WIRELESS NETWORK NEUTRALITY: TECHNOLOGICAL CHALLENGES AND POLICY IMPLICATIONS

Christopher S. Yoo[†]

ABSTRACT

One key aspect of the debate over network neutrality has been whether and how network neutrality should apply to wireless networks. The existing commentary has focused on the economics of wireless network neutrality, but to date a detailed analysis of how the technical aspects of wireless networks affect the implementation of network neutrality has yet to appear in the literature. As an initial matter, bad handoffs, local congestion, and the physics of wave propagation make wireless broadband networks significantly less reliable than fixed broadband networks. These technical differences require the network to manage dropped packets and congestion in a way that contradicts some of the basic principles underlying the Internet. Wireless devices also tend to be more heterogeneous and more tightly integrated into the network than fixed-line devices, which can lead wireless networks to incorporate principles that differ from the traditional Internet architecture. Mobility also makes routing and security much harder to manage, and many of the solutions create inefficiencies. These differences underscore the need for a regulatory regime that permits that gives wireless networks the flexibility to deviate from the existing architecture in ways, even when those deviations exist in uneasy tension with network neutrality.

DOI: https://dx.doi.org/10.15779/Z38HQ3RZ0S

[†] John H. Chestnut Professor of Law, Communication, and Computer & Information Science and Founding Director of the Center for Technology, Innovation and Competition, University of Pennsylvania. I would like to thank the New York Bar Foundation for its financial support. This paper benefited from presentations at ETH Zurich, University of Haifa, and the Free State Foundation. Responsibility for any errors remains with the author.

TABLE OF CONTENTS

I. INTRODUCTION

For the past decade, a single issue dominated Internet policy debates: network neutrality. The perceived need to protect network neutrality led the Federal Communications Commission (FCC) to adopt its first Open Internet Order in 2010 only to see that order overturned on judicial review in 2014.[1] The FCC issued its second Open Internet Order in 2015, which was upheld by the courts the following year.[2] On April 27 2017, the FCC announced its agenda for its May 18 Open Meeting, which included a Notice of Proposed Rulemaking that would revisit most of the key provisions of the second Open Internet Order.[3] The debate over network neutrality appears to be far from over.

Although myriad definitions of network neutrality exist,[4] they share a general commitment to preventing network providers (such as Verizon and Comcast) that offer broadband access to end users from discriminating against traffic based on its source, destination, or content, or based on its associated application, service, or device. From this point of view, all application-specific intelligence and functionality should be confined to the computers operating at the edge of the network, while the routers operating in the core of the network should be kept as simple as possible.

[1] Preserving the Open Internet, Report and Order, 25 FCC Rcd. 17905 (2010) [hereinafter 2010 Open Internet Order], *aff'd in part, vacated in part sub nom.* Verizon v. FCC, 740 F.3d 623 (D.C. Cir. 2014).

[2] Protecting and Promoting the Open Internet, Report and Order on Remand, Declaratory Ruling, and Order, 30 FCC Rcd. 5601 (2015) [hereinafter 2015 Open Internet Order], *aff'd sub nom.* U.S. Telecom Ass'n v. FCC, 825 F.3d 674 (D.C. Cir. 2016).

[3] Restoring Internet Freedom, Notice of Proposed Rulemaking, WC Docket No. 17-108 (FCC Apr. 27, 2017) [hereinafter 2017 Open Internet NPRM], https://apps.fcc.gov/edocs_public/attachmatch/DOC-344614A1.pdf.

[4] *See, e.g.*, Rachelle B. Chong, *The 31 Flavors of Net Neutrality: A Policymaker's View*, 12 INTELL. PROP. L. BULL. 147, 151–55 (2008) (identifying five distinct versions of network neutrality); Eli Noam, *A Third Way for Net Neutrality*, FIN. TIMES (Aug. 29, 2006, 5:26 PM), http://www.ft.com/cms/s/2/acf14410-3776-11db-bc01-0000779e2340.html (identifying seven distinct versions of network neutrality).

Designing the network in this manner is often regarded as essential to ensuring that the network remains open to all applications.[5]

One central issue in both Open Internet Orders was whether mobile broadband should be subject to less restrictive rules than fixed broadband. Specifically, the 2010 Open Internet Order adopted three rules, but restricted the application of one of the rules to mobile broadband and completely exempted mobile broadband from another rule.[6] The 2015 Open Internet Order took a different approach, choosing to apply the same rules to both fixed and mobile broadband. At the same time, the 2015 Order repeatedly recognized the existence of key technical differences between fixed and mobile broadband that must be considered when determining whether a particular network management practice is permissible.[7] The 2017 NPRM reopened this issue by "seek[ing] comment on whether mobile broadband should be treated differently from fixed broadband."[8]

Both orders explicitly suggest that technical dissimilarities might justify the use of network management practices on mobile broadband networks that would not be allowed on fixed broadband networks. Indeed, the 2015 Order requires that regulators grapple with the technical details when determining whether a particular practice violates its terms. Unfortunately, the technical aspects of mobile broadband have gone largely unexplored. So far, the academic commentary has focused almost exclusively on the economics of wireless network neutrality, debating whether wireless broadband providers have the economic means and incentive to restrict traffic from certain sources or applications in ways

[5] For the FCC's most extensive elaboration of this rationale, see Preserving the Open Internet, Notice of Proposed Rulemaking, 24 FCC Rcd. 13,064, 13,070 ¶ 19, 13,086 ¶ 56, 13,088–89 ¶ 63 (2009) [hereinafter 2009 Open Internet NPRM]. For subsequent restatements embracing this principle, see 2015 Open Internet Order, *supra* note 2, at 5702 n.570; Protecting and Promoting the Open Internet, Notice of Proposed Rulemaking, 29 FCC Rcd. 5561, 5629 ¶ 8, 5597 ¶ 102 & n.226, 5702 n. 570, 5803 ¶ 431 (2014); Preserving the Open Internet, Broadband Industry Practices, Report and Order, 25 FCC Rcd. 17,905, 17,909–10 ¶ 13 & nn.13–14 (2010) [hereinafter 2010 Open Internet Order].

[6] *See* 2010 Open Internet Order, *supra* note 5, at 17,956–62 ¶¶ 93–106; *Net Neutrality: Hearing Before the S. Comm. on Commerce, Sci. & Transp.*, 109th Cong. 9 (2006) (prepared statement of Vinton G. Cerf, Vice Pres. & Chief Internet Evangelist, Google Inc.) ("The remarkable success of the Internet can be traced to a few simple network principles—end-to-end design, layered architecture, and open standards . . . .").

[7] 2015 Open Internet Order, *supra* note 2, at 5611 ¶ 34, 5643 ¶ 101, 5651 ¶ 118, 5665 ¶ 148, 5701 ¶ 218, 5703–04 ¶ 223.

[8] 2017 Open Internet NPRM, *supra* note 3, at 30 ¶ 94.

that could harm consumers and innovation.[9] While one can debate the economic merits of prohibiting discrimination and prioritization, to date the literature has not grappled with the technical challenges that wireless broadband providers face in managing their networks.

An examination of the way wireless broadband networks actually work at a technical level is thus essential to understanding how network neutrality should be applied to mobile broadband. As discussed further below, differences in the ways that wireless broadband networks manage congestion and reliability necessarily introduce far more intelligence into the core of the network than is the case with fixed broadband networks. Moreover, mobile broadband networks are subject to bandwidth constraints that are much more restrictive than those faced by fixed broadband networks, and mobile operators choose to mitigate congestion by treating traffic differently depending on the applications with which it is associated. Indeed, the engineering literature is replete with observations listing support for mobility as one of the key network functions that the current architecture fails to perform well.[10] The National Science Foundation's Future Internet Architecture program is sponsoring a project to explore how the Internet might need to be redesigned to accommodate mobile broadband.[11]

[9] The debate over how to apply network neutrality to mobile broadband networks was initiated by Tim Wu. *See* Tim Wu, *Wireless* Carterfone, 1 INT'L J. ON COMM. 389 (2007). For later discussions, see Babette E.L. Boliek, *Wireless Net Neutrality Regulation and the Problem with Pricing: An Empirical, Cautionary Tale*, 16 MICH. TELECOMM. TECH. L. REV. 1 (2009); George S. Ford, Thomas M. Koutsky, & Lawrence J. Spiwak, *A Policy and Economic Exploration of Wireless* Carterfone *Regulation*, 25 SANTA CLARA COMPUT. & HIGH TECH. L.J. 647 (2008); Rob Frieden, *Hold the Phone: Assessing the Rights if Wireless Handset Owners and Carriers*, 69 U. PITT. L. REV. 675 (2008); Robert W. Hahn, Robert E. Litan & Hal J. Singer, *The Economics of Wireless Net Neutrality*, 3 J. COMPETITION L. & ECON. 399 (2007); Gregory L. Rosston & Michael D. Topper, *An Antitrust Analysis of the Case for Wireless Network Neutrality*, 22 INFO. ECON. & POL'Y 103 (2010); Marius Schwartz & Federico Mini, *Hanging Up on* Carterfone*: The Economic Case Against Access Regulation in Mobile Wireless* (May 2, 2007) (unpublished manuscript), http://ssrn.com/abstract=984240.

[10] *See, e.g.*, Mark Handley, *Why the Internet Only Just Works*, 24 BT TECH. J. 119, 120 (2006); Raj Jain, *Internet 3.0: Ten Problems with Current Internet Architecture and Solutions for the Next Generation*, PROC. MIL. COMM. CONF. (MILCOM 2006) (2007), http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4086425; Jon Crowcroft, *Net Neutrality: The Technical Side of the Debate*, ACM SIGCOMM COMPUTER COMM. REV., Jan. 2007, at 49, 51; Thrasyvoulos Spyropoulos et al., *Future Internet: Fundamentals and Measurement*, ACM SIGCOMM COMPUTER COMM. REV., Apr. 2007, at 101; Sixto Ortiz, Jr., *Internet Researchers Look to Wipe the Slate Clean*, COMPUTER, Jan. 2008, at 12.

[11] MobilityFirst Future Internet Architecture Project, *Overview*, MOBILITYFIRST http://mobilityfirst.winlab.rutgers.edu/ (last visited Feb. 9, 2016).

Many of the ways that wireless broadband networks operate differently from fixed broadband networks involve explicit prioritization of certain types of applications. Other aspects of wireless broadband networks violate many central tenets of the Internet's architecture, either by changing the semantics or by changing the basic principles around which the Internet is currently designed. Such changes reduce the interoperability of the network and create a much tighter integration between end users and the network. Even less transformative proposals are likely to affect different applications and end users differently and inevitably cause traffic on wireless and wireline networks to behave in a strikingly different manner. Understanding the technical space is thus essential to understanding how and when differential regulatory treatment between wireline and wireless networks may be justified and determining how the exception for reasonable network management should be applied to wireless networks.

The balance of this Article is organized as follows: Part II lays out the relevant regulatory history. Part III explains the basic architectural principles generally associated with the Internet, specifically nondiscrimination and the end-to-end argument. Part IV describes the more restrictive bandwidth constraints that mobile broadband networks face. Part V discusses quality of service. Part VI examines the heterogeneity of devices, and Part VII addresses the additional complexity of routing.

II. THE FCC'S SPECIAL TREATMENT OF MOBILE BROADBAND

The FCC's attempts to mandate network neutrality have consistently recognized that mobile broadband faces greater challenges than fixed broadband. Indeed, these differences have led the FCC to apply fewer restrictions to mobile broadband than to fixed broadband.

A. THE BASIC REGULATORY REGIME GOVERNING COMMUNICATIONS

The basic structure of the laws governing U.S. communications technologies was established by the Communications Act of 1934 ("1934 Act") and has remained largely unchanged ever since.[12] Title II of the 1934 Act governs telecommunications services, which have historically consisted primarily of traditional telephone service provided via fixed-line technologies. Under Title II, telecommunications carriers are subject to

[12] Communications Act of 1934, Ch. 652, 48 Stat. 1064 (codified as amended at 47 U.S.C. §§ 151, 202, 212, 311, 313, 314, 316, 317, 506, 521, 543 (2012)).

common carriage regulation,[13] which requires that they provide service to anyone who requests it on terms that are just, reasonable, and nondiscriminatory.[14] A subsequent amendment to Title II authorizes the FCC to use a process known as "forbearance" to excuse telecommunications carriers from having to comply with any regulations that the FCC finds are not necessary to protect consumers.[15]

Title III of the 1934 Act governs spectrum-based communications, which initially consisted solely of radio and television broadcasting transmitted over the air. A provision of the 1934 Act prohibited broadcasting from being treated as common carriers.[16] In *FCC v. Midwest Video Corp.* (*Midwest Video II*), the Supreme Court held that this statutory provision prohibited the FCC from requiring any service regulated under the broadcasting statute from making its facilities available on a nondiscriminatory basis.[17]

The emergence of cellular telephony upset this tidy regulatory taxonomy by making it possible to provide telephone service via spectrum. In response, Congress amended Title III to permit regulating spectrum-based communications technologies as common carriers only if they constituted Commercial Mobile Services (CMS). A CMS is any mobile service that makes interconnected services available to the public.[18] All other services are Private Mobile Services (PMS), which are exempt from common carriage regulation.[19]

The emergence of new services that combined the transmission associated with telephone service with the data processing and storage associated with modern computing raised the question of whether and how these technologies should be regulated. From the time these new services first emerged, the FCC consistently exempted them from most regulation.[20] As then-FCC Chairman William Kennard could observe in 1999, "[f]or the past 30 years, the FCC has created a deregulatory environment in which the Internet could flourish."[21] That said, the FCC

[13] 47 U.S.C. § 153(51) (1996).

[14] 47 U.S.C. § 202(a) (1989).

[15] 47 U.S.C. § 160(a) (1996).

[16] 47 U.S.C. § 153(11) (2010).

[17] 440 U.S. 689, 700–02, 707 (1979).

[18] 47 U.S.C. § 332(c)(1)(A) (1996).

[19] 47 U.S.C. § 332(c)(2) (1996).

[20] MTS and WATS Market Structure, Access Charge Reconsideration Order, 97 F.C.C.2d 682, 711–22 (1983).

[21] William E. Kennard, Chairman, Fed. Commc'ns Comm'n, Remarks Before the Federal Communications Bar, Northern California Chapter: The Unregulation of the

tried to avoid directly addressing the proper regulatory classification that would apply to broadband Internet access, which drew a sharp rebuke from two members of the Supreme Court in January 2002.[22] Finally, in March 2002, the FCC ruled that cable modem service was not a Title II service.[23]

The modern debate over network neutrality emerged in 2004, when a speech by FCC Chairman Michael Powell challenged the industry to preserve four "Internet freedoms."[24] The first three freedoms called for allowing consumers to access legal content, run applications, and attach devices as they saw fit, while the fourth held that consumers should receive meaningful information about their service plans.[25]

The Supreme Court's *Brand X* decision eliminated any uncertainty about the propriety of the FCC's 2002 decision regarding the regulatory classification of cable modem systems discussed above when it upheld the FCC's ruling that the Internet was not a Title II service.[26] Although the Supreme Court noted that the FCC had not yet decided whether to impose any specific regulatory obligations on cable modem systems,[27] most observers believed that broadband Internet access services would not be subject to open access obligations.[28] Shortly thereafter, the FCC issued

Internet: Laying a Competitive Course for the Future 2 (July 20, 1999), https://transition.fcc.gov/Speeches/Kennard/spwek924.doc.

[22] Nat'l Cable & Telecomms. Ass'n v. Gulf Power Co., 534 U.S. 327, 348–49, 353–56 & n.5 (2002) (Thomas, J., joined by Souter, J., concurring in part and dissenting in part).

[23] Inquiry Concerning High-Speed Access to the Internet over Cable and Other Facilities, Declaratory Ruling and Notice of Proposed Rulemaking, 17 FCC Rcd. 4798 (2002), *aff'd sub nom.* Nat'l Cable & Telecomm. Ass'n v. Brand X Internet Servs., 545 U.S. 967 (2005).

[24] Michael K. Powell, Chairman, Fed. Commc'ns Comm'n, Remarks on Preserving Internet Freedom: Guiding Principles for the Industry Delivered at the Silicon Flatirons Symposium 5–6 (Feb. 8, 2004), http://hraunfoss.fcc.gov/edocs_public/attachmatch/DOC-243556A1.pdf. For an earlier discussion, see Christopher S. Yoo, *Network Neutrality and the Economics of Congestion*, 94 GEO. L.J. 1847, 1857 (2006). The FCC considered the related issue of whether it should mandate open access to cable modem systems when clearing a series of cable industry mergers from 1999 to 2002. *See* Daniel F. Spulber & Christopher S. Yoo, *Access to Networks: Economic and Constitutional Connections*, 88 CORNELL L. REV. 885, 1015–18 (2003).

[25] Powell, *supra* note 24, at 5.

[26] Nat'l Cable & Telecomm. Ass'n v. Brand X Internet Servs., 545 U.S. 967 (2005).

[27] *Id.* at 996.

[28] *See, e.g.*, John Blevins, *A Fragile Foundation — The Role of "Intermodal" and "Facilities-Based" Competition in Communications Policy*, 60 ALA. L. REV. 241, 279 n.155 (2009) ("In practice, . . . Title I 'regulation' is essentially deregulation.").

decisions ruling that broadband access provided by telephone companies and mobile providers were also not Title II services.[29]

In 2005, the FCC issued a Policy Statement adopting four principles that echoed the four "Internet Freedoms" advanced in Powell's speech.[30] The FCC's first three principles mirrored Powell's first three freedoms, albeit subject to some caveats.[31] The FCC's Policy Statement also replaced Powell's transparency principle with an "entitle[ment] to competition among network providers, application and service providers, and content providers."[32]

The FCC was explicit that its Policy Statement was not a new set of rules. According to the FCC, the Policy Statement simply indicated its intention to "incorporate the above principles into its ongoing policymaking activities."[33] FCC Chairman Kevin Martin released a concurrent statement recognizing that "policy statements do not establish rules nor are they enforceable documents" and expressing his confidence "that the marketplace will continue to ensure that these principles are maintained" and "therefore, that regulation is not, nor will be, required."[34] Despite these concessions, the FCC invoked the Policy Statement as the basis for sanctioning Comcast for its use of Transmission Control Protocol (TCP) resets to slow down traffic generated by certain peer-to-peer file sharing applications in 2008.[35]

Because wireless had not yet emerged as an important broadband platform, Chairman Powell's four freedoms and the 2005 Policy Statement did not draw any distinctions between different broadband technologies.

[29] *See* Appropriate Framework for Broadband Access to the Internet over Wireline Facilities, Report and Order and Notice of Proposed Rulemaking, 20 FCC Rcd. 14,853 (2005), *petition for review denied sub nom.* Time Warner Telecom, Inc. v. FCC, 507 F.3d 205 (3d Cir. 2007); Appropriate Regulatory Treatment for Broadband Access to the Internet over Wireless Networks, Declaratory Ruling, 22 FCC Rcd. 5901 (2007).

[30] Policy Statement on Appropriate Framework for Broadband Access to the Internet over Wireline Facilities, Policy Statement, 20 FCC Rcd. 14,986, 14,988 (2005).

[31] Specifically, the Policy Statement made the right to access applications "subject to the needs of law enforcement." *Id.* It also limited the right to connect devices to "legal devices that do not harm the network." *Id.* All of the principles were "subject to reasonable network management." *Id.* at 14,988 n.15.

[32] *Id.* at 14,988.

[33] *Id.* at 14,988 & n.15.

[34] Kevin J. Martin, Chairman, Fed. Commc'ns Comm'n, Comments on Commission Policy Statement (Aug. 5, 2005), http://hraunfoss.fcc.gov/edocs_public/attachmatch/DOC-260435A2.pdf.

[35] Formal Complaint of Free Press and Public Knowledge Against Comcast Corporation for Secretly Degrading Peer-to-Peer Applications, Memorandum Opinion and Order, 23 FCC Rcd. 13,028, 13,050–58 ¶¶ 141–151 (2008), *rev'd sub nom.* Comcast Corp. v. FCC, 600 F.3d 642, 644 (D.C. Cir. 2010).

The impetus to apply less restrictive network neutrality regulations to mobile broadband did not emerge until the proceedings that led to the 2010 Open Internet Order.

B. THE 2010 OPEN INTERNET ORDER

The first recognition that mobile broadband might receive separate treatment appeared in the 2009 Notice of Proposed Rulemaking (NPRM) that led to the 2010 Open Internet Order.[36] The NPRM proposed codifying the four principles included in the 2005 Policy Statement, augmented by new rules that prohibited discrimination and required transparency.[37] The NPRM also included an exception for reasonable network management.[38] The NPRM explicitly sought comment on whether nondiscrimination and reasonable network management might apply differently to mobile broadband.[39]

The 2009 NPRM proved controversial from the outset.[40] It became even more so in April 2010, when the D.C. Circuit overturned the FCC's order sanctioning Comcast not because the FCC had failed to adopt formal rules, but rather because the FCC had failed to base its actions on any valid statutory grant of authority.[41]

Uncertainty about the FCC's jurisdiction over network neutrality led then FCC Chairman Julius Genachowski to float a proposal on May 6, 2010 that would have reversed the 2002 Cable Modem Declaratory Ruling upheld in *Brand X* and would have reclassified Internet access as a Title II service, thereby bringing the Internet within the regulatory regime that governs traditional telephone service.[42] Under the proposal, the FCC would also exercise its statutory forbearance authority to excuse broadband Internet access providers from all but six of the relevant statutory provisions.[43]

Genachowski's reclassification proposal proved even more controversial than the 2009 NPRM. On May 24, 2010, seventy-four House

[36] 2009 Open Internet NPRM, *supra* note 5.

[37] *Id.* at 13,100–11 ¶¶ 88–132.

[38] *Id.* at 13,112–15 ¶¶ 133–141.

[39] *Id.* at 13,123–24 ¶¶ 171–174.

[40] *See* Wendy Davis, *Controversy Continues as FCC Votes Unanimously to Consider Net Neutrality Rules*, MEDIA POST (Oct. 22, 2009, 6:21 PM), http://www.mediapost.com/publications/article/115959/controversy-continues-as-fcc-votes-unanimously-to.html.

[41] *See* Comcast Corp. v. FCC, 600 F.3d 642, 644 (D.C. Cir. 2010).

[42] Julius Genachowski, Chairman, Fed. Commc'ns Comm'n, The Third Way: A Narrowly Tailored Broadband Framework 5–6 (May 6, 2010), https://apps.fcc.gov/edocs_public/attachmatch/DOC-297944A1.pdf.

[43] *Id.*

Democrats signed a letter urging Genachowski not to reclassify broadband Internet access as a Title II service, warning that it would "jeopardize jobs" and "should not be done without additional direction from Congress."[44] Thirty-seven House Republicans filed a similar letter the same day.[45]

Undeterred, the FCC adopted a Notice of Inquiry on June 17, 2010, seeking comment on the possibility of reclassifying the Internet as a Title II service, again over the objections of the two Republican Commissioners.[46] Over the summer, the FCC convened a series of closed-door meetings attempting to find common ground among the key industry players.[47] Concurrently, reports began to emerge that Google and Verizon were on the verge of announcing a new joint position on network neutrality.[48] Rumors of the impending agreement caused the FCC to end its efforts to broker a compromise.[49]

Google and Verizon unveiled their joint proposal on August 10, 2010.[50] The joint proposal endorsed the FCC's vision of creating rules embodying the first three principles of the 2005 Policy Statement as well as the new rules mandating nondiscrimination and transparency.[51] More importantly for our purposes, it provided a ringing endorsement of subjecting mobile broadband to less stringent regulation. Only the transparency principle would apply to mobile broadband "[b]ecause of the unique technical and operational characteristics of wireless networks, and the competitive and still-developing nature of wireless broadband services."[52] On September 1, 2010, the FCC issued a further inquiry

[44] Declan McCullagh, *Congress Rebukes FCC on Net Neutrality Rules*, CNET (May 24, 2010, 9:46 PM), http://www.cnet.com/news/congress-rebukes-fcc-on-net-neutrality-rules/.

[45] *Id.*

[46] Framework for Broadband Internet Service, Notice of Inquiry, 25 FCC Rcd. 7866, 7889–95 ¶¶ 52–66 (2010).

[47] *See* Matthew Lasar, *A Peek Inside the "Secret, Backroom" Net Neutrality Meetings*, ARS TECHNICA (July 28, 2010, 3:56 PM), http://arstechnica.com/tech-policy/2010/07/fcc-secret-net-neutrality-meetings-continue-in-plain-sight/.

[48] *See* Edward Wyatt, *Web Deal Near on Paying Up to Get Priority*, N.Y. TIMES, Aug. 5, 2010, at A1.

[49] *See FCC Ends Net Neutrality Compromise Talks*, CBS NEWS (Aug. 5, 2010, 3:50 PM), http://www.cbsnews.com/news/fcc-ends-net-neutrality-compromise-talks/.

[50] *Verizon-Google Legislative Framework Proposal*, GOOGLE BLOG (Aug. 10, 2010), http://www.google.com/googleblogs/pdfs/verizon_google_legislative_framework_proposal_081010.pdf; *see also* Alan Davidson & Tom Tauke, *A Joint Policy Proposal for an Open Internet*, GOOGLE PUB. POL'Y BLOG (Aug. 9, 2010), http://googlepublicpolicy.blogspot.com/2010/08/joint-policy-proposal-for-open-internet.html.

[51] *Verizon-Google Legislative Framework Proposal*, *supra* note 50, at 1.

[52] *Id.* at 2.

seeking comment on how the proposed network neutrality rules should apply to mobile broadband in general and on the Google-Verizon joint proposal in particular.[53]

The idea of subjecting mobile broadband to less stringent regulation than fixed broadband became embodied in the Open Internet Order that the FCC adopted on December 23, 2010.[54] Consistent with the Google-Verizon joint proposal, the 2010 Order applied the transparency rule to mobile broadband, but refrained from applying the nondiscrimination rule to mobile broadband.[55] The 2010 Order did part with the Google-Verizon joint proposal in one respect, by subjecting mobile broadband to a modified no-blocking rule applicable only to websites and "applications that compete with the provider's voice or video telephony services."[56] The rules were subject to exceptions for reasonable network management and specialized services.[57] Regarding legal authority, the 2010 Order opted not to regulate under Title II, and instead asserted a welter of other statutory provisions.[58]

The FCC recognized that "mobile broadband presents special considerations that suggest differences in how and when open Internet protections should apply," specifically that mobile broadband represented an early-stage platform characterized more competition and greater operational constraints.[59] Chairman Genachowski echoed this reasoning, noting that key differences distinguished mobile broadband from fixed broadband, including "unique technical issues involving spectrum and mobile networks, the stage and rate of innovation in mobile broadband; and market structure."[60] The other two Democratic Commissioners expressed their wish that mobile broadband had been treated the same as fixed broadband, but nonetheless voted for the Order.[61] Network neutrality advocates were not so easily satisfied, bringing a number of challenges to the decision to apply a lighter touch to mobile broadband.[62]

[53] Further Inquiry into Two Under-Developed Issues in the Open Internet Proceeding, Public Notice, 25 FCC Rcd. 12,637, 12,640–42 (2010).

[54] 2010 Open Internet Order, *supra* note 5, at 17,958–62 ¶¶ 97–105.

[55] *Id.* at 17,958 ¶ 96, 17959 ¶ 98, 17,962 ¶ 104.

[56] *Id.* at 17,959 ¶ 99.

[57] *Id.* at 17,951–56 ¶¶ 80–92, 17,964–65 ¶¶ 112–114.

[58] *Id.* at 17,966–81 ¶¶ 115–137.

[59] *Id.* at 17,956–97 ¶¶ 94–95.

[60] *Id.* at 18,041.

[61] *Id.* at 18,046 (Copps, Comm'r., concurring), 18,082 (Clyburn, Comm'r., approving in part and concurring in part).

[62] *See, e.g.*, Free Press v. FCC, No. 11-2123 (1st Cir. filed Sept. 28, 2011); Mountain Area Info. Network v. FCC, No. 11-2036 (4th Cir. filed Sept. 27, 2011); People's Prod. House v. FCC, No. 11-3905 ag (2d Cir. filed Sept. 26, 2011); Media

The D.C. Circuit issued its decision resolving the various challenges to the 2010 Open Internet Order on January 14, 2014.[63] The court held that the FCC had the authority to regulate broadband Internet access under Section 706 of the Telecommunications Act of 1996,[64] but ruled that the FCC could not exercise that authority in a manner inconsistent with any other express statutory provisions, such as the section providing, "A telecommunications carrier shall be treated as a common carrier under this [Act] only to the extent that it is engaged in providing telecommunications services."[65] Nondiscrimination is the hallmark of common carriage regulation (indeed the FCC explicitly equated its nondiscrimination rule with the nondiscrimination contained in Title II),[66] and the Communications Act prohibits the FCC from regulating any provider as a common carrier unless it were classified as a Title II provider—a step the FCC specifically declined to take with respect to broadband Internet access.[67] The court recognized that it had previously held that another access regulation mandating access on "commercially reasonable" terms did not constitute nondiscrimination because the regulation left "'substantial room for individualized bargaining and discrimination in terms.'"[68] The FCC's reliance on the same rationale for both the nondiscrimination and the no-blocking rules led the court to strike down the no-blocking rule as well.[69] The court noted that the no-blocking rule might be reconstructed as a requirement of a minimum level of service, but found that argument barred by the FCC's failure to adopt such argument in the 2010 Order or to raise that argument in its briefs.[70]

The D.C. Circuit also singled out the attempt to regulate mobile broadband for separate discussion. As noted above, a separate statutory provision provides that the FCC can only subject a mobile service to common carriage if it constitutes as a CMS, while barring common carriage regulation of PMS.[71] Because the FCC had classified mobile

Mobilizing Project v. FCC, No. 11-3627 (3d Cir. filed Sept. 26, 2011); Access Humboldt v. FCC, No. 11-72849 (9th Cir. filed Sept. 26, 2011). On October 6, 2011, the Judicial Panel on Multidistrict Litigation consolidated all of these appeals in the D.C. Circuit. *In re* Fed. Commc'ns Comm'n, Preserving the Open Internet, Report and Order, No. 1:11-ca-01356 (J.P.M.L. Oct. 6, 2011) (order granting motion to consolidate).

[63] Verizon v. FCC, 740 F.3d 623 (D.C. Cir. 2014).

[64] 47 U.S.C. § 1302 (2015).

[65] *Verizon*, 740 F.3d at 635–50 (citing 47 U.S.C. § 153(51)).

[66] *Id.* at 657.

[67] *Id.* at 650–56.

[68] *Id.* at 652 (quoting Cellco P'ship v. FCC, 700 F.3d 534, 548 (D.C. Cir. 2012)).

[69] *Id.* at 658.

[70] *Id.*

[71] 47 U.S.C. § 332(c)(2) (1996).

broadband as a PMS, mobile broadband providers are statutorily immune from common carriage requirements "twice over."[72] The invalidation of the no-blocking and the nondiscrimination rules rendered moot challenges to the decision not to apply them equally to both fixed and mobile broadband.

C. THE 2015 OPEN INTERNET ORDER

On May 15, 2014, four months after the D.C. Circuit's decision overturning the 2010 Open Internet Order's no-blocking and nondiscrimination rules, the FCC, now under the leadership of Chairman Tom Wheeler, issued a new Notice of Proposed Rulemaking designed to establish new rules.[73] The NPRM explicitly noted that it was following the "blueprint" laid out by the D.C. Circuit[74] by replacing the nondiscrimination rule with a mandate of commercial reasonableness.[75] It also adopted "the revised rationale the court suggested" and reconstructed the no-blocking rule to establish a minimal level of access.[76]

The FCC tentatively decided to follow the approach taken by the 2010 Open Internet Order that subjected mobile broadband to a less stringent no-blocking rule and exempted mobile broadband from the nondiscrimination rule altogether, although the agency sought comment on whether it should revisit those decisions.[77] The FCC also sought comment on whether it should continue to classify mobile broadband as a CMS and if so, whether forbearance should apply.[78]

President Obama's November 20, 2014, endorsement of reclassifying broadband Internet access as a Title II service changed the course of the rulemaking dramatically.[79] Although Chairman Wheeler initially expressed some reluctance,[80] the Open Internet Order adopted by the FCC

[72] *Verizon*, 740 F.3d at 650 (quoting *Cellco*, 700 F.3d at 538).

[73] Protecting and Promoting the Open Internet, Notice of Proposed Rulemaking, 29 FCC Rcd. 5561 (2014).

[74] *Id.* at 5563 ¶ 4, 5618 ¶ 162; *see also id.* at 5647 (statement of Chairman Tom Wheeler) (observing that the NPRM was designed to follow the "roadmap laid out by the court").

[75] *Id.* at 5563 ¶ 3, 5594 ¶¶ 92–93, 5599–600 ¶¶ 110–111.

[76] *Id.* at 5595 ¶ 95.

[77] *Id.* at 5583–84 ¶ 62, 5598 ¶¶ 105–106, 5609 ¶ 140.

[78] *Id.* at 5613–14 ¶ 149–150, 5616 ¶¶ 153, 155.

[79] White House Office of the Press Secretary, Statement by the President on Net Neutrality (Nov. 10, 2014), https://www.whitehouse.gov/the-press-office/2014/11/10/statement-president-net-neutrality.

[80] Brian Fung & Nancy Scuola, *Obama's Call for an Open Internet Puts Him at Odds with Regulators*, WASH. POST (Nov. 11, 2014), https://www.washingtonpost.com/news/the-switch/wp/2014/11/11/the-fcc-weighs-breaking-with-obama-over-the-future-of-the-internet/.

on February 26, 2015, explicitly reclassified broadband Internet access as a Title II service.[81] Pursuant to this authority, the 2015 Order adopted three bright-line rules prohibiting blocking, throttling, and paid prioritization backed by a catch-all standard prohibiting unreasonable interference or disadvantage to consumers or edge providers.[82] The blocking and throttling rules as well as the catch-all standard remained subject to reasonable network management.[83] The 2015 Order self-consciously revised the FCC's approach to mobile broadband.[84] In contrast to both the 2010 Order and the 2014 NPRM, the 2015 Order opted to apply the same rules to both fixed and mobile broadband.[85] Consistent with this change, the FCC reclassified mobile broadband as a CMS or its functional equivalent instead of a PMS.[86] The FCC continued to recognize that mobile networks "must address dynamic conditions that fixed, wired networks typically do not, such as the changing location of users as well as other factors affecting signal quality," as well as more restrictive capacity constraints.[87] The 2015 Order thus explicitly recognized that these challenges must be taken into account when assessing whether a practice constitutes reasonable network management and cautioned that this inquiry must preserve mobile broadband operators' flexibility.[88] The D.C. Circuit upheld these aspects of the FCC's decision.[89] The 2017 NPRM sought comment on once again classifying mobile broadband as a CMS and reopened the question whether mobile broadband should be regulated differently from fixed broadband.[90]

* * *

The FCC's network neutrality regulations have consistently acknowledged that the challenges associated with mobile broadband justify subjecting mobile broadband to lighter touch regulation than fixed broadband. In particular, the current rules require a detailed, context-

[81] 2015 Open Internet Order, *supra* note 3, at 5618 ¶ 59, 5757–77 ¶¶ 355–87.

[82] *Id.* at 5607–09 ¶ 14–22, 5609 ¶ 25, 5638 ¶ 92, 5685 ¶ 192.

[83] *Id.* at 5699–704 ¶¶ 214–24.

[84] *Id.* at 5635–43 ¶¶ 86–101.

[85] *Id.* at 5609 ¶ 25, 5638 ¶ 92, 5685 ¶ 192. The FCC also sought comment on how the transparency rule should apply to mobile, *id.* at 5669 ¶ 155, and created a safe harbor for disclosures made in the format established by the Consumer Advisory Committee, *id.* at 5680 ¶ 179.

[86] *Id.* at 5615 ¶ 48, 5776–90 ¶¶ 388–408.

[87] *Id.* at 5703 ¶ 223 (footnotes omitted); *accord id.* at 5611 ¶ 34.

[88] *Id.* at 5611 ¶ 34, 5643 ¶ 101, 5651 ¶ 118, 5665 ¶ 148, 5701 ¶ 218, 5703–04 ¶ 223.

[89] U.S. Telecom Ass'n v. FCC, 825 F.3d 674, 713–26 (D.C. Cir. 2016).

[90] 2017 Open Internet NPRM, *supra* note 3, at 20–22 ¶¶ 55–62, 30 ¶ 94.

specific assessment to determine whether a mobile operator's particular practice constitutes reasonable network management.

III. THE BASIC ARCHITECTURAL COMMITMENTS UNDERLYING NETWORK NEUTRALITY

The FCC ruled that mandating network neutrality was necessary to preserve two architectural features that have proven essential to promoting innovation.[91] First, broadband Internet access providers had to be prevented from blocking or disadvantaging traffic associated with certain edge providers or applications.[92] Second, regulators had to preserve the end-to-end architecture.[93] Each will be discussed in turn.

A. THE (SUPPOSED) ABSENCE OF PRIORITIZATION/QUALITY OF SERVICE

Network neutrality advocates often assert that the Internet is also based on the commitment not to permit routers to prioritize traffic based on its source, content, or the application with which it is associated.[94] Such prioritization would allow broadband providers to harm innovation by "preferring their own or affiliated content, demanding fees from edge providers, or placing technical barriers to reaching end users."[95]

As a matter of history, the claim that the Internet's architecture did not permit prioritization is problematic.[96] Since its inception, the IP header has contained a six-bit *type of service field* designed to allow the attachment of different levels of priority to individual packets.[97] The original design

[91] Some of these commitments fall outside the scope of this paper. One prime example is the idea of protocol layering. *See* Christopher S. Yoo, *Protocol Layering and Internet Policy*, 161 U. PA. L. REV. 1707 (2013). Another example is modularity. *See* Christopher S. Yoo, *Modularity Theory and Internet Policy*, 2016 U. ILL. L. REV. 1.

[92] *See* 2010 Open Internet Order, *supra* note 5, at 17,915–23 ¶¶ 21–31.

[93] *See id.* at 17,909–10 ¶ 13 & n.13.

[94] *See, e.g.*, 2010 Open Internet Order, *supra* note 5, at 17,947 ¶ 76 ("pay for priority would represent a significant departure from historical and current practice"); LAWRENCE LESSIG, THE FUTURE OF IDEAS 37 (2002) (arguing that "the design effects a neutral platform—neutral in the sense that the network owner can't discriminate against some packets while favoring others").

[95] *See* 2015 Open Internet Order, *supra* note 2, at 5629 ¶ 80.

[96] *See* David D. Clark, *The Design Philosophy of the DARPA Internet Protocols*, ACM SIGCOMM COMPUTER COMM. REV., Aug. 1988, 108 ("The second goal [of the DARPA architecture after survivability] is that it should support . . . a variety of types of service. Different types of service are distinguished by differing requirements for such things as speed, latency and reliability."); *see also* Kai Zhu, Note, *Bringing Neutrality to Net Neutrality*, 22 BERKELEY TECH. L.J. 615, 619–21, 634–38 (2007) (observing that the Internet was never designed to be neutral).

[97] Info. Sci. Inst., *Internet Protocol: DARPA Internet Program Protocol Specification* 8, 18, 35–36 (Sept. 1981), http://tools.ietf.org/pdf/rfc791 (IETF Network

accommodated three levels of precedence as well as additional flags for particular needs regarding delay, throughput, and reliability, although subsequent changes now allow this field to be used even more flexibly.[98]

Moreover, claims that the Internet is hostile toward prioritization ignore certain realities about the routing architecture. Tier 1 ISPs share information about the routing architecture with one another through the Border Gateway Protocol (BGP). Enabling networks to engage in policy-based routing that alters the path that traffic takes based on its source or destination represented one of the principal motivations behind BGP's most recent redesign.[99]

Nor did efforts to support prioritization end there. Throughout the Internet's history, the Internet Engineering Task Force (IETF) has issued standards designed to allow networks to provide differential levels of quality of service, including Integrated Services (IntServ),[100] Differentiated Services (DiffServ),[101] MultiProtocol Label Switching (MPLS),[102] and such modern initiatives as Low Extra-Delay Batch Transport (LEDBAT).[103] Providing better support for quality of service (particularly for real-time data) was identified as one of the major goals of the transition to IPv6.[104] Indeed, IPv6 includes a *traffic class* field that is

Working Group Request for Comments no. 791); *see also* Info. Sci. Inst., *DoD Standard Internet Protocol* 12, 26–27, (Dec. 1979), http://128.9.160.29/ien/txt/ien123.txt (Internet Engineering Note no. 123).

[98] ANDREW S. TANENBAUM & DAVID J. WEATHERALL, COMPUTER NETWORKS 440 (5th ed. 2003).

[99] CHRISTIAN HUITEMA, ROUTING IN THE INTERNET 195 (1995); Kirk Lougheed, *A Border Gateway Protocol (BGP)* 1 (June 1981), http://tools.ietf.org/pdf/rfc1105 (IETF Network Working Group Request for Comments no. 1105). A leading textbook gives the following examples of policy-based routing: "1. Do not carry commercial traffic on an educational network. 2. Never send traffic from the Pentagon on a route through Iraq. 3. Use TeliaSonera instead of Verizon because it is cheaper. 4. Don't use AT&T in Australia because performance is poor. 5. Traffic starting or ending at Apple should not transit Google." TANENBAUM & WEATHERALL, *supra* note 98, at 479.

[100] *See* Robert Braden et al., *Integrated Services in the Internet Architecture: An Overview* (June 1994), https://tools.ietf.org/html/rfc1633 (IETF Network Working Group Request for Comments no. 1633).

[101] *See* Steven Blake et al., *An Architecture for Differentiated Services* (Dec. 1998), https://tools.ietf.org/html/rfc2475 (IETF Network Working Group Request for Comments no. 2475).

[102] *See* Eric C. Rosen et al., *Multiprotocol Label Switching Architecture* (Jan. 2001), https://tools.ietf.org/html/rfc3031, (IETF Network Working Group Request for Comments no. 3031).

[103] *See* Stanislav Shalunov et al., *Low Extra Delay Background Transport (LEDBAT)* (Dec. 2012), https://tools.ietf.org/html/rfc6817 (IETF Network Working Group Request for Comments no. 6817).

[104] Scott Bradner & Allison Mankin, *IP: Next Generation (IPng) White Paper Solicitation* 4 (Dec. 1993), http://tools.ietf.org/pdf/rfc1550 (IETF Network Working

analogous to the type of service field in IPv4.[105] Moreover, IPv6 added a *flow label* field similar to the labels used by MPLS to incorporate prioritization and other routing policies.[106]

To say that prioritization has a long historical pedigree is not to say that it has won the day. Just as quality of service has its advocates within the engineering community, it also has its detractors. If the presentations in the leading textbooks on network engineering are any guide, the controversy over quality of service shows no signs of abating, with people on both sides of the argument holding strong views.[107] This Article is not intended to take sides in this debate. Instead, the goal is simply to emphasize that the debate over the relative merits of prioritization remains far from settled. In any event, as the following discussion demonstrates, the arguments in favor of prioritizing certain applications over others becomes increasingly compelling when wireless networks are involved.

B. THE END-TO-END ARGUMENT

Another architectural principle often regarded as essential to enhancing innovation is known as the end-to-end argument.[108] In end-to-end system designs, the routers operating in the middle of the network are not optimized for any particular application; instead, any functionality needed to support particular applications is confined to the hosts operating at the edges of the network.[109] Restricting application-specific intelligence to the edges of the network allows developers of new applications to focus exclusively on the software running in the hosts and to avoid having to modify any application-specific programs running in the core of the network.[110] This gives entrepreneurs the confidence that they will remain

Group Request for Comments no. 1550); *accord* DOUGLAS E. COMER, INTERNETWORKING WITH TCP/IP 563 (5th ed. 2006); LARRY L. PETERSON & BRUCE S. DAVIE, COMPUTER NETWORKS: A SYSTEMS APPROACH 319 (4th ed. 2007); TANENBAUM & WEATHERALL, *supra* note 98, at 456.

[105] Stephen E. Deering & Robert M. Hinden, *Internet Protocol, Version 6 (IPv6) Specification* 25 (Dec. 1998), http://tools.ietf.org/pdf/rfc2460 (IETF Network Working Group Request for Comments no. 2460).

[106] *Id.*

[107] *See* COMER, *supra* note 104, at 510, 515.

[108] The seminal statement of the end-to-end argument is found in J.H. Saltzer, D.P. Reed & D.D. Clark, *End-to-End Arguments in System Design*, 2 ACM TRANSACTIONS ON COMPUTING 277 (1984). For another leading statement, see Brian E. Carpenter, *Architectural Principles of the Internet* 2–3 (June 1996), http://tools.ietf.org/pdf/rfc1958 (IETF Network Working Group Request for Comments no. 1958) [hereinafter RFC 1958].

[109] 2009 Open Internet NPRM, *supra* note 5, at 13,070 ¶ 19.

[110] *Id.*

free to innovate without having to seek permission from any broadband Internet access providers.[111]

Although end-to-end system design is sometimes treated as if it were an absolute mandate, it should instead be treated as a pragmatic rule of thumb that should give way under appropriate circumstances.[112] Even the IETF document that is most strongly associated with the principle recognizes that the continuous nature of technological change means that architecture principles inevitably change as well.[113] This document observed that "[p]rinciples that seem sacred today will be deprecated tomorrow" and that "[t]he principle of constant change is perhaps the only principle of the Internet that should survive indefinitely."[114] As a result, the document rejected the idea that the end-to-end argument represented "dogma about how Internet protocols should be designed."[115] Indeed, the document recognized that circumstances might cause the Internet Protocol to change altogether.[116]

The end-to-end argument is operationalized through two principles relevant to this Article. First, in an end-to-end design, routers do not maintain any information associated with any particular traffic. This is known as flow state or per-flow state.[117] Second, each host should have a unique address that is visible to all other machines.[118]

1. *The Absence of Per-Flow State*

One of the central commitments around which the Internet is designed is that the routers operating in the core of the network store the individual segments comprising larger communication (known as packets) for the minimum time needed to forward them toward their final destination. As soon as the routers have finished forwarding the packets, the routers discard all information associated with them. Two corollaries of this principle are that each router makes its own decision about the direction to route any particular packet and that each packet travels through the

[111] *Id*. at 13,089 ¶ 63; *accord* Protecting and Promoting the Open Internet, Notice of Proposed Rulemaking, 29 FCC Rcd. 5561, 5564 ¶ 8; 2010 (2014); 2010 Open Internet Order, *supra* note 5, at 17,909–10 ¶ 13 & n.13.

[112] Christopher S. Yoo, *Would Mandating Network Neutrality Help or Hurt Broadband Competition?: A Comment on the End-to-End Debate*, 3 J. ON TELECOMM. & HIGH TECH. L. 23 (2004).

[113] RFC 1958, *supra* note 108, at 1.

[114] *Id.*

[115] *Id.* at 2.

[116] *Id.* at 3.

[117] Clark, *supra* note 96, at 113 (flow state); Christopher S. Yoo, *The Changing Patterns of Internet Usage*, 63 FED. COMM. L.J. 7, 86 (2010) (per-flow state).

[118] RFC 1958, *supra* note 108, at 5.

network independent of the packets preceding or following it in the data stream. This concept represented a sharp change from the architecture around which the telephone network was designed, which established dedicated circuits between end users and channeled all of the data associated with that communication along that circuit. The switches in the core of such a circuit-switched network, such as the telephone network, must necessarily retain a lot of information about each flow passing through the network. This information about where packets came from or where they are routed to is called *per-flow state*.[119]

The Internet's origins as a military network meant that the architects placed the highest priority on *survivability*, measured by the network's continuing ability to operate despite the loss of nodes within the network.[120] Networks that rely on a large amount of per-flow state tend not to be particularly robust in this manner. Consider what occurs when a switch in the middle of a telephone network fails. When the switch is lost, so too is all of the information maintained by the switch with respect to each flow. The loss of this per-flow state means that neither the network nor the end user can recover from this event. As a result, the communication fails, and the only way to reestablish it is by placing a new call. Designing the network to avoid per-flow state in the core of the network increased the network's survivability.[121]

That said, some entity involved in the communication must maintain per-flow state in order to monitor whether the communication was ever delivered. Should that entity fail the communication would necessarily fail as well. The Internet architects assigned responsibility for these function to the computers operated by end users at the edge of the network, called *hosts*, a practice that has become known as *fate sharing*. The rationale is that if the hosts involved in the communication fail, there is probably no need to finish the communication.[122]

Although survivability represented the original justification for avoiding having routers operate in the core of the network to maintain per-flow state, this rationale has little applicability to the modern Internet. While the loss of nodes may be a real concern in the hostile environments in which the military operates, the destruction of nodes is not typically a major concern in commercial networks.[123] Instead, the modern rationale for avoiding the maintenance of per-flow state in the core of the network

[119] *Id.*
[120] *See* Clark, *supra* note 96, at 106–07.
[121] *Id.* at 108.
[122] *Id.*; RFC 1958, *supra* note 108.
[123] Clark, *supra* note 96, at 107.

is to facilitate the interconnection of networks that operate on very different principles.

The manner in which the absence of per-flow state facilitates interconnection is well illustrated by the history of the Advanced Research Projects Agency Network (ARPANET), which is widely regarded as the predecessor to the Internet.[124] In the ARPANET, all of the routers operating in the core of the network, called Interface Message Processors or IMPs, were manufactured by a single company based on the same computer and ran the same software, and were interconnected by the same technology—telephone lines.[125] The IMPs were responsible for a wide variety of tasks. For example, consistent with the standard approach of day,[126] IMPs were responsible for making sure that the packets were successfully delivered to the next IMP and, if not, for correcting any errors by resending the packets.[127] In addition, IMPs were responsible for congestion control.[128]

As a result, IMPs had to maintain a large amount of information about the current status of the packets passing through its network. Although these tasks were often quite complex, the fact that all IMPs were constructed with the same technology and operated on the same principles made them very easy to interconnect. The architects encountered greater problems when they attempted to interconnect the ARPANET with the two other packet networks sponsored by the Defense Department: the San Francisco Bay Area Packet Radio Network (PRNET) and the Atlantic Packet Satellite Network (SATNET). Differences in transmission technologies, throughput rates, packet sizes, and error rates made these networks remarkably difficult to interconnect. In addition, every network would have to maintain the same per-flow state information as the other network with which it wanted to interconnect and would have to

[124] *See* JANET ABBATE, INVENTING THE INTERNET 113–33 (1999).

[125] F.E. Heart et al., *The Interface Message Processor for the ARPA Computer Network*, 36 AFIPS CONF. PROC. 551, 552 (1970).

[126] *See* Geoff Huston, *The End of End to End?*, ISP COLUMN (May 2008), at 1, http://www.potaroo.net/ispcol/2008-05/eoe2e.pdf (noting that the predominant approach to digital networking during the 1970s and 1980s required that each switch in a path store a local copy of the data until it received confirmation that the downstream switch has received the data).

[127] John M. McQuillan & David C. Walden, *The ARPANET Design Decisions*, 1 COMPUTER NETWORKS 243, 282 (1977).

[128] Christopher S. Yoo, *Protocol Layering and Internet Policy*, 161 U. PA. L. REV. 1707, 1758 (2013).

understand its expected response when receiving a communication from another router.[129]

The International Network Working Group (INWG) considered a variety of solutions to these problems.[130] It rejected as too cumbersome and too error-prone approaches that would have required every host to run simultaneously every protocol used by other types of networks[131] or would have required each system to translate the communication into another format whenever it crossed a boundary between autonomous systems as too cumbersome and error-prone.[132] Instead, Vinton Cerf and Robert Kahn's seminal article creating the Internet Protocol (IP) established a single common language that all networks could understand.[133] To facilitate its use by multiple networks, this common language was kept as simple as possible and included only the minimum information needed to transmit the communication.[134] All of this information was placed in an internetwork header that every gateway could read without modifying it.[135] The fact that all of the information needed to route a packet was contained in the IP header eliminated the need for any router to know anything about the design of the upstream network delivering the packet to it or about the design of the downstream network to which it was delivering the packet.

This in turn meant that functions previously handled by routers, such as reliability, were now assigned to the hosts operating at the edge of the network. Even friendly observers have conceded that at the time this approach was regarded as "heresy,"[136] "unconventional,"[137] and "odd."[138] Over time, it has become an accepted feature of the network.

[129] *See* ABBATE, *supra* note 124.

[130] *Id.* at 131–32.

[131] Vinton G. Cerf & Robert E. Kahn, *A Protocol for Packet Network Interconnection*, 22 IEEE TRANSACTIONS ON COMM. 637, 638 (1974) ("The unacceptable alternative is for every HOST or process to implement every protocol . . . that may be needed to communicate with other networks.").

[132] *See* ABBATE, *supra* note 124, at 128; Vinton G. Cerf & Peter T. Kirstein, *Issues in Packet-Network Intercommunication*, 66 PROC. IEEE 1386, 1399 (1978).

[133] Cerf & Kahn, *supra* note 131, at 638.

[134] *See* Barry M. Leiner et al., *The DARPA Internet Protocol Suite*, IEEE COMM., Mar. 1985, at 29, 31 ("The decision on what to put into IP and what to leave out was made on the basis of the question 'Do gateways need to know it?'.").

[135] Cerf & Kahn, *supra* note 131, at 638–39.

[136] Huston, *supra* note 126, at 1.

[137] ABBATE, *supra* note 124, at 125.

[138] Ed Krol & Ellen Hoffman, *FYI on "What Is the Internet?"* 2, 4 (May 1993), http://tools.ietf.org/pdf/rfc1462 (IETF Network Working Group Request for Comments no. 1462).

2. *Unique, Universal Addresses Visible to All Other Machines*

The interconnection of different networks was further complicated by the fact that each network tended to employ its own idiosyncratic scheme for assigning addresses to individual hosts and routers.[139] The Internet's architects solved this problem by requiring that that all networks employ a single, uniform addressing scheme common to all networks.[140] This scheme included the address information in the header of every IP packet so that every router could access the address information directly instead of having to maintain per-flow state. Moreover, hosts operating at the edge of the network must make their IP addresses visible to the rest of the network.[141]

IV. TRAFFIC GROWTH, BANDWIDTH CONSTRAINTS, AND NETWORK MANAGEMENT

The sharp increase in bandwidth consumption poses one of the biggest challenges to wireless networks. Since 2010, the number of mobile broadband subscribers has exceeded the number of subscribers of all other broadband technologies combined.[142] Moreover, industry observers estimate that wireless traffic will grow at an annual rate of 57% from 2014 to 2019, as compared with a growth rate of 23% forecast for fixed Internet service.[143] When traffic saturates the available capacity, packets are forced to wait in queues. These queues become sources of jitter and delay, which degrades the quality of service provided by the network.

The increase in the number of mobile broadband subscribers and the growth in wireless broadband traffic have increased the need for network providers to engage in network management. As a general matter, there are two classic approaches to managing explosive traffic growth. One

[139] *See* Cerf & Kahn, *supra* note 131, at 637.

[140] *See* Cerf & Kirstein, *supra* note 132, at 1393, 1399 (discussing the common internal address structure required for packet-level interconnectivity); Cerf & Kahn, *supra* note 131, at 641 ("A uniform internetwork TCP address space, understood by each GATEWAY and TCP, is essential to routing and delivery of internetwork packets.").

[141] Tony Hain, *Architectural Implications of NAT* 7–8, 18 (Nov. 2000), http://tools.ietf.org/pdf/rfc2993 (IETF Network Working Group Request for Comments no. 2993).

[142] Fed. Commc'ns Comm'n, Internet Access Services: Status as of December 31, 2013, https://transition.fcc.gov/Daily_Releases/Daily_Business/2014/db1016/DOC-329973A1.pdf.

[143] *See* CISCO SYS., INC., CISCO VISUAL NETWORKING INDEX: FORECAST AND METHODOLOGY, 2014–2019, at 5 tbl.1 (2015), http://www.cisco.com/c/en/us/solutions/collateral/service-provider/ip-ngn-ip-next-generation-network/white_paper_c11-481360.pdf.

solution is simply to increase network capacity.[144] The presence of additional headroom makes it less likely that spikes in traffic will saturate the network, which in turn allows the packets to pass through the network without any delay. The other solution employs network management to give a higher priority to traffic associated with those applications that are most sensitive to delay.[145]

For example, traditional Internet applications, such as email and web browsing, are essentially file transfer applications. Because file transfer applications typically display their results only after the last packet is delivered, delays in the delivery of intermediate packets typically do not adversely affect their performance. This contrasts with real-time, interactive applications, such as Voice Over Internet Protocol (VoIP), video conferencing, and virtual worlds, which are becoming increasingly important on the Internet. The performance of these applications depends on the arrival time and spacing of every intermediate packet, with delays of as little as one third of a second being enough to render the service unusable.[146] As such, these applications are considerably more vulnerable to network congestion.[147]

Networks can help protect the operation of time-sensitive applications either by expanding capacity or by giving their packets a higher priority. In the latter case, it is conceivable that the network need only to rearrange the order of the intermediate packets without affecting when the last packet will arrive. If so, network management can improve the performance of the time-sensitive application without having any adverse impact on the application that is less time sensitive. Even if small delays occur, with non-time-sensitive applications such as file-transfer, delays of a fraction of a second are virtually undetectable.

A review of leading computer networking textbooks reveals that the choice between these two approaches has long been a source of

[144] For a representative statement appearing in the engineering literature, see Yaqing Huang & Roch Guerin, *Does Over-Provisioning Become More or Less Efficient as Networks Grow Larger?*, PROC. 13TH IEEE INT'L CONF. ON NETWORK PROTOCOLS (ICNP) 225 (2005). For a similar statement appearing in the legal literature, see, for example, LESSIG, *supra* note 94, at 47 (arguing in favor of addressing bandwidth scarcity by increasing capacity instead of implementing quality of service).

[145] Christopher S. Yoo, *Beyond Network Neutrality*, 19 HARV. J.L. & TECH. 1, 21–23 (2005).

[146] International Telecommunication Union, ITU Recommendation G.114 (2003).

[147] The problem is most acute for interactive video, such as video conferencing. For linear video (whether prerecorded or live), media players can ameliorate the jitter caused by congestion by delaying playback to buffer a quantity of packets so they may be released in a steady stream. Yoo, *supra* note 117, at 71.

controversy in the engineering community with respect to wireline networks.[148] In the wireline context, engineering studies indicate that the amount of headroom needed to preserve quality of service without prioritization can be substantial.[149] Expanding bandwidth thus maintains simplicity, but requires the incurrence of significant capital costs. The additional cost associated with nonprioritized solutions increases the number of subscribers that a bandwidth expansion needs to breakeven, which in turn limits broadband deployment in ways that are likely to exacerbate the digital divide.[150] Network management, on the other hand, substitutes operating costs for capital costs, which allows them to be recovered as they are incurred. It does have the side effect of adding complexity to the network.

The tradeoff between these two approaches plays out much differently in the context of wireless networking. As an initial matter, wireless networks face limits that wireline networks do not face with regards to the number of end users that can be served in a particular area. A person connected to the Internet via a wireline technology (whether fiber, coaxial cable, or twisted pairs of copper) employs a signal that is narrowly channeled through space. This geographic limitation allows multiple end users to avoid interfering with one another even if they are sitting side by side.[151]

Wireless signals propagate quite differently. Unlike wireline signals, wireless signals propagate in an unchanneled manner in all directions.[152] The signals of one user are thus perceived as noise by other end users. As Claude Shannon recognized in 1948, the increase in noise reduces the amount of usable bandwidth available to those other users.[153] The greater the density of users becomes, the more constricted the bandwidth

[148] *See* COMER, *supra* note 104, at 510, 515.

[149] *See* M. Yuksel et al., *Quantifying Overprovisioning vs. Class-of-Service: Informing the Net Neutrality Debate*, PROC. 9TH INT'L CONF. ON COMPUTER COMM. & NETWORKS (2010), http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5560131.

[150] Christopher S. Yoo, *Network Neutrality, Consumers, and Innovation*, 2008 U. CHI. LEGAL FORUM 179, 188, 229–32.

[151] The fact that any electrical current creates some degree of radio frequency interference means that adjacent usage does create some interference. Any such interference occurs at very low power and can be minimized by proper shielding of the cables and the equipment.

[152] Piyush Gupta & P.R. Kumar, *The Capacity of Wireless Networks*, 46 IEEE TRANSACTIONS ON INFO. THEORY 388 (2000).

[153] Claude E. Shannon, *Communication in the Presence of Noise*, 37 PROC. INST. RADIO ENGINEERS 10 (1949).

becomes. This implies that there is an absolute limit to the density of end users who can use wireless broadband in any particular geographic area.[154]

Even more importantly, wireless providers' options for expanding capacity are much more limited than for wireline networking. Wireless providers can increase bandwidth by deploying a larger number of microwave base stations operating at lower power or by deploying increasingly sophisticated receiving equipment. Such solutions are typically quite costly. Moreover, the gains from such strategies are finite. Once they are exhausted, the restrictions on the amount of spectrum allocated to any particular service sharply limits network providers' ability to expand capacity any further.[155]

These bandwidth limitations require wireless networks to engage in extensive network management.[156] Specifically, if a subscriber in a low-bandwidth location is speaking on the telephone, the wireless network will prioritize the voice traffic and hold all email and other data traffic until the subscriber moves to a higher-bandwidth location or ends the call.[157] Other services rate-limit or prohibit video and other high-bandwidth services to ensure that a small number of users do not occupy all of the available bandwidth.[158] Technologies such as T-Mobile's Binge On adopt a different approach: it uses a strategy pioneered by MetroPCS[159] to reduce the bandwidth needed to convey video by reducing the resolution of all video to 480p. The bandwidth reductions are so significant that T-Mobile is able to exempt this traffic from counting towards end-users' data caps.[160] From a technical standpoint, this scheme does not work for non-video applications and thus cannot be employed in an application-agnostic way. A prohibition on prioritization based on applications would obstruct these types of network management tools from being deployed.

[154] Gupta & Kumar, *supra* note 152, at 391–92. Wireless operators can reduce this interference by using directional transmitters and receivers. Such solutions work only if you know the location of every sender and receiver. As such, they are poorly suited to wireless networking of mobile devices.

[155] Charles Jackson et al., *Spread Spectrum Is Good—But It Does Not Obsolete* NBC v. U.S.*!*, 58 FED. COMM. L.J. 245, 253–59 (2006).

[156] *See* Charles L. Jackson, *Wireless Efficiency Versus Net Neutrality*, 63 FED. COMM. L.J. 445, 477 (2011).

[157] Yoo, *supra* note 117, at 78.

[158] *Id.* at 78–79.

[159] Christopher S. Yoo, Wickard *for the Internet: Network Neutrality after* Verizon v. FCC, 66 FED. COMM. L.J. 415, 458–59 (2014).

[160] Jason Cipriani, *T-Mobile's Binge On Streams 480p Video. Does It Matter?*, FORTUNE (Nov. 11, 2015, 12:55 PM), http://fortune.com/2015/11/11/tmobile-480p-video/.

Prioritizing certain applications over others requires tight integration of the network and the device. The FCC noted as much when repealing the regulation barring network providers from bundling telecommunications services with the devices used by end-user, also known as customer premises equipment or CPE. The FCC recognized that the equipment that increasingly serves as enhancements to the network requires sophisticated interactions between the network and the device that was being impeded by the unbundling requirement.[161] In other words, the device was part of the functionality of the network itself, a fact that renders calls for mandating that wireless broadband networks be open to all devices problematic.[162]

A. Differences in wireline and wireless Quality of Service and Reliability

Wireline and wireless broadband networks also differ in terms of their reliability. As anyone who has suffered through dropped calls on their mobile telephone recognizes, wireless technologies are much less reliable than wireline technologies. Part of the problem is the difficulty of seamlessly handing off a communication when a mobile wireless user transfers from one base station to another. Other problems are due to the physics of wave propagation, which can cause interference in wireless networks to arise in much more transient and unpredictable ways than in wireline networks.

These differences in reliability have implications for many basic architectural decisions for the Internet. For example, although the current network relies on hosts to correct errors by resending packets that are dropped, in a wireless world it is often more efficient to assign responsibility for those functions to routers operating in the core of the network. In addition, wireline networks rely on hosts to manage congestion on the Internet. For reasons discussed below, wireless networks' lack of reliability means that that the traditional approach to congestion management will not work well on wireless. The result is that basic functions such as recovery from errors and managing congestion—two of the most fundamental functions performed by the network—operate far differently on wireless networks than on wireline networks.

[161] Policy and Rules Concerning the Interstate, Interexchange Marketplace, Report and Order, 16 FCC Rcd. 7418, 7427 ¶ 16 (2001).

[162] *See* Wu, *supra* note 9, at 395–401.

B. DIFFERENT DIMENSIONS OF QUALITY OF SERVICE

The performance guarantees provided by different networking technologies, known as quality of service or QoS, can vary widely. Most commentators discuss quality of service in terms of guaranteed throughput rates. As a preliminary matter, it bears mentioning that the engineering community typically views quality of service as occupying more dimensions than mere bandwidth. In addition, networks vary in terms of their reliability (i.e., the accuracy with which they convey packets), delay or latency (i.e., the amount of time it takes for the application to begin functioning after the initial request is made), and jitter (i.e., variations in the regularity of the spacing between packets).[163]

Interestingly, applications vary widely in the types of quality of service they demand. For example, the transfer of health records is not particularly bandwidth intensive and can accept millisecond latencies and jitter without much trouble, but is particularly demanding in terms of reliability. Voice over Internet Protocol (VoIP) is also not bandwidth intensive and tolerates unreliability, but is quite sensitive to latency and jitter. Financial transactions have low bandwidth requirements, but must have latency guarantees in the microseconds and perfect reliability. Interactive video applications (such as video conferencing and virtual worlds) are bandwidth intensive and intolerant of jitter and latency, but can allow a degree of unreliability.

Furthermore, network systems can improve certain dimensions of quality of service, but only at the expense of degrading other dimensions.[164] For example, streaming video works best when packets arrive in a steady stream. As a result, it is quite sensitive to jitter. Irregularities in the spacing between packets can be largely eliminated by placing all of the arriving packets in a buffer for some length of time and beginning to release them later. The presence of an inventory of backlogged packets allows them to be released in a nice even pattern. The cost, however, is to create a delay before the application begins to run.

C. CAUSES OF POOR QUALITY OF SERVICE ON WIRELESS BROADBAND NETWORKS

Quality of service on wireless broadband networks can degrade for a wide variety of reasons not applicable to wireline networks. These reasons

[163] TANENBAUM, *supra* note 98, at 405.

[164] CHRISTOPHER S. YOO, THE DYNAMIC INTERNET: HOW TECHNOLOGY, USERS, AND BUSINESSES ARE TRANSFORMING THE NETWORK 25–27 (2012).

include bad handoffs between base stations, local congestion, and the physics of wave propagation.

1. Bad Handoffs

Bad handoffs represent an important cause of poor quality of service in mobile broadband networks. In order to receive service, a wireless device must typically establish contact with some base station located nearby. Circumstances may require a device to transfer its connection from one base station to another. For example, the mobile host may have moved too far away from the original base station. Alternatively, the current base station may have become congested or environmental factors may have caused the signal strength between the current base station and the mobile host to have deteriorated.[165] For reasons discussed more fully below, transferring responsibility for a mobile host from one base station to another has proven to be quite tricky. It is not unusual for wireless networks to make bad handoffs, which can cause communications to be dropped.

2. Local Congestion

In addition, because wireless technologies share bandwidth locally, they are more susceptible to local congestion than many fixed-line services, such as DSL and fiber to the home. Local congestion makes end users acutely sensitive to the downloading behavior of their immediate neighbors. Other technologies, such as cable modem systems, are also subject to local congestion. The more restrictive bandwidth limitations make this problem worse for wireless networks, as does the fact that wireless networks are typically designed so that data and voice traffic share bandwidth, unlike wireline telephone and cable modem systems which place their data traffic in a different channel from their core business offerings. As a result, wireless broadband networks are particularly susceptible to spikes in demand.

These limits have led many wireless providers to limit or ban bandwidth intensive applications, such as video and peer-to-peer downloads, in order to prevent a small number of users from rendering the service completely unusable. For example, some providers using unlicensed spectrum to offer wireless broadband in rural areas have indicated that they bar users from operating servers for this reason.[166]

[165] JAMES F. KUROSE & KEITH W. ROSS, COMPUTER NETWORKING: A TOP-DOWN APPROACH 572–74 (6th ed. 2013).

[166] *See, e.g.*, *Ensuring Competition on the Internet: Net Neutrality and Antitrust: Hearing Before the Subcomm. on Intellectual Prop., Competition, and the Internet of the*

United blocks video on its airplanes. Amtrak similarly blocks video and restricts large downloads on its train, while permitting such traffic in its stations where bandwidth is less restricted.[167]

3. The Physics of Wave Propagation

The unique features of waves can cause wireless technologies to face interference problems that are more complex and fast-changing than anything faced by wireline technologies. Anyone who has studied physics knows that waves have some distinctive characteristics. These characteristics can reinforce each other in unexpected ways, as demonstrated by unusual echoes audible in some locations in a room and by whispering corners, where the particular shape of the room allows sound to travel from one corner to the other even though a person speaks no louder than a whisper. As noise-reducing headphones and cars demonstrate, waves can also cancel each other out. Waves also vary in the extent to which they can bend around objects and pass through small openings, depending on their wavelength. The discussion that follows is necessarily simplified, but it is sufficient to convey the intuitions underlying some of the considerations that make wireless networking so complex.

For example, wireless signals attenuate much more rapidly with distance than do wireline signals, which makes bandwidth much more sensitive to small variations in how distant a particular user is from the nearest base station. This requires wireless providers to allocate bandwidth by dynamically requiring individual transmitters to adjust their power. The physics of wireless transmission can also create what is known as the "near-far" problem, where a transmitter can completely obscure the signal of another transmitter located directly behind it by broadcasting too loudly.[168] WiFi networks similarly adjust the power of individual users dynamically to help allocate bandwidth fairly.[169] Again, the solution is to require the nearer transmitter to reduce its power, and accordingly its available bandwidth, in order for the other transmitter to be heard.

H. Comm. on the Judiciary, 112th Cong. 55 (2011) (prepared testimony of Laurence Brett ("Brett") Glass, Owner and Founder, LARIAT).

[167] Yoo, *supra* note 147, at 79 n.39.

[168] *See, e.g.*, Mahesh K. Varanasi & Behnaam Aazhang, *Optimally Near-Far Multiuser Detection in Differentially Coherent Synchronous Channels*, 37 IEEE TRANSACTIONS ON INFO. THEORY 1006 (1991).

[169] *See, e.g.*, Huazhi Gong & JongWon Kim, *Dynamic Load Balancing Through Association Control of Mobile Users in WiFi Networks*, 54 IEEE TRANSACTIONS ON CONSUMER ELEC. 342 (2008).

Moreover, in contrast to wireline technologies, there is an absolute limit to the density of wireless users that can operate in any particular area. Shannon's Law dictates that the maximum rate with which information can be transmitted given limited bandwidth is a function of the signal-to-noise ratio.[170] Unlike wireline transmissions, which travel in a narrow physical channel, wireless signals propagate in all directions and are perceived as noise by other receivers. That means that when more people use wireless broadband, the amount of bandwidth available to others operating in the same area is reduced. At some point, the noise becomes so significant that the addition of any additional wireless radios becomes infeasible.

Managing wireless networks is further complicated by the fact that waves are also subject to refraction and diffraction. Refraction is a change in speed and direction that occurs whenever the transmission medium through which the wave is passing changes, such as when a wave travelling through the air passes through a wall and then back into the air. The change in speed necessarily causes a change in the wave frequency. Diffraction occurs when a wave tries to bend around an obstacle or passes through a slit that is comparable in size to its wavelength. It has long been recognized that diffraction can cause complex patterns of interference.

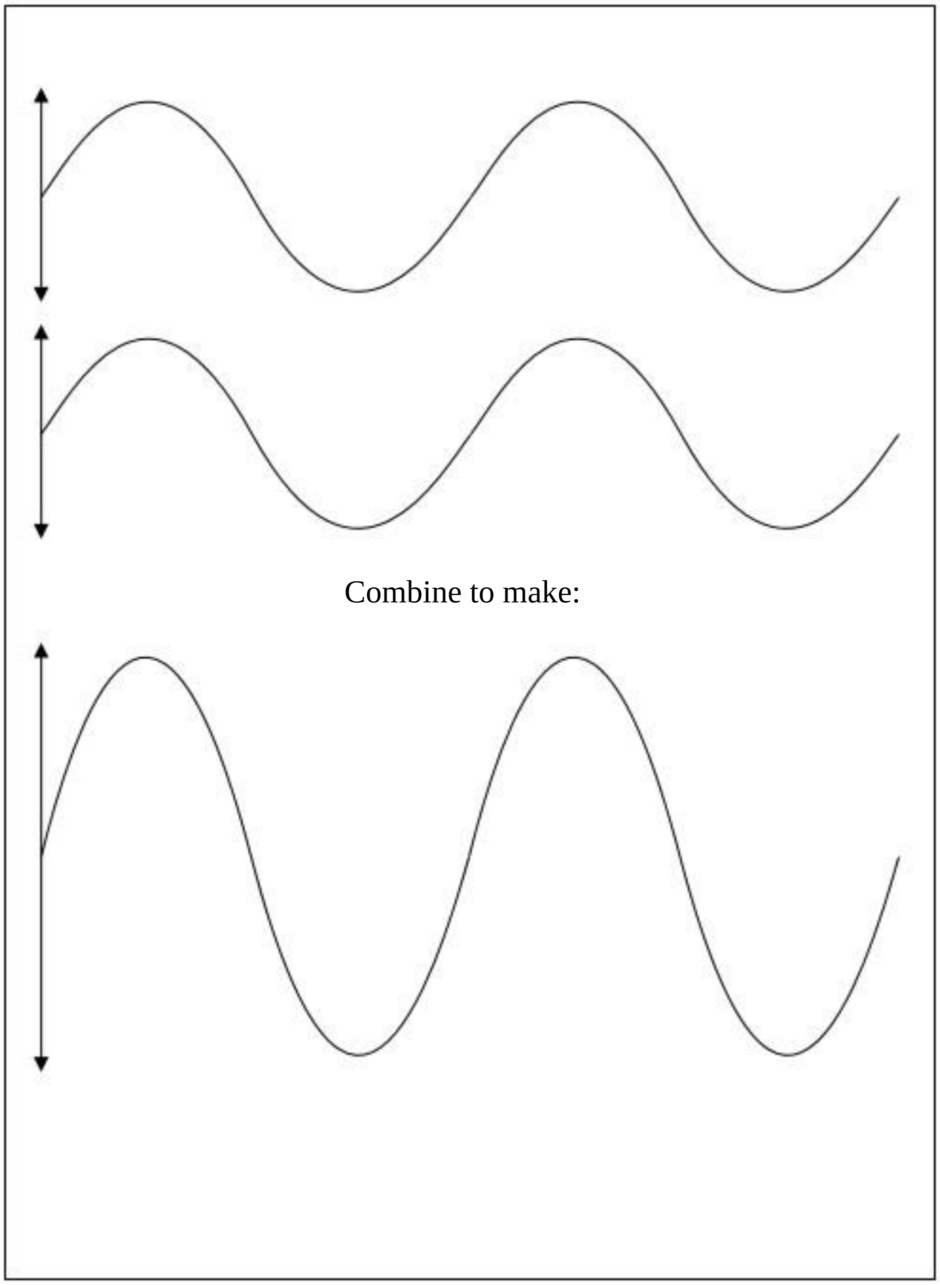
Wireless transmissions also suffer from what are known as "multipath problems" resulting from the fact that terrain and other physical features can create reflections that can cause the same signal to arrive at the same location multiple times. Unless the receiver is able to detect that it is receiving the same signal multiple times, it will perceive multipathing as an increase in the noise floor that reduces the available bandwidth.[171]

When reflections cause the same signal to arrive by different paths, the signal can arrive either in phase (with the peaks and the valleys of the wave form from the same signal arriving at exactly the same time) or out of phase (with the peaks and the valleys of the wave form from the same signal arriving at different times). When waves reflecting off a hard surface arrive in phase, the signal reinforces itself, creating a localized hot spot in which signal is unusually strong.

[170] C. E. Shannon, *A Mathematical Theory of Communication* (pt. 1), 27 BELL SYS. TECH. J. 379 (1948); C. E. Shannon, *A Mathematical Theory of Communication* (pt. 2), 27 BELL SYS. TECH. J. 623 (1948).

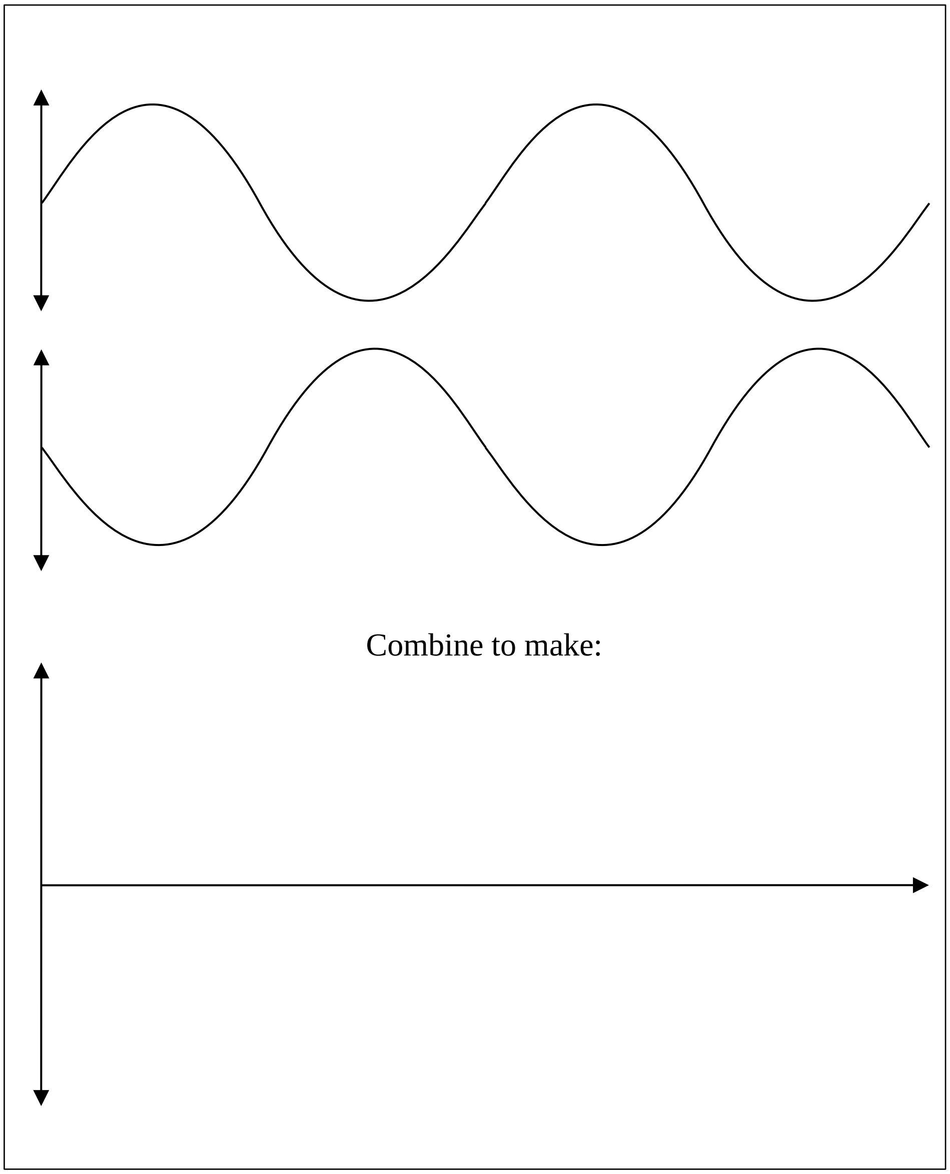
[171] Jørgen Bach Andersen et al., *Propagation Measurement and Models for Wireless Communications Channels*, IEEE COMM., Jan. 1995, at 42.

Figure 1: Reinforcement of Two Wave Forms That Are in Phase



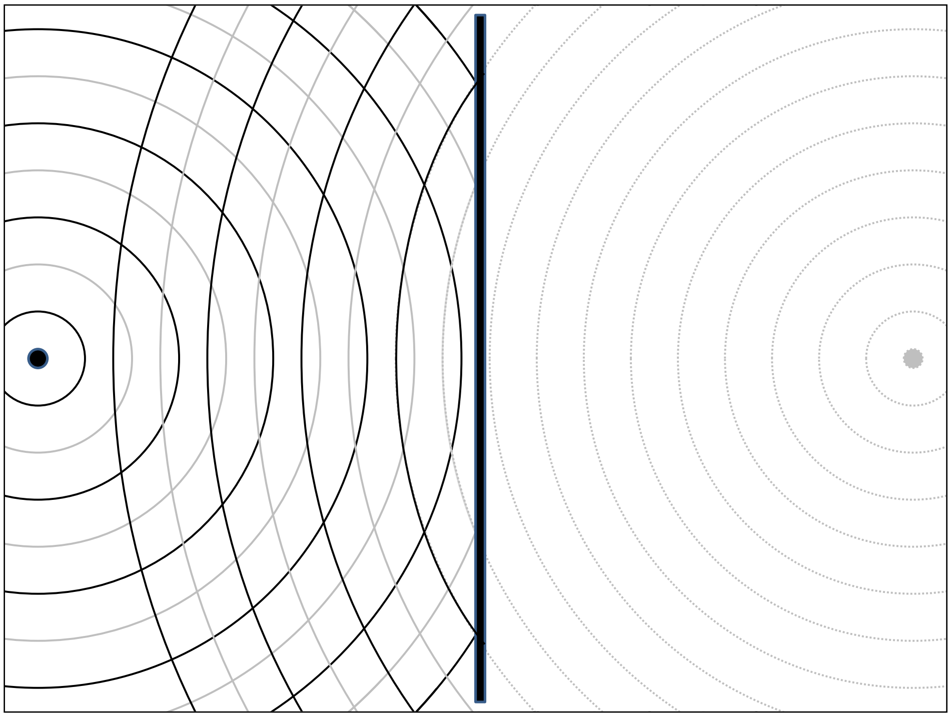
When reflected waves arrive out of phase, they can dampen the signal. When they arrive perfectly out of phase (i.e., 180º out of phase), the reflection can create a dead spot by canceling out the wave altogether. Although smart transmitters and receivers can avoid these problems if they know the exact location of each source and can even use the additional signal to extend the usable transmission range, they cannot do so if the receiver or the other sources are mobile devices whose locations are constantly changing.

Figure 2: Cancellation by Two Wave Forms That Are 180º Out of Phase



A standard result in any physics textbook is that a reflection creates waves that are identical to a point source that is equidistantly located on the other side of the reflective surface and the signal strength is quite unpredictable. Consider the simple diagram in Figure 3, in which the black circles represent the peaks of the wave form, while the grey circles represent the valleys. The points where two black circles or two grey circles cross represent hot spots where signals reinforce one another. The locations where a black circle crosses a grey circle represent dead spots where waves tend to cancel one another out.

Figure 3: Interference Caused by the Reflection of Waves



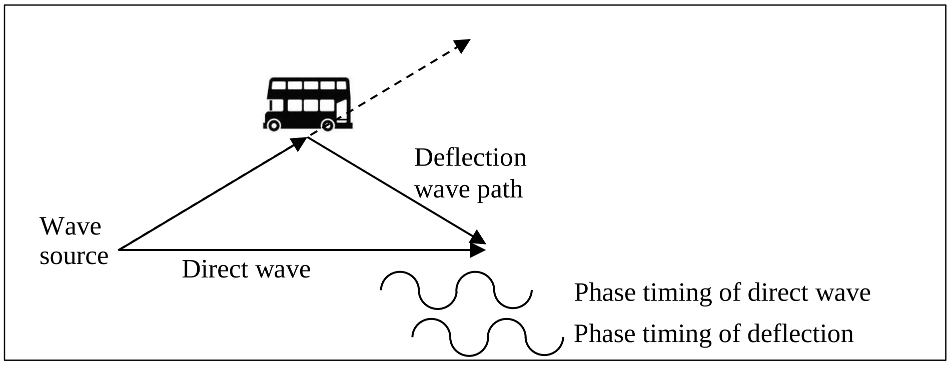
Obviously, individuals traversing a room might pass through a variety of hot and cold spots. In addition, wave reflections can result not only from immobile objects, such as terrain and buildings, but also from mobile objects, such as cars and trucks.[172] The result is that the amount of bandwidth available can change dynamically on a minute-by-minute basis.

A participant at a May 2010 conference held at the University of Pennsylvania related a particularly vivid example of this phenomenon. While living in London, he had an apartment overlooking the famous Speakers' Corner in Hyde Park. Thinking that those in the Speakers' Corner might enjoy having WiFi service, he established a WiFi hotspot and pointed a directional antenna at the location only to find that his signal was intermittently blocked even though nothing ever passed directly between his apartment and the Corner. He eventually discovered that the interference arose whenever a double-decker bus was forced to stop at a

[172] *Id*.

nearby traffic light. Even though the bus did not directly obstruct the waves travelling to and from the Speakers' Corner, it created a multipath reflection that periodically cancelled out the direct signal.[173]

Figure 4: The Problem of Multipath Propagation



The result is that interference from other sources can be quite unpredictable and change rapidly from minute to minute. For these reasons, many wireless providers implement protocols that dynamically manage their networks based on the available bandwidth, giving priority to time-sensitive applications during times when subscribers are in areas of low bandwidth, such as by holding back e-mail while continuing to provide voice service. They have to implement these protocols much more aggressively and dynamically than do wireline providers.

D. IMPLICATIONS OF THE LOWER QUALITY OF SERVICE IN WIRELESS NETWORKS

The difference in the quality of service provided by wireless and wireline networks necessarily requires that the two networks be managed differently. In particular, wireless networks handle error correction and congestion in a manner that is quite different from wireline networks.

1. Error Correction

Wireless networks sometimes run afoul of the standard approach to ensuring reliability on the wireline Internet. The workhorse transport protocol on the Internet known as the Transmission Control Protocol or TCP ensures reliability by calling for every host to set a retransmission

[173] Christian Sandvig, Assoc. Professor of Commc'n, Univ. of Ill., Remarks presented at the Center for Technology, Innovation and Competition's conference on "Rough Consensus and Running Code: Integrating Engineering Principles into the Internet Policy Debates," *How to See Wireless* (May 7, 2010). For a description of the project, see PHILIP N. HOWARD, NEW MEDIA CAMPAIGNS AND THE MANAGED CITIZEN xi–xii (2006).

timer based on the expected round-trip time between the sending host and the receiving host.[174] Receiving hosts are supposed to send acknowledgements for every packet they successfully receive. If the sending host does not receive an acknowledgment when its retransmission timer expires, it resends the packet and repeats the process until it is successfully transmitted.[175]

In many ways, relying on feedback loops and end-to-end retransmission is quite inefficient. Resending packets from the source requires the consumption of significant network resources. In addition, waiting for the retransmission timer to expire can cause significant delays. Such overhead costs become higher as the packet loss rates increase. If loss rates become sufficiently high, it may make sense for networks to employ network-based error recovery mechanisms instead of relying on end-to-end error recovery.

The lower reliability of wireless networks thus can lead system designers to deploy functionality in the core of the network to ensure reliability and error recovery. For example, PRNET employed a network-based reliability system known as forward-error correction.[176] The higher loss rates in wireless technologies also explains why wireless broadband networks are increasingly deploying network-based reliability systems, such as Automatic Repeat reQuest (ARQ), that detect transmission errors and retransmit the missing data from the core without waiting for the host-based retransmission timer to expire and without consuming the additional network resources needed to retrieve the packet all the way from the host.[177] Other techniques that allow routers in the core to participate in the transport layer exist as well.[178]

2. *Congestion Management*

The lack of reliability also requires that wireless technologies employ a significantly different approach to managing congestion. The primary mechanism for controlling congestion on the Internet was developed in the late 1980s shortly after the Internet underwent a series of congestion collapses. As noted earlier, TCP requires that receiving hosts send acknowledgments every time they successfully receive a packet. If the

[174] TANENBAUM & WEATHERALL, *supra* note 98, at 569–70.

[175] *Id.* at 568.

[176] Robert E. Kahn et al., *Advances in Packet Radio Technology*, 66 PROC. IEEE, 1468, 1492 (1978).

[177] KUROSE & ROSS, *supra* note 107, at 207–15; TANENBAUM & WEATHERALL, *supra* note 98, at 222–26.

[178] *See* KUROSE & ROSS, *supra* note 107, at 575–77.

sending host does not receive an acknowledgement within the expected timeframe, it presumes that the packet was lost and resends it.[179] The problem is that the host now has sent twice the number of packets into a network that was already congested. Once those packets also failed to arrive, the host introduced still another duplicate packet. The resulting cascade would bring the network to a stop.

Because congestion is a network-level problem that is the function of what multiple end users are doing simultaneously rather than the actions of any one end user, some proposed addressing it through a network-level solution. This was done in the original ARPANET through networks running asynchronous transfer mode (ATM) and many other early corporate networks.[180] However, the router hardware of the time made network-based solutions prohibitively expensive. On the other hand, hosts can also stop congestion collapse if they cut their sending rates in half or more whenever they encounter congestion. The problem is that congestion is the product of what multiple hosts are doing, whereas any individual host only knows what it is doing. Thus the hosts operating at the edge of the network typically lack knowledge of when the network is congested.

Van Jacobson devised an ingenious mechanism by which hosts operating at the edge of the network can infer when the core of the network has become congested based on the information they were able to see. Jacobson noted that packet loss typically occurs for only two reasons: (1) transmission errors, or (2) discard by a router where congestion has caused its buffer to become full.[181] Because wireline networks rarely drop packets due to transmission errors, hosts operating at the edge of the network could infer that the failure to receive an acknowledgement within the expected time was a sign of congestion. Hosts could then take this as a signal to reduce congestion by slowing down their sending rates exponentially.[182]

However, this inference is invalid for wireless networks. Wireless networks drop packets due to transmission error quite frequently, either because of a bad handoff as a mobile user changes cells or because of the interference problems discussed above. When a packet is dropped due to a

[179] *Id.* at 240.

[180] Raj Jain & K.K. Ramakrishnan, *Congestion Avoidance in Computer Networks with a Connectionless Network Layer: Concepts, Goals and Methodology*, PROC. COMPUTER NETWORKING SYMPOSIUM 134 (1988), http://www.cse.wustl.edu/~jain/papers/ftp/cr1.pdf.

[181] Van Jacobson, *Congestion Avoidance and Control*, 18 ACM SIGCOMM COMPUT. & COMM. REV. 314, 319 (1988).

[182] *Id.*

transmission error, reducing the sending rate exponentially only serves to degrade network performance. Instead, the sending host should resend the dropped packet as quickly as possible without slowing down. In other words, the optimal response for wireless networks may well be the exact opposite of the optimal response for wireline networks.

E. RESPONSES TO THE LOWER QUALITY OF SERVICE IN MOBILE BROADBAND NETWORKS

In short, the deployment of wireless broadband is putting pressure on the traditional mechanisms for managing error correction and congestion, two of the most basic functions performed by the network. The higher loss rates make the traditional approach to error recovery more expensive and make it impossible to regard packet loss as a sign of congestion.

As a result, the engineering community is experimenting with a variety of alternative approaches.[183] One approach allows local recovery of bit errors through some type of forward error recovery.[184] One such solution places a "snoop module" at the base station that serves as the gateway used by wireless hosts to connect to the Internet that keeps copies of all packets that are transmitted and monitors acknowledgments passing in the other direction. When the base station detects that a packet has failed to reach a wireless host, it resends the packet locally instead of having the sending host do so.[185] A second approach calls for the sending host to be aware of when its transmission traverses wireless links. Dividing the transaction into to two internally homogeneous sessions makes it easier to infer the current status of the network.[186] A third approach splits the wireless and the wireline approaches into separate TCP or UDP sessions.[187]

Many of these approaches violate the semantics of TCP, since the packets are not addressed to the receiving hosts. Many of them introduce

[183] KUROSE & ROSS, *supra* note 107, at 576–77.

[184] Ender Ayanoglu et al., *AIRMAIL: A Link-Layer Protocol for Wireless Networks*, 1 WIRELESS NETWORKS 47 (1995).

[185] *See generally* Hari Balakrishnan et al., *Improving Reliable Transport and Handoff Performance in Cellular Wireless Networks*, 1 WIRELESS NETWORKS 469 (1995).

[186] Ajay Bakre & B.R. Badrinath, *I-TCP: Indirect TCP for Mobile Hosts*, 1995 PROC. 15TH INT'L CONF. ON DISTRIBUTED COMPUTING SYS. (ICDCS '95) 136, 137; Hari Balakrishnan et al., *A Comparison of Mechanisms for Improving TCP Performance Over Wireless Links*, 5 IEEE/ACM TRANSACTIONS ON NETWORKING 756, 760 (1997).

[187] Wei Wei et al., *Inference and Evaluation of Split-Connection Approaches in Cellular Data Networks*, PROC. ACTIVE & PASSIVE MEASUREMENT WORKSHOP (2006); Raj Yavatkar & Namrata Bhagwat, *Improving End-to-End Performance of TCP over Mobile Internetworks*, PROC. WORKSHOP ON MOBILE COMPUTING SYS. & APPLICATIONS 146, 147 (1994).

intelligence into the core of the network and violate the principle of avoiding per-flow state. The split connection approach violates the principle of end-to-end connectivity. All of them require introducing traffic management functions into the core of the network to a greater extent than originally envisioned by the Internet's designers.

V. THE HETEROGENEITY OF DEVICES

Starting with Michael Powell's 2004 four freedoms speech, every network neutrality proposal has called for broadband Internet access networks to be open to all legal devices. Indeed, the 2015 Open Internet Order included devices within the no blocking, no throttling, and no paid prioritization rules as well as the catchall prohibiting unreasonable interference and disadvantage.[188]

In stark contrast to the fixed line world, wireless devices are not universally compatible with every network. For example, Verizon's wireless broadband network is based on a protocol known as Evolution-Data Optimized (EV-DO) operating in the traditional cellular portion of the spectrum. Sprint's wireless broadband network also employs EV-DO, but operates in the band of spectrum originally allocated to the second-generation wireless technology known as Personal Communications Services (PCS). AT&T's wireless broadband networks use a different format known as High Speed Packet Access (HSPA). Each has different technical characteristics. Indeed, the greater compatibility of HSPA with the iPhone is part of what led Apple initially to deploy the iPhone exclusively through AT&T.

Instead of relying on a personal computer, wireless broadband subscribers connect to the network through a wide variety of smart phones. These devices are much more sensitive to power consumption than are PCs, which sometimes leads wireless network providers to disable certain functions that shorten battery life to unacceptable levels, for example because they either employ analog transmission or search constantly for an available connection. In addition, wireless devices have much less processing capacity and employ less robust operating systems than do the laptop and personal computers typically connected to wireline services. As a result, wireless devices are more sensitive to conflicts generated by multiple applications, which can cause providers to be much more careful about which applications to permit to run on them. This compels wireless broadband networks to manage devices and applications to a greater extent than wireline networks.

[188] 2015 Open Internet Order, *supra* note 2, at 5607–09 ¶¶ 15–21.

Wireless devices also tend to be much more heterogeneous in terms of operating systems and input interfaces including keyboards and touch screens. As a result, the dimensions and levels of functionality offered by particular wireless devices vary widely. It seems too early to predict with any confidence which platform or platforms will prevail. Furthermore, as noted earlier, many wireless networks address bandwidth scarcity by giving a higher priority to time-sensitive applications, which typically requires close integration between network and device. These features underscore the extent to which variations in particular devices are often an inextricable part of the functionality of the network.[189]

These differences in compatibility and functionality call into question the provisions mandating that all broadband Internet access networks be open to all devices. Simply put, modern wireless devices prioritize traffic on the basis of application and are properly regarded as part of the network's functionality.

VI. ROUTING

Routing on wireless broadband networks is also very different from routing on fixed broadband networks. In particular, mobile broadband networks often exchange traffic with Internet gateways and rely on a legacy telephone technology to deliver traffic to end users instead of treating smartphones as IP-enabled devices. In addition, mobile broadband interferes with both the stability of routing tables and the compactness of the address space. Although potential solutions exist, such as the identity/locator split, they have yet to be implemented. As a result, wireless broadband networks must rely on a suite of protocols known as mobile IP, which introduce a wide range of intelligence into the core of the network in ways that violate the end-to-end argument.

A. THE USE OF INTERNET GATEWAYS

One of the realities of wireless broadband networks is that they introduce a great deal of intelligence into the network in ways that fit less comfortably with the end-to-end argument. Recall that one of the Internet's foundational principles is that each host connected to the Internet must have a unique IP address that is visible and accessible to all other hosts. In addition, all of the routers within the network are supposed to route traffic on the basis of this address.

[189] Charles L. Jackson, *Wireless Efficiency versus Net Neutrality*, 63 FED. COMM. L.J. 445, 476–77 (2011).

It bears mentioning that until recently, wireless networks have not routed traffic in this manner. Unlike devices connected to wireline networks, which have IP addresses that are visible to all other Internet-connected hosts, third-generation wireless devices did not have IP addresses. Instead, Internet connectivity is provided by an IP gateway located in the middle of the network that connects to individual wireless devices using a legacy telephone-based technology rather than IP. This means that for most of their history, wireless devices did not have the end-to-end visibility enjoyed by true Internet-enabled devices and instead connected through a virtual circuit between the Internet gateway and the wireless device. Fourth-generation wireless technologies such as LTE connect through IP. Until 3G is retired, some wireless devices will necessarily connect to the Internet on different and less open terms than devices connected through wireline networks.

This reality means that many wireless broadband devices violate the principle that each device has a unique IP address that is visible to all others. In addition, part of the connection operates using a different address system and employing circuit-based technologies that deviate from the Internet's commitment to store and forward routing. Simply put, traffic bound for and received from wireless devices will not pass through the network on the same terms as traffic going to and from hosts connected to the network through wireline technologies.

B. ACCELERATION IN THE PACE OF CHANGES IN ROUTING ARCHITECTURE

The mobility inherent in wireless broadband networks necessarily requires more frequent updates to routing tables than is the case for fixed broadband networks. Although solutions exist that could simplify this process, both the traditional version of the Internet Protocol, known as IPv4, as well as the new version, known as IPv6, rely on a mobile IP approach that requires a great deal of intelligence in the network.

A key feature of the current routing architecture is that it is updated on a decentralized basis. Every backbone router periodically informs its adjacent neighbors of the best routes by which it can reach every location on the Internet. This means that initially any changes to the network architecture will only be advertised locally. During the next update cycle, routers that have been informed of the change will inform the routers located the next level away. Over time, the information will spread out in all directions until the entire network is aware of the change. When this occurs, the routing table is said to have reached equilibrium.

Before the routing table has reached equilibrium, however, some parts of the network may not know of certain changes that have occurred in other parts of the network. Suppose, for example, that one host in one corner of the network drops off the network. A host in a distant corner will not find out about that for quite some time. In the meantime, it could keep sending packets to a host that is no longer there, which wastes resources and unnecessarily adds to network congestion.

The efficient functioning of the network thus depends on the routing architecture being able to reach equilibrium. Whether it does so is largely a function of the speed with which locations change compared to the speed with which information about that change can propagate through the entire network. Moreover, the current architecture is built on the implicit assumption that Internet addresses change on a slower timescale than do communication sessions. So long as the address architecture changes at a slower timescale, any particular Internet-based communication may take the address architecture as given.

Mobility, however, increases the rate at which the address architecture changes. In addition, because addressing is handled on a decentralized basis, information about changes in the address architecture takes time to spread across the Internet. Increases in the rate with which the address space changes can cause communications sessions to fail and create the need for a new way to manage addresses.

C. COMPACTNESS OF THE ADDRESS SPACE

As a separate matter, wireless technologies are also causing pressure on the way the amount of resources that the network must spend on keeping track of Internet addresses. To understand why this is the case, one must keep in mind that routers typically follow one of two strategies in keeping routes. Some routers keep *global routing tables* that identify the outbound link that represents the most direct path to every single host on the Internet. Other routers avoid the burden of maintaining complete routing tables by only keeping track of a limited number of paths. All traffic bound for locations for which this router does not maintain specific information is sent along a *default route* to a *default router*, which is responsible for identifying the route for delivery of all other traffic to its final destination.

The presence of default routes in a routing can give rise to a potential problem. For example, routers using default routes could point at one another, either directly or in a loop, which would cause the packets to pass back and forth indefinitely. The Internet ensures that traffic does not travel indefinitely through the network by assigning a *time to live* to each packet that limits the total number of hops that any packet may traverse before dropping off the network. Eventually, any packet caught in such a cycle will reach its maximum and drop off the network.[190]

The best way to prevent such roads to nowhere is to ensure that at least some actors maintain global routing tables, which by definition are routing tables that do not include any default routes. This role is traditionally played by the major backbone providers, known as Tier 1 ISPs. More than the economic relationships (such as peering), many regard the maintenance of default free routing tables as the defining characteristic of Tier 1 ISPs.[191]

Sustaining a global routing table that maintained a separate entry for the best path to every location on the Internet has proved to be very difficult. The expansion of the Internet meant that the size of the routing table grew at a very fast rate. In fact, it grew faster than the routers could keep up.[192]

The solution was an innovation called Classless InterDomain Routing (CIDR).[193] For our purposes, the important aspect of CIDR is that it allowed routers to use "route aggregation" to prevent routing tables from growing out of control. This mechanism can be illustrated by analogy to the telephone system. Consider an individual in Los Angeles who attempts to call the main telephone number for the University of Pennsylvania, which is (215) 898-5000. So long as all phones in the 215 area code are located in Philadelphia, a phone switch in Los Angeles could represent all of the telephone numbers in that area code ((215) xxx-xxxx) with a single entry in its routing table. Indeed, one can think of the millions of telephone numbers in the 215 area code as lying within the cone of telephone numbers represented by that entry.

Similarly, so long as all telephone numbers in the 898 directory within the 215 area code are connected to the same central office, switches within Philadelphia need not maintain separate entries for each phone number in that directory. Instead, they can represent the cone of all ten thousand telephone numbers located in (215) 898-xxxx with a single entry.

[190] Paul Milgrom et al., *Competitive Effects of Internet Peering Policies*, *in* THE INTERNET UPHEAVAL 175, 179–80 (Ingo Vogelsang & Benjamin M. Compaine eds., 2000).

[191] Peyman Faratin et al., *The Growing Complexity of Internet Interconnection*, 72 COMM. & STRATEGIES 51, 54 (2008).

[192] Geoff Huston, *Analyzing the Internet BGP Routing Table*, 4 INTERNET PROTOCOL J., Mar. 2001, at 2, 3, http://ipj.dreamhosters.com/wp-content/uploads/issues/2001/ipj04-1.pdf.

[193] Yoo, *supra* note 117, at 82.

CIDR adopts a similar strategy to reduce the size of the routing tables maintained by Tier 1 ISPs. For example, the University of Pennsylvania has been assigned all of the addresses in the 128.91.xxx.xxx prefix (covering 128.91.0.0 to 128.91.255.255). Various locations have individual addresses falling within this range, with the main website for the University of Pennsylvania being covered by 128.91.34.233 and 128.91.34.234. Assuming that all of the hosts associated with these IP addresses are located in the same geographic area, a Tier 1 ISP could cover all of the one million addresses within this prefix with a single entry.

The success of this strategy depends on the address space remaining compact. In other words, this approach will fail if the 215 area code includes phone numbers that are not located in Philadelphia. If the telephones associated with those numbers sometimes lie outside the Philadelphia area, the telephone company will have to maintain separate entries in its call database for all phones located outside the area. Similarly, if some hosts with the 128.91.xxx.xxx prefix reside outside the Philadelphia area, Tier 1 ISPs will have to track those locations with additional entries in their routing tables.

The advent of mobile telephony and mobile computing means, of course, that telephones and laptops will often connect to the network outside their home locations. This in turn threatens to cause the routing tables to grow faster again. Other developments, including multihoming, the use of provider-independent addresses, and the deployment of IPv6, are further reducing the compactness of the routing table.

D. THE IDENTITY/LOCATOR SPLIT

A solution does exist that would not require introducing intelligence into other parts of the network to accommodate routing. This solution is known as the identity/locator split.[194] The idea gained new impetus by the Report from the Internet Architecture Board (IAB) Workshop on Routing and Addressing, which reflected a consensus that such a split was necessary.[195] The International Telecommunication Union (ITU) has also embraced the need for the ID/locator split in Next Generation Networks

[194] For an early statement, see Jerome H. Saltzer, *On the Naming and Binding of Network Destinations* (Aug. 1993), http://tools.ietf.org/pdf/rfc1498 (IETF Network Working Group Request for Comments no. 1498) (identifying the potential need for separate names for nodes and network attachment points).

[195] David Meyer et al., *Report from the IAB Workshop on Routing and Addressing* 22–23 (Sept. 2007), http://tools.ietf.org/pdf/rfc4984 (IETF Network Working Group Request for Comments no. 4984).

(NGNs).[196] Additionally, it is the focus of a major research initiative sponsored by the National Science Foundation's Future Internet Architecture Program.[197]

The proposal is based on the insight that an IP address currently plays two distinct functions. It simultaneously serves as an *identifier* that identifies a machine, and it serves a *locator* that identifies where that machine is currently attached to the network topology. When all hosts were connected to the Internet via fixed telephone lines, the fact that a single address combined both functions was not problematic. The advent of mobility caused the unity of identity and location to break down. A single mobile device may now connect to the network through any number of locations. Although the network could constantly update the routing table to reflect the host's current location, doing so would require propagating the updated information to every router in the network as well as an unacceptably large number of programs and databases.

Others have proposed radical changes in the addressing and routing architecture. One approach would replace the single address now employed in the network with two addresses: one to identify the particular machine and the other to identify its location.[198] Others criticize such proposals as unnecessarily complicated.[199]

If deployed, the identity/locator split would represent a radical deviation from the existing architecture. Whatever solution is adopted would represent a fundamental change in the network layer than unifies the entire Internet. It would require a change in the way we approach routing and addressing and require reconfiguring every device attached to the network. If implemented, it would eliminate some of the asymmetries in the way that routing to mobile hosts is done. To date, however, the identity/locator split has not yet been implemented, and any future

[196] INT'L TELECOMM. UNION, TELECOMM. STANDARDIZATION SECTOR, RECOMMENDATION ITU-T Y.2015: GENERAL REQUIREMENTS FOR ID/LOCATOR SEPARATION IN NGN (2009), http://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-Y.2015-200901-I!!PDF-E&type=items.

[197] MobilityFirst Future Internet Architecture Project, *supra* note 11.

[198] *See* Chakchai So-In, *Virtual ID: ID/Locator Split in a Mobile IP Environment for Mobility, Multihoming and Location Privacy for the Next Generation Wireless Networks*, 5 INT'L J. INTERNET PROTOCOL TECH. 142 (2010) (surveying alternative approaches to the ID/locator split).

[199] *See, e.g.*, Dave Thaler, Keynote Address at the 3rd ACM International Workshop on Mobility in the Evolving Internet Architecture (MobiArch 2008): Why Do We Really Want an ID/Locator Split Anyway? (Aug. 22, 2008), http://conferences.sigcomm.org/sigcomm/2008/workshops/mobiarch/slides/thaler.pdf.

implementations would require an extended transition time during which networks would have to operate both modes.

E. MOBILE IP

Instead of relying on solutions that would have kept the network simple, the modern Internet relies on a complex system of protocols operating in the core of the network to accommodate mobility. The most straightforward approach to addressing mobility would be to assign a mobile host a new IP address whenever it changes location. This would put significant strain on the network by requiring that it inform the rest of the network about the change. To the extent that it disrupts the compactness of the address space, it may create additional pressure on the routing architecture by causing the routing table to grow. In addition, dynamically changing IP addresses in the middle of an application may cause many applications to fail.[200]

How, then, do we handle mobility without having to update the routing tables constantly and without causing the size of routing tables to grow out of control? The Internet currently solves these problems through a regime known as *mobile IP*. Under mobile IP, each mobile user has a *home network*, with all other networks labelled *foreign networks*. The mobile host designates a router located on its home network as the contact point for all IP-based communications directed to the mobile host. This contact point is called the *home agent*. Anyone seeking to contact the mobile host, called the *correspondent*, simply sends the packets to the home agent, which then forwards the communication to the mobile host. If the mobile host moves from one foreign network to another, it simply notifies its home agent, which then routes any new packets it receives to the new location.

Although this solution sounds relatively simple, actually implementing it can be quite complex. For example, the home agent has to know to where the mobile host is currently located. This is relatively easy when the mobile host initiates the transaction. It is more complicated when a third party is attempting to contact the mobile host. Stated in the example of mobile telephony, networks can easily discover where a particular cellular user is located when it is that user that is initiating the call. The simple fact of establishing contact with the local microwave tower announces the location. The situation is different when the mobile user is receiving the call. To terminate this call, the network has to know where the mobile user is even when it is simply sitting around waiting.

[200] PETERSON & DAVIE, *supra* note 104, at 290.

This means that if a mobile host is to receive traffic, it must constantly announce its presence to the network serving its current location so that the network knows that it is there. This can be accomplished by designating a router located on the foreign network as the *foreign agent* responsible for managing mobile IP. Every mobile host must regularly register with the foreign agent serving the local foreign network in order to receive communications. This can happen by the foreign agent sending an advertisement notifying mobile nodes located in its service area that it is prepared to facilitate mobile IP or by the mobile node sending a solicitation to see if any foreign agents are located nearby capable of supporting mobile IP. Once a foreign agent registers the presence of a mobile host, it must then notify the home agent about the mobile host's current whereabouts so that the home agent knows where to forward any packets that it receives. Mobile IP works best if mobile nodes deregister when they leave the foreign network.

So how does the home agent send the packets to the foreign agent for delivery? It could alter the IP address contained in the packet. But as Cerf and Kahn noted, doing so is prone to errors and risks making the communication non-transparent to the sending host. Instead, the home agent encapsulates these packets in another IP packet addressed to the foreign agent where the mobile host is currently located. That way the application receiving the datagram does not know that the datagram was forwarded by the home agent. Once the foreign agent decapsulates the packet, it cannot simply send it to the address contained in the IP header. That would cause the packets to be routed back to the home network. Instead, it checks to see if the packets are addressed to a mobile host that has registered locally and routes the packets to the mobile host.

Mobile IP thus requires that the network perform three distinct functions:

- A protocol by which mobile nodes can register and deregister with foreign agents.
- A protocol by which foreign agents can notify home agents where the mobile node is currently located.
- Protocols for home agents and foreign agents to encapsulate and decapsulate datagrams they receive.

Unfortunately, this approach suffers from a number of well-known inefficiencies and issues relating to security, handoffs and triangle routing.

1. *Security*

The ability to register from remote locations raises major security concerns. For example, a malicious user could attempt to mislead the home agent into thinking it was the proper recipient. If so, it could receive all of the packets addressed to the IP address.[201] Although the architects considered making security a basic feature of IPv6, they eventually decided against doing so.

2. *Handoffs*

Mobile IP also must find a way to manage the network when a mobile host moves from one base station to another. One solution is to update the home agent. Any tardiness in the update can cause packets to become lost. Another solution is to designate the first foreign agent in a particular transaction as the *anchor foreign agent* that will be the location where the home agent will send all packets. Should the mobile host shift to a different foreign network, the anchor foreign agent can forward the packets to the new location.

3. *Triangle Routing*

By envisioning that all traffic will travel to the home agent and then be forwarded to the foreign agent, mobile IP employs a form of indirect routing that can be very inefficient. For example, when a person with a home network located in Philadelphia travels to Los Angeles, any packets sent to her while she is in Los Angeles will have to travel across the country to the home agent located in Philadelphia and then be rerouted back to Los Angeles. This can result in the inefficiency of what is sometimes called "triangle routing."[202]

The home agent can eliminate triangle routing by passing the mobile host's current location on to the sender so that the sender may forward subsequent packets to it directly. The initial communications must still bear the inefficiency of triangle routing. Moreover, such solutions become much more difficult to implement if the mobile agent is constantly on the move.[203] The network must have some way to notify the correspondent that the mobile host has changed location. The usual solution is that much as the home network and the foreign network have agents, the correspondent attempting to contact the mobile host also has a *correspondent agent*. The correspondent agent queries the home agent to learn the location of the mobile host. It then encapsulates the datagram in a new datagram addressed to the foreign agent. The foreign agent then

[201] *See* KUROSE & ROSS, *supra* note 107 at 556; PETERSON & DAVIE, *supra* note 104, at 294; TANENBAUM & WEATHERALL, *supra* note 98, at 488.

[202] PETERSON & DAVIE, *supra* note 104, at 293.

[203] COMER, *supra* note 104, at 339–46; KUROSE & ROSS, *supra* note 104, at 559–63; TANENBAUM & WEATHERALL, *supra* note 98, at 386–89, 485–88.

decapsulates the new datagram and passes the original datagram to the mobile host.

The problem arises if the mobile host moves from one foreign network to another. Under indirect routing, the mobile host simply notifies its home agent of the change of location. Under direct routing, however, the correspondent agent is responsible for encapsulating datagrams and forwarding them to the mobile host, not the home agent. At this point, the mobile node needs a way to update the correspondent agent as to its new location. This in turn requires two more protocols:

- A protocol by which correspondent agents can query the home agent as to the mobile node's current location.
- A protocol by which the mobile host that changes foreign networks can notify the correspondent agent about its new location.

The additional complexity is sufficiently difficult to implement that direct routing was not included in the upgrade to IPv6. The net result is that modern mobile broadband networks employ far more intelligence in their core than the end-to-end argument would suggest.

VII. CONCLUSION

The limited ability to add more spectrum and the absolute limit to density of people who can use wireless phones in the same location means that mobile broadband networks must manage their traffic much more aggressively than fixed broadband networks. As noted above, wireless networks often prioritize time-sensitive applications such as voice over non-time-sensitive applications such as email. In addition, certain solutions, such as the one being advanced by T-Mobile's Binge On, may reduce network congestion, but must do so in an application-specific manner. Bad handoffs, local congestion, and the physics of wave propagation necessitate that mobile broadband networks are subject to highly variable quality of service that requires introducing greater intelligence into the network. The greater heterogeneity of devices and differences in networking standards in the mobile broadband world also limits the feasibility of the prohibition against blocking or throttling devices. Finally, the greater complexity of routing in wireless networks requires introducing a greater degree of intelligence in the core of the network.

The net result is that mobile wireless broadband networks operate on principles that are quite different from those governing the rest of the Internet. Bandwidth limitations require that wireless providers manage their networks more intensively than those operating networks based on other technologies. Because many smartphones do not have IP addresses and wireless networks suffer higher rates of packet loss than fixed networks, wireless broadband networks need to employ virtual circuits and embed intelligence in the network to a greater extent than fixed broadband networks. The unpredictability of signal strength resulting from the physics of wave propagation can necessitate more extensive supervision than other technologies require, as do the realities of system conflicts and power consumption. Lastly, mobility is placing pressure on the routing and addressing space that may soon require more fundamental changes. The industry has not yet reached consensus on the best approach for addressing all of these concerns. In its consideration of regulatory interventions, the FCC must be careful to create a regime that takes these differences into account.