaps all application providers whose applications require connections capable of minimizing jitter or latency.

B. Priority in the Backhaul Network

The above discussion has described how treating different packets differently on the wireless access link can deliver more service or better service to consumers for a given level of investment. The same is true for the backhaul network—treating different types of packets differently can deliver better service for a given level of investment.

1. Separation of Control Signaling and User Information

In the early telephone network, control information was sent over the same links as those that carried the telephone call. In the very early days, that control was a human voice: a user would pick up a telephone and, in response to the operator's query "Number, please," would tell the operator the phone number one wished to call. Operators would speak to one another in a similar fashion in order to route calls. Later, the voice communications were replaced with digital signals transmitted in the voice band. In the mid-1970s, systems were deployed that separated the control information from the user information and transmitted the control information on a separate network. This was called common-channel interoffice signaling (CCIS). CCIS provided many advantages. For

example, in the older technology, a long-distance telephone call had to be set up all the way to the terminating switch before the call began to ring, and that long-distance connection was then tied up during the time that the destination telephone rang. This always wasted a few seconds of expensive long-distance capacity on every call—and because a large fraction of calls go unanswered, there was additional wastage. The most widely used CCIS system is known as Signaling System 7 (SS7), which is a packet network that is designed to be highly reliable.[59] Communications systems that separate the user information from the control signaling are often referred to as having a *control plane* and a *user plane.*

In the wireless industry, the term *backhaul network* refers to the communications links that run from the cell sites back to the mobile switching center and to connections to the PSTN and Internet. In early wireless systems, there were separate backhaul circuits for control signaling and user communications—the control plane and the user plane. For example, GSM uses SS7 for control-plane signaling.[60]

When networks were built using the Internet protocol, it was natural to mix control information and user information on the same packet network. Researchers had limited resources and the packet network could easily carry the control information. Building a second parallel network for control purposes would have substantially increased project cost. Combining control information and user data in a single packet network creates one major disadvantage: congestion caused by user traffic could choke off control traffic. Thus, if a misconfigured router were causing congestion problems, those congestion problems might prevent the network operator from sending reconfiguration information to the router.

The designers of the Internet protocol foresaw this problem. Their solution was to put a mechanism in the internet protocol to give network management traffic priority over other traffic. Specifically, the original 1981 standard for the Internet protocol, RFC 791, defined a precedence field that was carried in each packet.[61] The precedence field had eight values ranging from seven, the highest, for network management to zero, the lowest for routine traffic. There was also another single bit field that

59. Signalling System No. 7 is the most widely used network control standard in the telephone world. An introduction to it is provided in INT'L TELECOMM. UNION, ITU-T RECOMMENDATION Q.700 (1994), *available at* http://www.itu.int/rec/T-REC-Q.700/en.

60. GSM is the most widely used wireless standard in the world with more than three billion handsets operating on GSM networks. *Market Data Summary*, GSM WORLD, http://www.gsmworld.com/newsroom/market-data/market_data_summary.htm (last visited Feb. 21, 2011). Both AT&T and T-Mobile use GMS in the U.S.

61. *See* RFC 791, *supra* note 39, at 12.

defined whether a packet was to be processed with normal delay or low delay.[62]

2. Converged Networks

As is now common knowledge, data networking using the Internet protocol has become enormously successful and is often the best choice for implementing a communications network. The combination of voice, video, and data on a single network using the Internet protocol is sometimes called *convergence*. State-of-the-art 4G wireless networks use a converged backhaul network that combines all types of traffic—control, voice, video, and data—on a single internet protocol network.[63] Such combining of traffic has two significant advantages: (1) efficiencies arise from the need to run only one network rather than two or three; and (2) widely used Internet protocol routers and networking hardware can be used to build the combined network, rather than building the network using more expensive, specialized equipment such as SS7 packet switches that are built in relatively small volumes.

However, a converged backhaul network creates two problems. First, at times of heavy load, user traffic could create congestion that would hamper the flow of network control information. The consequence of this would be dropped calls or the inability to place a call. Second, the converged backhaul network will carry many types of traffic—most importantly voice and data. Voice is extremely sensitive to delay, whereas most data applications are not. Giving priority to voice over data would deliver more value to consumers. Moreover, there are different classes of users. Giving public safety or government emergency communications priority over general traffic allows those high-priority users to be served over a single shared network, providing great efficiencies.

3. Network Neutrality and Backhaul Networks

What would be the consequences of imposing network neutrality on wireless backhaul networks? There are two aspects of this to consider—the short-run efficiency concerns and the long-run incentives for network design and innovation.

In the short run, the impact depends somewhat on the exact definition of network neutrality that is adopted. If network neutrality meant that every IP datagram traveling the backhaul network had to be treated the same, then network management would lose any priority. The only way to assure

---

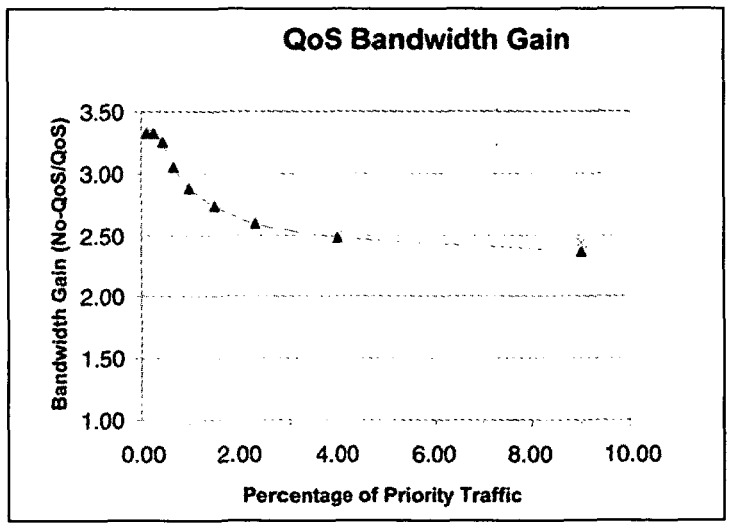
62. *Id.*

63. *See* Liu Xiheng, *Backhaul Technology in the IP Era*, HUAWEI COMMUNICATE, June 2009, at 25, 25–26, *available at* http://www.huawei.com/publications/view.do?id=5895&cid=10864&pid=61.

that management traffic would get through would be to carefully manage the level of traffic allowed and to drop user traffic whenever congestion appeared to rise. Either the quality or capacity for voice traffic would decline, or significant new investment would be needed in the network. If network neutrality allowed precedence for management data but required all user data to be treated equally, then public safety and government emergency communications could not depend on public wireless networks.

Moreover, if all applications were to be treated the same, substantial additional investment would be needed to assure that voice traffic would not be delayed. Figure 2 is a slide presented by Paul Sanchirico, vice president of Cisco Service Provider Systems Unit, at the FCC's Workshop on Broadband Network Management on December 8, 2009.[64] That slide illustrated the economic benefit of allowing voice traffic to have precedence over less-urgent data traffic.

Figure 2. Capacity benefits of priority routing



It shows the benefits of giving less delay-tolerant traffic priority over more delay-tolerant traffic. Specifically, a network with nine percent higher-priority traffic and ninety-one percent lower-priority traffic, but without any priority routing requires almost 2.5 times more capacity than does a network with priority routing, in order to meet the needs of both the

64. Paul Sanchirico, A Discussion with the FCC on the Open Internet 17 (Dec. 8, 2009) (unpublished Powerpoint slides), http://www.openinternet.gov/workshops/docs/ws_tech_advisory_process/Cisco%20FCC%20Network%20Management%20Presentation%20120809.pdf.

higher-priority and lower-priority applications.

In the long run, under any network neutrality regime, the substantial efficiencies created by separating network management traffic, higher-priority traffic, and lower-priority traffic would push for separation of the control plane from the user plane—a return to the control architecture of first-generation and second-generation wireless. These efficiencies would also push for separation of voice and data networks. Such separate voice and data networks would each be network neutral—the voice network would operate with a relatively light load, so the network would rarely experience excessive delay; the data network would tolerate increased delay, allowing the network to be used more intensely. In combination, the separate networks would be more expensive than one network employing priority to match service quality to application needs. Instead of one converged network, there would be four separate networks: a user-plane voice network, a control-plane voice network, a user-plane data network, and a control-plane data network.

C. *Cross-Layer Design*

Cross-layer design refers to the design of network elements, such as wireless access links, that take into account information from other layers to optimize performance. Cross-layer design gets its benefits at the cost of avoiding the simplifications created by the layering principal. Often this results in explicitly distinguishing between packets—something that some network regulation proposals would prohibit.

An example illustrates how cross-layer design can aid efficiency. Consider a radio link carrying two streams of traffic to and from the Internet. One stream is VoIP; the other is a TCP transfer of a web page. VoIP traffic can tolerate little delay, but an occasional packet can be lost without significant harm to the conversation.[65] The web page transfer is more tolerant of delay, but if a packet is lost, the TCP software will retransmit it until proper reception occurs.

Because radio links have much higher error rates than wired LANs, it is common for radio links to include error-detecting and error-correcting capabilities at the link level.[66] Suppose a packet is transmitted over the radio link and is found at the receiver to have arrived in error. The receiver can request partial retransmission of that packet using a technology called Hybrid-ARQ.[67] In Hybrid-ARQ retransmission, the transmitter sends

65. Typically, about one-fiftieth of a second of voice is encoded in a single packet; a packet carries only part of a single syllable.

66. *See* Y. JAY GUO, ADVANCES IN MOBILE RADIO ACCESS NETWORKS 60–68 (2004).

67. *See id.* at 64; *see also Hybrid Automatic Repeat Request*, WIKIPEDIA, http://en.wikipedia.org/wiki/Hybrid_automatic_repeat_request (last visited Feb. 21, 2011).

information, such as additional error-correcting coding, that supplements the original transmission rather than retransmitting the entire packet.

In this situation, if the receiving system detects that a packet has become corrupted on the radio link, the efficient action for the receiving system may depend on the type of packet that was received in error. If the packet is part of the TCP stream, then the receiving system should request link-level retransmission. A Hybrid-ARQ retransmission uses significantly less of the resources of the radio system than does a retransmission at the TCP level. In contrast, it might be reasonable for the receiving system to discard the VoIP packet that was received in error. Retransmitting the VoIP packet could add delay to the voice stream without any corresponding increase in the quality of the voice connection. Such a "nonneutral" link increases efficiency and improves customer's Internet experience without any harmful effects.[68] Thus, consumers get more for their money.

Somewhat related to cross-layer design is the use of cross-layer processing to improve service quality. Several manufacturers offer Ethernet switches that inspect Ethernet frames and route those frames, taking into account level three or level four protocol information. Cisco touts its ESW 500 series of switches for small business for their ability to give VoIP priority, saying, "QoS level assures that voice-over-IP (VoIP) traffic takes precedence."[69]

An analogous service could be provided in the public Internet. For example, with deep packet inspection, a carrier could examine packets to see if they represented an attempt to set up a voice call to 911 and give that call-setup attempt priority in the network. A sufficiently smart network would also be able to give priority to voice traffic to and from 911.[70]

Proposals that ISPs and wireless carriers only provide "dumb pipes"—pipes that are not smart enough to choose the most efficient retransmission and routing policies—would eliminate such potentially useful practices. Worse yet, they would stifle innovation in the development and use of such practices.

---

68. This example is illustrative. Wireless networks contain a subsystem, called the *scheduler* that manages transmissions. The exact algorithms used by the schedulers in various systems are proprietary to the manufacturers.

69. *See Cisco ESW 500 Series Switches: Small Business*, CISCO, http://www.cisco.com/en/US/prod/collateral/switches/ps5718/ps10143/data_sheet_c78-521740.html (last visited Feb. 21, 2011).

70. For example, the network could note the preliminary packets (SIP messages) from a user attempting to set up a call to 911 and could give priority to all telephony traffic from that user. (SIP is the acronym for the Session Initiation Protocol that is defined in J. Rosenberg et al., *The Stream Control Transmission Protocol (SCTP) as a Transport for the Session Initiation Protocol (SIP)*, IETF RFC 4168 (rel. Oct. 2005). SIP defines a method for setting up telephone call over the Internet.).

D. *Efficiency*

Wireless handsets are not analogous to telephone handsets. Unlike the case in wired telephony, in wireless telephony the features and quality of the handsets used on the network can have a substantial impact on the cost and quality of the wireless service, not only for the individual subscriber, but for all consumers. If User A uses an inferior wireless phone—even if that inferior phone was state of the art a few years ago—he may deny service to User B who is sitting next to him or may degrade service for other users a mile away. Widespread use of inferior handsets would substantially degrade wireless service—such as by increasing the number of coverage holes and dropped calls—or would require a significant increase in the capital plant used by wireless carriers. In either case, consumers would suffer. Wireless carriers have strong incentives to ensure that consumers use handsets that economize on total costs (capital costs and handset costs combined). In contrast, if one uses a poor-quality wireline handset, it does not degrade one's neighbor's wireline telephone service. In the economist's jargon, poor-quality wireless handsets can create substantial negative externalities, but poor-quality wireline handsets do not.

The wireless industry has seen enormous innovation and technical advancement over the last two decades. Many of these innovations have made the networks more efficient, expanding capacity and avoiding the otherwise rigid limits on capacity imposed by the finite spectrum made available for wireless service.[71] Innovations have also made new service capabilities—including data applications—available to consumers.[72] These innovations require interaction between the network and handsets to an extent that is unparalleled in wireline telephony. Seeding the market with handsets that provide expanded capabilities is an essential step in fostering the rapid adoption of more efficient or more capable wireless services. Adoption of capacity-expanding innovations would be far slower if carriers did not provide handsets supporting new capabilities. Similarly, the adoption of new services would also take longer absent carrier support of handset supply.

Various security features built into modern wireless handsets make cloning, fraud, and activation of stolen handsets far more difficult than was the case with earlier technologies. In particular, locking a handset to a network makes theft almost pointless. The adoption of such features was prompted in part by a request by responsible law enforcement agencies,

71. A variety of innovations have increased spectrum efficiency and thereby expanded capacity and lowered cost. These innovations are often known by the names of systems embodying them such as CDMA, EV-DO, and LTE.

72. New services include high-speed data services such as those provided using technologies with names like HSDPA, LTE, and Wi-MAX.

including the Federal Bureau of Investigation and the British government,[73] that wireless handsets be resistant to cloning and to easy activation after theft or robbery.

The FCC imposes several requirements on wireless carriers to support 911 calls. For example, wireless carriers must deliver all 911 calls—even calls placed by nonsubscribers.[74] The FCC also requires wireless carriers (1) to provide the location of wireless callers to 911 to the affected public safety access point (a capacity generally referred to as E911); and (2) to support communications from TTY devices used by the deaf.[75] For many carriers, meeting these two requirements is only possible if handsets contain specific features and meet minimum performance standards. As is more generally true, there is a tradeoff between handset performance and network performance in providing the location information capability. Widespread consumer use of handsets that perform the E911 functions better than industry standards may be necessary for a carrier to meet its legal obligations under the FCC's E911 accuracy requirements.

Wireless carriers provide help-desk support to their subscribers. Some modern handsets rival a personal computer of a few years ago in complexity and features. Providing help-desk support to unfamiliar or unknown handsets is difficult and costly.

Summing up, multiple technical factors, with the most important probably being the fundamental role of handsets in determining overall system efficiency and capital costs, create strong, efficiency-serving incentives for wireless carriers to control the nature and characteristics of the handsets used by their subscribers.

E. Handset Attributes and System Capacity

1. Receiver Sensitivity

The sensitivity of the radio receiver in the consumer handset is one handset feature that, if impaired, imposes costs on others. In CDMA systems, a base station transmits telephone calls to multiple subscribers using a single complex signal. That signal has fixed maximum power—typically near twenty watts. The base station divides that power among the various subscribers, transmitting to each subscriber at just above the minimum power needed to communicate with that subscriber. Consequently, base stations transmit at lower powers to subscribers near

73. *See, e.g.*, VICTORIA HARRINGTON & PAT MAYHEW, HOME OFFICE RESEARCH STUDY 235: MOBILE PHONE THEFT (2001); *Hearing Regarding Cellular Telephone Fraud: Hearing Before the Subcomm. on Crime, Terrorism, and Homeland Security* (1997) (statement of John Navarrete, Deputy Assistant Directory, Federal Bureau of Investigation).

74. *See* 47 C.F.R. § 20.18(b).

75. *See* 47 C.F.R. § 20.18(e)–(j).

the base station and at higher powers to subscribers who are more distant or who are in hard-to-reach locations—such as deep inside buildings.[76]

The sensitivity of a handset is defined by the minimum power needed to receive an acceptable signal. Consider two handsets, A and B, identical in all respects except that handset B is less sensitive than handset A—specifically, handset B requires twice as much received power to perform acceptably. A CDMA base station designed to serve twenty simultaneous conversations to type-A handsets could serve only ten simultaneous conversations to type-B handsets.[77] Looking at the problem another way, such a base station could serve twenty simultaneous conversations to type-B handsets only if those handsets were, on average, located closer to the base station. If one analyzes coverage using a simple and widely accepted model of radio propagation, one finds that a base station that could serve twenty type-A handsets spread over the area within one mile from the base station would be able to serve the same number of type-B handsets spread over an area about thirty percent smaller—the area within only 0.85 miles of the base station.[78] A wireless carrier could compensate for such a reduction in range by installing more base stations—in this case, approximately a thirty-percent increase in base stations would be needed. The base stations, the backhaul equipment needed for each base station, and the termination of backhaul at the wireless switch comprise the bulk of the capital cost in modern wireless systems.[79] A thirty-percent increase in the number of required base stations would, upon a first approximation, result in a thirty-percent increase in the capital cost of a wireless system,

---

76. Handset sensitivity in CDMA systems provides a particularly clear example of a handset feature that, if poorly implemented, reduces the network performance for other subscribers. However, in the GSM standard there are handset options, such as the AMR vocoder, that, if present and activated, permit a base station to serve more subscribers or subscribers at greater distances from the base station than would be the case otherwise. The GSM standard was originally developed by the European Telecommunications Standards Institute and is now maintained by the 3rd Generation Partnership Project (3GPP). *3GPP Specifications*, 3GPP.ORG, www.3gpp.org (last visited Feb. 21, 2011). The AMR vocoder was first specified in GSM Release 98. The current version is 3RD GENERATION PARTNERSHIP PROJECT, TECHNICAL SPECIFICATION GROUP SERVICES AND SYSTEM ASPECTS; MANDATORY SPEECH CODEC SPEECH PROCESSING FUNCTIONS; AMR SPEECH CODEC; GENERAL DESCRIPTION (RELEASE 9) 3GPP TS 26.071 V9.0.0 (2009).

77. This example is simplified. Many CDMA systems are limited by capacity on the reverse (mobile-to-base) link, not by forward-link capacity. However, were the sensitivity impairments significant, forward-link capacity would become limiting. In the high-speed data service EVDO, forward-link capacity is often limiting. EVDO is the third-generation version of the CDMA standard used by Verizon and Sprint. For more information on these standards, visit *3GPP Specifications*, 3GPP.ORG, www.3gpp.org (last visited Feb. 21, 2011).

78. The analysis is based on using an inverse fourth-power propagation law. The reduction in spacing is actually by a factor of 0.8409.

79. "Backhaul" is the transportation of wireless traffic from the cellular station to a mobile switching office from which it can be sent on to its destination.

and consequently would significantly increase the cost of wireless service.[80]

Closely related to sensitivity is the quality of the antenna on a handset. A poor antenna degrades handset performance in much the same way as does reduced sensitivity. Similarly, given that retractable antennas often fail, a service provider requirement that retractable antennas be field replaceable would make it easier for consumers to repair handsets with broken antennas. Easier repair would mean that fewer consumers will have handsets with defective antennas that consume excessive network resources.

2. Vocoder Performance

Another handset feature that has a major impact on network capacity is the performance of the voice compression subsystem in the handset. This subsystem, known as the voice coder or *vocoder*, determines how many bits per second are generated to represent a speech signal. Continuing research has resulted in the development of vocoders that perform adequately using fewer bits per second than those originally used in CDMA and GSM. These better vocoders permit more subscribers to be served over a given number of radio channels. Thus, better vocoders expand system capacity and, if better vocoders are sufficiently low cost, widespread use of better vocoders will lower total costs of wireless service.

The CDMA standard now includes vocoders called the Enhanced Variable Rate Coder (EVRC), the Selectable Mode Vocoder (SMV), and improved version of EVRC known as EVRC-B and a wideband version of EVRC known as EVRC-WB.[81] Because these are variable-rate vocoders, the network can command the handset to reduce the number of bits that are used to encode speech. The widespread use of variable rate vocoders such as the EVRC and EVRC-B vocoders in consumer handsets gives network operators several valuable options. First, the network operator can expand network capacity in times of emergency or sudden overload. Second, the

---

80. The factor-of-two difference in sensitivity between the two handsets discussed above is not an unreasonable difference from the point of view of practical receiver engineering. In late 2004, CTIA, the wireless industry association, filed with the FCC reports of recent tests of PCS handsets performed by independent laboratories. These tests showed, among other things, that the tested handsets were on average able to pick up signals less than half as strong as the weakest signals that could be picked up by a handset just meeting the requirements of the industry standard. *See* Comments of CTIA–The Wireless Association, Service Rules for Advanced Wireless Services in 1915-1920 MHz, 1995-2000 MHz, 2020-2025 MHz and 2175-2180 MHz Bands, FCC WT Docket No. 04-356 (rel. Dec. 9, 2004).

81. *See generally* Venkatesh Krishnan, Vivek Rajendran, Ananthapadmanabhan Kandhadai & Sharath Manjunath, *EVRC-Wideband: The New 3GPP2 Wideband Vocoder Standard, in* 2 IEEE INTERNATIONAL CONFERENCE ON COMMUNICATIONS 333 (2007).

network operator can compensate for delays in network expansion, such as might be caused by difficulty obtaining the proper zoning for a new cell site or by extended bad weather. In an area of limited coverage—such as might develop after a brush fire destroyed the equipment at a cell site—the network could command subscriber handsets to reduce the network capacity each handset uses, thereby providing more capacity for others. For example, the industry claims that the SMV vocoder increases system capacity by thirty-four percent while delivering the same quality as the EVRC vocoder.

The GSM world has a similar variable rate capability called the adaptive multirate (AMR) vocoder. It allows the wireless system to adjust the traffic generated by the handsets to better match the system capacity. Use of the AMR vocoder also permits a carrier to serve handsets at greater distance from a cell site or deeper inside office buildings than would otherwise be possible.

Closely related to the variable rate concept is the discontinuous transmission concept—the engineer's way of referring to handsets that turn off the transmitter when the user is in a conversation and is listening but not talking. Shutting off the handset transmitter in such situations not only extends battery life but reduces the interference that the handset generates to other users on the system.

Receiver sensitivity and vocoder performance are two handset attributes that directly substitute for network investment. Reduced receiver sensitivity reduces the transmission range from base stations, and requires more base stations for equivalent coverage. Vocoders that squeeze a conversation into half as many bits per second double the number of conversations that can fit into a wireless system—or cut in half the electronics required at the base station. Investments in improved receiver sensitivity and vocoder performance are direct substitutes for investment in network physical infrastructure.

3. Other Handset Attributes That Affect System Capacity

Handset sensitivity is not the only handset characteristic that affects the amount of system resources that a handset will consume. There are a number of handset attributes (including receiver sensitivity) that, if less than optimum, consume excessive system resources and thereby reduce the wireless system's capacity or coverage.

The first cellular technology used in the United States, AMPS, did not have the tight link between handset quality and system capacity that current systems exhibit.[82] Indeed, to a first approximation, in that early technology,

82. AMPS is an acronym for Advanced Mobile Phone Service—the name of the analog FC cellular standard first used in the U.S. Prior to 2002, the FCC required cellular carriers to

system capacity was independent of handset quality. Unlike modern CDMA and OFDMA systems that serve multiple subscribers from a single transmitter-receiver pair, those early systems used a separate transmitter and receiver for each conversation. Transmitting more power to one handset did not diminish the power available to other handsets.

Modern wireless handsets often support web browsers and other connections to the Internet. Many of the standard rules for communicating over the Internet were designed under the assumption that communications capacity was relatively plentiful and inexpensive—consequently, standard Internet communications often contain substantial redundancy. Recognizing that this assumption is not always appropriate, the Internet standards community developed add-on capabilities that permit more efficient use of the communications links at the expense of additional processing in the handset and the network. The most well known of these is *Van Jacobson TCP/IP header compression*, but there are several others.[83] Requiring these features in a handset lowers the handset's use of network resources.

4. Handset Attributes and Service Quality

Many of the capabilities or attributes of handsets affect not only the efficiency of the network, but also the quality of the service delivered to subscribers. For example, a handset with poor sensitivity loses calls at locations where a phone with better sensitivity could permit the conversation to continue Similarly, speech delivered by a handset with a poor voice coding subsystem (vocoder implementation) or a low-quality speaker does not sound as good as speech delivered by a higher-quality handset. Some handset impairments that harm other consumers or consume system resources have no direct negative impact on the user of the impaired handset.

5. Poor Handsets or Poor Networks?

Consumers are unable to distinguish between many handset limitations (such as poor sensitivity or weak uplink power) and related network limitations (such as poor coverage). The symptoms of these particular network and handset impairments are exactly the same—dropped calls, regions of poor or no service, and poor voice quality on a call. Because consumers cannot readily distinguish between network weakness and handset shortcomings, consumers with poor handsets may mistakenly blame service providers for the resulting poor service. Wireless carriers

---

support AMPS handsets. *See* 47 C.F.R. § 22.901.

83. V. Jacobson, *Compressing TCP/IP Headers for Low-Speed Serial Links*, IETF RFC 1144 (rel. Feb. 1990), http://www.rfc-editor.org/rfc/rfc1144.pdf.

concerned with protecting their reputation have an incentive to control the handsets used by their subscribers.

Wireless service is a new service—it is still in the process of rapid technical evolution. Furthermore, because of the rapid growth of the number of subscribers and their use of the service, wireless service providers are constantly building out and upgrading their networks. The wireless transmission facility—the radio paths to and from the base station—is created, in part, by the handset. Unlike the case with wired telephone service, the consumer cannot unplug the handset to test the line. With wireless, the handset and the wire are one and the same.

Handsets affect service quality in another way, as well. Customers often call their wireless carrier for assistance with configuring their handsets or dealing with service features. A customer using a handset that the help-desk staff is not familiar with would pose unusual and difficult challenges, especially if the customer were trying to use one of the less-common features.

6. Network Standards Evolution

Wireless service providers in the United States have used multiple standards—AMPS, TDMA, CDMA, GSM, WCDMA, and cdma2000—and have had to transition their systems from one standard to another. All U.S. wireless carriers continuously face such standards transitions—the problem is the need to manage the transition from one generation of technology to the next. All cellular carriers had to shift from analog to digital. Today, wireless carriers face the problem of moving from second-generation systems (GSM, CDMA) to third-generation systems (UMTS, cdma2000) and now confront the transition to fourth-generation systems. Providing customers with a mix of dual-mode handsets is an important tool in such a transition.[84]

Note that individual consumers have no incentive to buy new-technology handsets—the service delivered to new-technology and old-technology handsets is exactly the same. If it is the case that (1) the adoption of new-technology base stations and handsets is the efficient way

84. It should be noted that some nations do not permit wireless carriers to move from one generation of technology to the next within their licensed spectrum. Rather, carriers in a specific band are locked into a specific technology. See Telefonica O2 UK Unlimited v. Office of Comm., [2010] CAT 25 (Eng.), http://www.catribunal.org.uk/files/1154_Telefonica_Judgments_071010.pdf, for a statement of the U.K. policy limiting technology in the bands used for GSM. The more rigidly a nation controls the technology used in wireless, the weaker the arguments for carrier control of handsets used with the carrier's network become. At the same time, such rigid controls undercut the innovation process. It should be no surprise that the CDMA technology underlying all 3G system designs was developed under the flexible regulatory regime in the United States.

to expand network capacity and (2) new-technology handsets are more expensive than old-technology handsets, the efficient network/handset choice will not be made unless the carrier provides an incentive to consumers to use the more efficient handset technology. The usual theory of congestion pricing teaches that service price is one such incentive—the carrier could offer discounts to users who used the new-technology handsets in locations served by new-technology base stations during peak times. Unfortunately, such pricing would run directly counter to consumer preferences for simple price schedules.[85] Another approach is for the carrier to subsidize the sale of new-technology handsets to those who are likely to make calls in areas served by the new-technology base stations. Tying and handset subsidies are good tools for ensuring rapid consumer adoption of new-technology handsets.

IV. SCHEDULING AND PRIORITY ROUTING IN SATELLITES, ELECTRICITY, AND WIRELESS

It may be instructive to consider how our economy copes with congestion and capacity limits in other services. Nature has imposed similar random fluctuations on the capacity of other types of important services. The capacity of some geostationary communications satellites comes in physical units called transponders. A satellite might have twenty-four transponders. Satellite providers often sell the capacity of an entire transponder to a customer. Unfortunately, transponders are like computers or refrigerators—they can work fine for months or years and then unexpectedly fail. Satellite carriers and satellite users have a good idea of the probability of these failures. Thus, at the time that a twenty-four-transponder satellite is launched, a planner might expect that five years later there would be a 100 percent chance that the satellite would have twenty or more working transponders, a fifty percent chance of having twenty-two or more working transponders, and a ten percent chance of having all twenty-four transponders working.

As is the case for the wireless channels described above, the capacity of a satellite varies randomly. The satellite industry deals with this uncertainty by offering three types of transponder services—protected, unprotected, and preemptible. *Protected service* provides the highest reliability. If a protected transponder fails, the user's traffic is transferred to a different transponder that is still working. *Unprotected service* provides less reliability but costs less. If an unprotected transponder fails, the user is out of luck—the user loses the satellite link through that transponder.

85. *See* ANDREW ODLYZKO, AT&T LABS, INTERNET PRICING AND THE HISTORY OF COMMUNICATIONS (2001), http://www.dtc.umn.edu/~odlyzko/doc/history.communications1b.pdf.

Preemptible service provides the least reliability. When a protected transponder fails, a user of a preemptible transponder may see service terminated in order to free up a transponder for the user with protected service. If there were a rule that all satellite transponders had to be offered on the same terms, then either (1) a user who needed highly reliable service, say a TV programming service, would need to rent multiple transponders in order to ensure access to backup capacity, or (2) the satellite operator would need to keep the backup transponders idle. Giving some transponder users priority over others increases the total value delivered by the satellite system. Moreover, it makes available to users several price/service quality options.

Electrical power systems also have uncertain capacity because generators fail, transmission lines fail, river flows vary, and the wind is stronger at some times than at others. Naturally enough, wholesale electric power producers sell products such as firm power and interruptible power.[86] Interruptible power would be unacceptable for most homes and businesses. However, some commercial uses of electricity, such as refining aluminum or pumping water for irrigation, can be operated efficiently on interruptible power.

A wireless system engineered to support human conversation may have no more capacity for telephone calls but may still have capacity to carry delay-tolerant packets. Because some Internet applications are far more tolerant of delay than are human conversations, this additional capacity can be used to deliver useful service to consumers. A rule prohibiting any differential treatment of packets—that is, that no priority be afforded to one class of packets over another—would block consumer access to this additional capacity and prevent the efficient use of the radio spectrum and of the base stations and radios used to communicate across that spectrum.

Demand variations create essentially identical concerns in the wireline and wireless worlds. For example, it is well known that when multiple users go online at the same time—such as when kids leave school in the afternoon—the resulting congestion can affect the latency and jitter experienced by cable modem users competing for the finite and shared

86. *See Glossary of Terms Used in Subscription Power Product Descriptions*, BONNEVILLE POWER ADMIN. (Nov. 5, 1997), http://www.bpa.gov/power/pl/ subscription/ prodglos.htm. The power industry also faces variations in demand and offers a variety of user-pricing mechanisms designed to limit peak demand or to move demand from peak to off-peak times. The application of congestion pricing to energy through Advanced Metering Infrastructure is a key part of the Department of Energy's Smart Grid policy. *See The Smart Grid: An Introduction*, DEPT. OF ENERGY, http://www.oe.energy.gov/SmartGridIntroduction.htm (click on any graphic for more information) (last visited Feb. 21, 2011).

resource. In that context as well, approaches that differentiate between latency-sensitive traffic and other traffic could yield substantial consumer benefits and enable services that otherwise might not function well or at all at times of congestion.

V. CONCLUSION

Priority-enforcing technologies offer the opportunity to combine all communications on a single broadband link to the Internet.[87] In contrast, any prohibition on priority routing would steer traffic away from smaller service providers that operate only one network. For example, a hospital cannot use the Internet for latency-sensitive traffic, such as a medical monitoring service, if it must live with the threat that another user's rogue application can seriously degrade or cut off service.[88] Rather, a hospital would need to purchase dedicated connections from a provider able to provide such service on a network separate from the public Internet.

Any form of network regulation that prohibits priority routing or other approaches to assuring service quality would make it necessary for the United States to have multiple networks for voice, high-priority data, and general Internet data. The requirement to connect to and use multiple networks may not be a significant burden for a large corporation in an office building in Manhattan—fiber runs to the basement of the building, and the organization has sufficient scale to operate three networks efficiently. Smaller organizations, however, would face proportionately larger costs to manage the multiple networks and pay the various fixed costs. The development of applications that require high-quality network service would be handicapped, as such applications would perform better on dedicated networks than over the public Internet. Aggressive but delay-tolerant applications would thrive, and latency-sensitive applications would stumble along. In such cases, regulation and the physics of networks rather than consumer preferences would determine which firms and applications succeed in the market.

There is no simple rule that can identify when priority routing should be applied or to which flows it should be applied. In the above discussions of priority in wireless and of cross-layer design, this Article provided examples of well-accepted practices that give preferential processing to one

87. Larry Roberts, one of the true pioneers of the Internet, described the benefits from improved routing in a seminar at Stanford in 2009, saying, "[R]ecent improvement in flow technology . . . maintains information for each active flow, insures [sic] quality voice/video, allows utilization in the 95% region, and maintains unprecedented fairness." Seminar Announcement, Lawrence G. Roberts, Upgrading the Internet with Flow Technology (Jan. 17, 2008), http://netseminar.stanford.edu/seminars/01_17_08.html.

88. Recall that the BitTorrent white paper said that BitTorrent software does exactly this at times. *See* Shalunov, *supra* note 21.

category of packet over another, effectively expanding capacity and improving efficiency in the use of a limited resource. As discussed above, a careful analysis of the nature of the application and of the higher-level protocols permits doing more with the limited resources of broadband networks.

Likewise, consistent with widely accepted practices, differentiation among packets can combat the real problem of congestion. Congestion was a severe problem in the Internet in the mid-1980s. The solution to that congestion was the adoption of improved versions of TCP that incorporated congestion control. Unfortunately, this is congestion control on the honor system. Some current web browsers and peer-to-peer applications bend or break the honor system, permitting them to deliver better service to their users but at the expense of more congestion for other users. No simple rule regarding priority for one class of packets can encompass this complexity.

Congestion can also arise from network equipment failures, software features, and malicious software. This Article described four recent incidents of such congestion failures, though there were likely many more that went unpublicized.[89] In three of these examples, the ability of networks to manage congestion-causing traffic permitted most uses of the network to continue in a close-to-normal fashion.[90] Consumers benefit if networks have these capacities during times of congestion, whether that congestion is caused by normal patterns of use, hardware failures, software failures, or malicious software.

Although this Article has focused on technical issues—such as how priority scheduling expands wireless capacity or how packet inspection limits denial-of-service attacks—one should remember that there is also an economic argument for priority. Just as it makes sense to give an ambulance priority over commuters' cars, it makes sense to give packets carrying VoIP 911 calls priority over packets carrying music downloads.

Although some have urged the adoption of policies that would prohibit service providers from distinguishing between packets or ever favoring one packet over another, their analysis was silent on the many costs and unintended consequences that this policy would impose.[91] Indeed,

89. See the anomaly case studies list at SLAC for a few examples. *Case Studies for Wide Area Network Problems*, INTERNET END-TO-END PERFORMANCE MONITORING, https://confluence.slac.stanford.edu/display/IEPM/Case+Studies+for+Wide+Area+Network+Problems (last visited Feb. 21, 2011).

90. I have not seen any account of the countermeasures used for the July 4, 2009 cyberattacks.

91. *See, e.g.*, Reply Comments of Center for Media Justice, Consumers Union, Media Access Project, and New America Foundation, Preserving the Open Internet Broadband Industry Practices, FCC GN Docket No. 09-191 (rel. Apr. 26, 2010).

some essentially argued that it would impose no costs.[92] But, as the above discussion shows, it is difficult to conceive that an informed engineer or economist would consider priority scheduling of packets to be a zero-sum game. Today, ISPs, wireless carriers, and private networks use a variety of technologies to defend networks against malicious traffic and to give priority to traffic that is sensitive to delay or jitter. Prohibiting or restricting such technologies would harm consumers and pose risks to the economy and to public safety. Perhaps worst of all, it would hamper innovation and create artificial incentives to have multiple, fragmented networks.

Phrases like *net neutrality* and *cellular Carterfone* sound good—neutrality has positive connotations and it is widely accepted that the FCC's *Carterfone* decision served consumers well.[93] However, such concepts have to be reviewed carefully, as artful coinage of terms may mislead about their ultimate impacts on consumers. Many who have opposed any form of congestion control or priority-routing mechanism that would favor one class of packets over another or otherwise differentiate between packets have failed to identify or discuss the many costs that would flow from adopting such a policy. Net neutrality—whether wired or wireless—would impose substantial costs on consumers. Such policies should not be adopted without understanding and acknowledging such costs.

92. For example, in BEN SCOTT, MARK COOPER & JEANNINE KENNEY, WHY CONSUMERS DEMAND INTERNET FREEDOM 4 (2006), http://www.freepress.net/files/nn_fact_v_fiction_final.pdf, the authors state: "But network prioritization is a zero-sum game. The fact is that every time one Web site is sped up, another must be slowed down." But, of course, that assertion is only true if all network traffic is equally time sensitive.

93. *See* Use of the Carterfone Device in Message Toll Telephone Service, *Decision*, 13 F.C.C.2d 420 (1968). It is less well recalled that that FCC decision did not occur until well after the D.C. Circuit Court of Appeals had made it clear in its 1956 *Hush-A-Phone* decision that the law required the FCC to follow the basics of *Carterfone*. *See* Hush-A-Phone Corp. v. United States, 238 F.2d 266 (D.C. Cir. 1956).