EDITOR'S NOTE

Welcome to the second Issue of the sixty-third Volume of the *Federal Communications Law Journal*, the nation's premier communications law journal and the official journal of the Federal Communications Bar Association. The *Journal* staff is excited to present the Symposium Articles and the Notes in this Issue.

The Issue begins with a series of Articles presenting analysis of the intersection between engineering principles in the Internet and broadband policy. The Articles included here are the result of a Symposium that took place at the University of Pennsylvania's Center for Technology, Innovation and Competition on May 6–7, 2010. The conference, "Rough Consensus and Running Code: Integrating Engineering Principles into the Internet Policy Debates," brought together a number of engineering, policy, regulatory, academic, and Internet experts to discuss the architecture of the Internet and the interaction between that technical structure and broadband policy.

Christopher Yoo, professor of law at the University of Pennsylvania, provides an introduction to the following pieces, as well as an overview of the presentations made at the conference last May. He emphasizes the discussions' focus on the technical considerations of the Internet and the important role of such considerations on the formation of broadband policy.

The first Article in the Symposium series, which represents a sampling of the presentations from the conference, is by Marjory Blumenthal, associate provost and academic at Georgetown University, and David Clark, senior research scientist at the MIT Computer Science and Artificial Intelligence Laboratory. They present an expansion on the discussion of the end-to-end argument, focusing on the role of trust in decision making with respect to applications that use the Internet. Through this interpretation of the end-to-end argument, they emphasize the importance of the end user's control over trust decisions.

Next, Andrea Matwyshyn, assistant professor of Legal Studies and Business Ethics at the Wharton School at the University of Pennsylvania, presents a discussion of the ways users interact with technology. She frames her argument in developmental psychology and discusses the implications of these interactions on data privacy law. Professor Matwyshyn calls for user resilience on the Internet as a basis for a more secure information technology marketplace.

Dirk Grunwald, the Wilfred and Caroline Slade Endowed Professor in the Department of Computer Science, the Department of Electrical and Computer Engineering, and the Interdisciplinary Telecommunications Program at the University of Colorado, next tackles the network neutrality debate. In doing so, he explores the architecture of "cloud computing" and other services that can enable competition for services, content, and innovation. He argues for a cautious approach to the regulation of potentially anti-competitive practices in the Internet.

In the next Article, Charles Jackson, an electrical engineer and adjunct professor at The George Washington University who has extensive experience in communications and wireless, presents a discussion of congestion and congestion control in the Internet. Dr. Jackson discusses the potential role for priority routing in wireless and the impact of regulation on such priority routing. He emphasizes the possibility of increased efficiency and decreased costs through the implementation of priority routing and other forms of congestion control.

In the final Symposium Article, Christian Sandvig, associate professor of communication at the University of Illinois at Urbana-Champaign, turns the discussion to spectrum and the potential impact of the open spectrum model on wireless telecommunications and communication law. Professor Sandvig explores the subject through case studies involving private entrepreneurs and competing Wireless Internet Service Providers, providing theories as to what these case studies mean for the future of open spectrum regimes.

The Issue then shifts to our Notes, written by third-year members of the *Journal* staff. The first Note is authored by myself, and focuses on shifts in the media landscape and how those shifts impact the access to media test in defamation law. I specifically focus on the ubiquity of social networking in today's media environment and the effect on the distinction between public and private figures. Next, Alicia Sanders offers an argument regarding the Computer Fraud and Abuse Act and the unique nature of e-book purchases. She argues that the CFAA offers consumers protections that are tailored to the unique nature of e-book ownership, which amounts to a bargained-for set of rights in a file. Finally, Joshua Robare takes a close look at television accessibility for the visually impaired. Mr. Robare argues that new legislation or increased funding would afford video descriptions the same wide acceptance and use as closed captioning.

The Editorial Board would like to express its appreciation to the Authors for their contributions and cooperation throughout the editing process. We would also like to extend our gratitude to the Federal Communications Bar Association for its continued support and guidance. Finally, we would like to thank the editors and staff of the *Journal* who worked diligently over these past few months to prepare these Articles and Notes for publication.

The *Journal* is committed to providing its readership with broad coverage of interesting and important communications topics, and we sincerely appreciate the continued support of contributors and readers alike. As always, we welcome your feedback or submissions—any questions or comments you might have about this Issue or our future issues can be sent to fclj@indiana.edu, and any submissions for consideration in our future issues should be sent to fcljsae@indiana.edu. Finally, this Issue and past issues can be accessed at http://law.indiana.edu/fclj.

Ann E. O'Connor
Editor-in-Chief

FEDERAL COMMUNICATIONS LAW JOURNAL

VOLUME 63 2010–2011 NUMBER 2

Published by the INDIANA UNIVERSITY MAURER SCHOOL OF LAW
and the FEDERAL COMMUNICATIONS BAR ASSOCIATION

Federal Communications Law Journal

The *Federal Communications Law Journal* is co-published by the Federal Communications Bar Association and the Indiana University Maurer School of Law. The *Journal* publishes three issues per year, including articles, student notes, commentaries, and book reviews examining a wide range of domestic and international communications and information issues, including telecommunications, the First Amendment, broadcasting, telephony, computers, mass media, intellectual property, communications and information policy making, and related fields.

As the official journal of the Federal Communications Bar Association, the *Journal* is distributed to the Association's more than 2,000 members and almost 400 additional legal practitioners, industry experts, government officials, and academics. The *Journal* is also distributed by Westlaw and Lexis and is available on the Internet at http://www.law.indiana.edu/fclj.

The *Journal* is managed by a student Editorial Board, in cooperation with the Law Journal Committee and Editorial Advisory Board of the FCBA, and a Faculty Advisor.

Federal Communications Bar Association

The Federal Communications Bar Association is a nonprofit member-supported organization of attorneys and other professionals, including engineers, consultants, economists, and government officials, involved in the development, interpretation, and practice of communications law and policy. The FCBA promotes fairness and efficiency in the development and application of communications law and policy at all levels of government; excellence and integrity in the profession; education and training for those involved in communications law and policy; and equality of opportunity in the profession of law.

Founded in 1936, the FCBA has more than 2,500 members, the majority of whom are lawyers who practice in the metropolitan Washington, D.C. area. The FCBA's roster also includes members from almost all of the fifty states, several territories, and many foreign countries. The FCBA is represented as an affiliated organization in the House of Delegates of the American Bar Association.

The FCBA regularly conducts educational seminars, which apprise members of legal, technological, and policy developments in communications and related fields. The FCBA also monitors and reviews legislative, agency, and court developments relevant to the practice of communications law and makes submissions to various government agencies on matters of interest to the membership.

Indiana University Maurer School of Law

The Indiana University Maurer School of Law, founded in 1842, prepares students from the United States and foreign countries for careers as lawyers—in private practice, government, business, and other national organizations—academics, and other professionals. The Law School's over 670 full-time J.D. students are drawn from more than 190 undergraduate colleges and universities throughout the United States and abroad.

Located on the main campus of Indiana University, one of the nation's largest public universities, the Maurer School of Law works closely with other schools and departments, including the Schools of Business, Journalism, and Public and Environmental Affairs. The Law School's library is one of the largest in the country, with more than 725,000 volumes, and nine full-time librarians. The library's collection is accessible through an online catalog, which also lists the 11.2 million holdings in the Indiana University libraries. The library is a U.S. government depository as well as one of only eleven libraries to receive copies of all U.S. Supreme Court briefs.

The Maurer School of Law offers a broad curriculum in communications and information law, including courses in print and electronic media, the First Amendment, information law and policy, copyright, patent, trademark, international telecommunications, internet law, and related fields. Students may also take courses in Indiana University's nationally ranked Telecommunications Department and School of Journalism.

ADMINISTRATIVE OFFICERS

Michael A. McRobbie, *President of the University*
Kenneth R.R. Gros Louis, *Chancellor*
Lauren K. Robel, *Dean and Val Nolan Professor of Law*
Hannah L. Buxbaum, *Executive Associate Dean and Walter W. Foskett Professor of Law*
Leonard D. Fromm, *Associate Dean*
Mark S. Hilycord, *Assistant Dean*
Frank Motley, *Assistant Dean*
Michael Keller, *Assistant Dean*
Colleen K. Pauwels, *Law Library Director and Associate Professor*

The *Federal Communications Law Journal* is published three times a year by the Indiana University Maurer School of Law and the Federal Communications Bar Association. Editorial and general offices are located at 211 South Indiana Avenue, Bloomington, Indiana 47405; mailing address is the *Federal Communications Law Journal*, Indiana University Maurer School of Law, 211 South Indiana Avenue, Bloomington, Indiana, 47405. The *Federal Communications Law Journal* can also be contacted via e-mail at fclj@indiana.edu. Address all correspondence with the Federal Communications Bar Association to the Federal Communications Bar Association, 1020 19th Street, N.W., Suite 325, Washington, D.C., 20036-6101.

Subscriptions: Subscriptions are $30.00 per year (Domestic), $40.00 per year (Canada and Mexico), and $50.00 per year (International). Subscriptions are to be paid in U.S. currency. Subscriptions are accepted only on a volume basis, starting with the first issue. All subscriptions will be renewed automatically unless the subscriber provides timely notice of cancellation. Address changes must be made at least one month before publication date. Please provide an old mailing label or the entire old address. Address changes or other requests for subscription information should be directed to the Executive Editor.

Single and Back Issues: Each issue of the current volume of the *Journal* can be purchased for $15.00 (Domestic, Canada, and Mexico), and $20.00 (International), paid in U.S. currency, from the *Federal Communications Law Journal* (check must accompany order). For back issues from Volumes 1 through 62, inquire of William S. Hein & Co., Inc., 1285 Main Street, Buffalo, New York 14209, (800) 828-7571.

Manuscripts: The *Journal* invites the submission of unsolicited articles, comments, essays, and book reviews. Manuscripts cannot be returned unless a self-addressed, postage-paid envelope is submitted with the manuscript.

Production: The citations of the *Journal* conform to the *Bluebook: A Uniform System of Citation* (19th ed. 2010), copyright by the *Columbia*, *Harvard*, and *University of Pennsylvania Law Reviews* and *The Yale Law Journal*. Variations exist for purposes of clarity and at the editors' discretion. The *Journal* is printed by Western Newspaper Publishing Co., in Indianapolis, Indiana.

Citation: Please cite this issue as 63 FED. COMM. L.J. __ (2011).

FEDERAL COMMUNICATIONS LAW JOURNAL

Formerly
FEDERAL COMMUNICATIONS BAR JOURNAL

VOLUME 63 MARCH 2011 NUMBER 2

Policy debates about the evolution of the Internet show varying degrees of understanding about the underlying technology. A fundamental principle of the design of the Internet, from the early 1980s, is the so-called "end-to-end argument" articulated in a seminal technical paper. Intended to provide guidance for what kind of capability is built into a network as opposed to the devices that use the network, the end-to-end argument has been invoked in discussions about "freedom," "neutrality," and other qualities that may be associated with the supply and use of the Internet and with related public policy. This Article builds on the technical discussions of end-to-end to address the design of applications that use the Internet. It explores the role of trust as a factor in decisions about the structure of applications and their interaction with the Internet as part of a larger system.

In the discourse on communications and new media policy, the average consumer—the user—is frequently eliminated from the equation. This Article presents an argument rooted in developmental psychology theory regarding the ways that users interact with technology and the resulting implications for data privacy law. Arguing in favor of a user-centric construction of policy and law, the Author introduces the concept of resilience. The concept of resilience has long been discussed in terms of the structure of technology systems themselves; but, the resilience of the human users of these systems—though equally if not more important to their functioning—has been neglected. The goal of fostering user resilience should be explicitly included in the discourse on technology policy with respect to data privacy and information security; a

a variety of informal arrangements there, including jamming and extortion. It concludes that wireless may be increasingly subject to extralegal allocation, and the outcomes of federal spectrum policy may in fact rest in local hands.

Notes

Plaintiffs' access to media has long been a factor in defamation cases, enabling courts to determine whether that plaintiff is a public figure who must meet the actual malice standard, or whether that plaintiff is a private figure worthy of greater protection from defamation. This component of the public-private distinction can no longer be applied with clear precision, given the advent of social networking and today's world of widespread media access. In light of the massive changes that have taken place in the media world, the access to media test must be revisited and appropriately retailored to avoid an inappropriate assessment of an otherwise private figure's social networking capabilities. This Note explores the history of the access to media test and the rise of social networking in today's media landscape, and argues a reconfiguring of the test is the only way to continue to draw the distinction between public figures and private figures that the Supreme Court originally envisioned—and that it deemed so important. Despite an individual's excessive number of Facebook friends or Twitter followers, a presence on social networking sites does not necessarily equate to a visible presence in the media as imagined by the *Gertz* Court.

In 2009, Amazon.com decided to correct a potential copyright violation by deleting e-books by George Orwell and Ayn Rand from the Kindles of users who had already purchased the offending texts. Two of those users, Justin Gawronski and Antoine Bruguier, claimed that Amazon.com had violated the Computer Fraud and Abuse Act (CFAA) by accessing their Kindles without authorization. The plaintiffs also relied on other causes of action, including breach of contract and trespass to chattels. Although the dispute quickly settled, the Gawronski lawsuit remains a useful case study that shows why the CFAA is a useful protection for consumers. Recently, courts have begun to restrict the application of the CFAA to cases of computer hacking instead of its more expansive applications in employment law. If the statute were restricted along these lines, consumers would lose the unique protections of the CFAA. As it currently exists, the CFAA provides several advantages to consumers that other causes of action do not. First, the CFAA provides a way for consumers to access federal courts, which can ensure a more uniform treatment of Internet-based contracts than does state law. Furthermore, the CFAA also has the conceptual advantage of conceiving of e-book ownership as a bargained-for set of rights in a file. This concept more accurately reflects the reality of the e-book market than common law approaches. To take further advantage of these benefits, a revision of the CFAA expressly creating a cause of action for tethered e-book readers should be added. Such an amendment would prevent companies from attempting to contract around the CFAA.

Video descriptions allow people who have visual impairments to get the full benefits from television. Through voiceovers those who have problems seeing are told what is happening on screen allowing them to get the most out of viewing television. However, the Federal Communications Commission currently lacks the authority to require broadcasters to create video descriptions for their programs following the decision in *Motion Picture Association of America, Inc. v. Federal Communications Commission*. This situation contrasts with closed caption which allows viewers with hearing problems read the dialog being said on screen. The FCC retained the power to regulate closed captions and as a result they are widely used. Many of the court's reasons in *Motion Picture Association of America* are no longer compelling as a result of digital television transition. Video descriptions can become as widely used as closed captioning as a result of new legislation or increased funding.

SYMPOSIUM INTRODUCTION

Rough Consensus and Running Code: Integrating Engineering Principles into Internet Policy Debates

Christopher S. Yoo*

On May 6–7, 2010, the University of Pennsylvania's Center for Technology, Innovation and Competition hosted the conference, "Rough Consensus and Running Code: Integrating Engineering Principles into the Internet Policy Debates."[1] This conference brought together members of

* Professor of Law, Communication, and Computer & Information Science and Founding Director of the Center for Technology, Innovation and Competition at the University of Pennsylvania. Special thanks to David Clark, Jonathan Smith, and Anna Gavin for their help in putting this conference together and to the staff of the *Federal Communications Law Journal* for their willingness to publish this special conference issue.

1. The full program and video of the panels are available at *Rough Consensus and Running Code: Integrating Engineering Principles into the Internet Policy Debates*, CENTER FOR TECH., INNOVATION & COMPETITION (2010), http://www.law.upenn.edu/cf/institutes/ctic/conferences/internetpolicy.html.

the engineering community, regulators, legal academics, and industry participants in an attempt to provide policymakers with a better understanding of the Internet's technical aspects and how they influence emerging issues of broadband policy.

At various points during the recent debates over broadband policy, observers both inside and the outside the government have acknowledged that the debate has yet to reflect a full appreciation of the engineering principles underlying the Internet and the technological opportunities and challenges posed by the existing architecture. The level of discourse is reminiscent of the days when economic arguments first began to be advanced in during regulatory proceedings, when participants in policy debates lacked a sufficient vocabulary and an understanding of the underlying intuitions to engage in a meaningful discourse about the relevant insights.

The conference's title, "Rough Consensus and Running Code,"[2] also emphasizes that network engineering has long been a pragmatic rather than a theoretical discipline that does not lend itself to abstract conclusions. Network engineers recognize that there is no such thing as the perfect protocol. Instead, optimal network design varies with the particular services, technologies, and flows associated with any particular scenario. In other words, network engineering is more about shades of gray than absolutes, with any solution being contingent on the particular circumstances and subject to change over time as the underlying context shifts. Policymaking is better served by an understanding of the relevant tradeoffs than by categorical endorsements of particular architectural structures as being the foundation for the Internet's success.

Another side effect of the lack of technical sophistication in the current debate is a tendency to defer to opinions advanced by leading members of the engineering community. People without technical backgrounds often regard strong statements of scientific conclusions as possessing a high degree of conclusiveness. Yet anyone who reads broadly in the technical literature quickly realizes that members of the engineering community often disagree sharply over the best way to move forward and that many seemingly authoritative declarations are actually positions in technical debates that are hotly contested and still ongoing. Just as in

---

2. For the seminal statement, see David Clark, *A Cloudy Crystal Ball – Visions of the Future*, 24 PROC. INTERNET ENGINEERING TASK FORCE 539, 543 (1992), https://www.ietf.org/proceedings/24.pdf ("We reject: kings, presidents and voting. We believe in: rough consensus and running code.").

economics and law, where there are often as many different positions as there are people offering opinions, so too in network engineering. At the same time, many areas over which policymakers are now struggling are regarded by the engineering community as completely uncontroversial and long settled.

Understanding how technical considerations should influence Internet policy thus requires a better understanding of the principles on which the Internet is based and an appreciation of the current areas of agreement and dispute within the engineering community. Toward this end, the conference program brought together engineers representing the full range of views on various issues currently confronting policymakers, as well as industry participants who have actual experience in deploying and running networks.

I. TUTORIAL

The conference began with a tutorial designed to provide an introduction to the basic engineering concepts underlying the Internet and to provide a flavor of the tradeoffs underlying the architectural choices. Major topics included the differences between host-to-host protocols, such as the Transmission Control Protocol (TCP) and the User Datagram Protocol (UDP); the edge-based approach currently used to manage network congestion, known as Additive Increase Multiplicative Decrease (AIMD); the deployment of active queue management techniques such as Random Early Discard (RED); the role of Classless Inter-Domain Routing (CIDR) to solve emerging routing problems; the challenges posed by network address translators (NATs); the role of the Border Gateway Protocol (BGP) in routing traffic; and the history of scheduling through techniques such as Integrated Services (IntServ), Differentiated Services (DiffServ), MultiProtocol Label Switching (MPLS), Explicit Congestion Notification (ECN), and emerging techniques such as Low Extra Delay Background Transport (LEDBAT). It offered some observations about current demands that the Internet is not designed to perform well, such as cost allocation, efficiency, security, mobility, and multicasting. It also offered some examples of how architectural decisions that are locally rational can create unexpected and potentially problematic interactions as traffic scales.

II. THE CONTINUING DEBATE OVER NETWORK MANAGEMENT AND QUALITY OF SERVICE

Over the past two decades, some engineers have proposed a series of enhancements to the Internet's architecture to provide more reliable quality of service than the current "best efforts" architecture permits.[3] Other engineers believe that instead of deploying new forms of network management, the better solution is simply to add more capacity.[4] This panel reexamined this debate in light of recent changes to the technological and competitive environment.

David Clark, who served as DARPA's chief protocol architect during the 1980s and currently serves as senior research scientist at the Computer Science and Artificial Intelligence Laboratory at MIT, expressed annoyance that the term "management" had been co-opted in the current debate, given that networks have always been managed. He also criticized the term "network neutrality" given that the Internet is not now and never has been neutral.[5] Instead, the issue is how to manage scarcity, which leads to congestion. Interestingly, the latency that degrades the performance of many time-sensitive applications is often caused by routers deployed by end users in their home networks (a phenomenon called "self congestion") in ways that is alleviated, but not eliminated, by increasing the bandwidth of the access link. It can also arise in other locations on a steady state or intermittent basis. Clark also indicated that concerns about strategic uses of discrimination to create artificial scarcity are overblown, in part because network providers do not need quality of service (QoS) techniques to create scarcity and in part because providing QoS would help innovation. The QoS techniques designed into the protocols that run the Internet ensure that decisions about prioritization are made by end users rather than network operators.

Deke Kassabian, senior technology director for networking and telecommunications at the University of Pennsylvania, described how

---

3. For textbook discussions of these proposals, see, e.g., 1 DOUGLAS E. COMER, INTERNETWORKING WITH TCP/IP: PRINCIPLES, PROTOCOLS, AND ARCHITECTURE 510–14 (5th ed. 2006); JAMES F. KUROSE & KEITH W. ROSS, COMPUTER NETWORKING: A TOP-DOWN APPROACH 602–04, 660–72 (5th ed. 2010).

4. *See, e.g.*, COMER, *supra* note 3, at 511; KUROSE & ROSS, *supra* note 3, at 603, 629–31.

5. *See* David Clark, Written Statement to the En Banc Public Hearing on Broadband Network Management Practices Before the FCC (Feb. 25, 2008), http://www.fcc.gov/broadband_network_management/022508/clark.pdf ("The Internet is not neutral, and has not been neutral for a long time.").

network architectures of large research universities are designed. Penn ensures that its user community has flexible and affordable access to network capacity by maintaining a private line connection to the nearest carrier hotel, where it can obtain easy access to a wide variety of service providers. In terms of performance management, Penn's basic approach is to add bandwidth rather than actively manage QoS. Penn does engage in some bandwidth management, however, by limiting students' Internet access on a per-address basis as well as capping the total amount available to students. Penn occasionally protects other users by limiting the bandwidth consumed by major research projects, sometimes diverting network intensive research projects onto Internet2's Interoperable On-demand Network (ION), which can establish dedicated circuits on a temporary basis.[6] In terms of security, rather than relying on a border firewall, Penn minimizes the impact on other users by deploying security as close as possible to the asset being protected through hardened server configurations, dedicated firewalls in front of a server, or broader use of authentication. Kassabian summarized the essence of this approach captured with the mantra, "open networks, closed servers, protected sessions."

Paul Dauby, vice president and chief operating officer of the Perry-Spencer Rural Telephone Cooperative (PSC), described the efforts of a remarkable rural cooperative serving six counties in southwest Indiana. Despite serving a territory with only 10.3 access lines per square mile and 2.98 subscribers per route mile, PSC supports a dazzling variety of services.[7] It offers digital subscriber line (DSL) service to all of its customers; fixed wireless broadband through unlicensed spectrum;[8] fiber-to-the-home to 560 customers in areas where it operates as a competitive local exchange carrier (CLEC);[9] limited multichannel video to its broadband customers via a virtual local area network (VLAN); and a ten-gigabit regional Ethernet transport that serves area hospitals. In order to make wireless broadband work on unlicensed spectrum, it limits the

6. In private conversations, Kassabian indicated that Penn also prioritizes traffic associated with public safety communications and environmental controls.

7. By way of comparison, Dauby indicated that if a city with the geographic footprint of Washington, D.C., had equivalent subscriber density as the service area in which PSC operates as an incumbent local exchange carrier (ILEC), it would only have seven hundred total subscribers.

8. PSC uses its wireless network for backhaul as well as for providing direct end user connections.

9. Dauby reports that PSC recently received a $29 million grant from the Rural Utilities Service to provide fiber-to-the-premises to its ILEC customers as well.

bandwidth available to peer-to-peer applications, restricting them to no more than ten sessions. PSC currently does not rate limit its wireline offerings despite the fact that it pays transit costs that are several times the cost in larger cities. The advent of over-the-top video is placing increasing financial pressure on their ability to continue its policy of nondiscrimination.

Paul Misener, vice president for global public policy at Amazon.com, remarked about what he saw as a surprising level of agreement on network neutrality. Specifically, both sides of the debate agree that openness is good, that a fair amount of concentration exists at the edges, and that switching costs restrict end users' ability to change providers. In addition, the industry had been in a state of détente during which few untoward activities had occurred, which he attributed to the network providers' fear of regulation. He argued that topological solutions—such as moving servers nearer to end users, buying private line service to closer interconnection points, and contracting with content distribution networks (CDNs) like Akamai—did not violate network neutrality so long as they involve new investments that are incremental to the facilities used to provide existing services. During the question and answer session, he argued that networks should be permitted to favor time sensitive applications such as voice over Internet protocol (VoIP) over less time sensitive applications such as file transfers.

III. CHANGING TECHNOLOGY AND THE LIMITS OF THE LAYERED AND END-TO-END MODELS

Network engineers have long explored alternatives to the layered, edge-based approach that dominates the network's current architecture.[10] This shift is motivated in part by one of the most distinctive characteristics of networks, specifically the interactions between individual flows and the underlying protocols as networks scale. It also reflects the emergence of management and security solutions that require the aggregation of information about the behavior of multiple endpoints and flows. This panel, chaired by the late W. David Sincoskie, professor of electrical and computer engineering and director of the Center for Information and Communication Sciences at the University of Delaware, who tragically

10. *See, e.g.*, R. Bush & D. Meyer, *Some Internet Architectural Guidelines and Philosophy*, IETF RFC 3439, at 7 (rel. Dec. 2002), http://tools.ietf.org/pdf/rfc3439; *The Rise of the Middle and the Future of End-to-End: Reflections on the Evolution of the Internet Architecture*, IETF RFC 3724 (J. Kempf & R. Austein eds., rel. Mar. 2004), http://tools.ietf.org/pdf/rfc3724.

passed away on October 20, 2010, explored the implications of those changes. Sincoskie shared anecdotes of his experiences in the telecommunications industry. He also offered the observation that the Internet is no longer end-to-end and that layering is an abstract concept that when strictly enforced does not perform well in reality.

Matt Mathis, who recently served as senior networking engineering specialist at the Pittsburgh Supercomputing Center, explained how new implementations designed to make TCP run faster are causing congestion in parts of the network. For example, the auto-tuning feature of Windows Vista, Windows 7, Linux, and Mac O/S causes end users running those operating systems to obtain a greater proportion of the available bandwidth than end users running older versions of Windows, such as Windows XP. In addition, TCP allocates bandwidth in inverse proportion to the roundtrip time of the underlying TCP connection. This allows end users located relatively close to their data to consume up to ninety percent of the capacity of the relevant link. Also, the new implementations are designed to expand their transmission windows until they fill all of the available links. Thus, unlike previous implementations of TCP, new implementations inevitably create congestion at some location in the network. This makes performance unstable and unpredictable and makes it extremely difficult for network providers to outbuild the load, particularly when applications are designed to prefetch data. The result is that the network has to play a more active role in allocating network capacity through techniques such as weighted fair queuing.

Jason Livingood, executive director for Internet systems engineering, National Engineering and Technical Operations, Comcast Cable Communications, noted the vehement disagreement among engineers over the relative merits of edge-based versus network-based solutions, pointing out that the decision the two approaches should not be regarded as a binary choice. Instead, engineering's emphasis on tradeoffs and optimality means that any particular solution makes sense for particular circumstances and is necessarily subject to change over time. He gave several examples of functions that previously were provided by the hosts operating at the edge of the network were migrating into the core—including cloud computing, antispam filtering, congestion management, security, and some type of relay to provide global access to content during the transition from IPv4 to IPv6. Other developments were shifting functions in the opposite direction, such as the Session Initiation Protocol (SIP), which was shifting primary responsibility for the functions traditionally associated with telephone

switches operating in the core of the network into the hosts operating at the edge.

Kevin Werbach, assistant professor of legal studies and business ethics at the Wharton School of the University of Pennsylvania, observed that the layered approach that the engineering community uses to frame network design contrasts sharply with the siloed, technology-specific approach reflected in the federal statutes governing communications law. In addition, he pointed out that the layered model does not prescribe certain architectures and that the real world frequently does not conform to the theoretical model. He identified several risks in the current debate, including superficially applying engineering concepts to policymaking, thinking in terms of absolutes, and oversimplifying. He also pointed out a number of ways in which the network has changed since the Internet's primary protocols were designed in the 1970s, including the growing importance of wireless networks, cloud computing, online gaming, video, the Internet of things, and the Internet as a platform for commerce, advertising, and media distribution. He called for a better understanding of the incentives of network players and the relationships between them, better translation of engineering principles into the legal discourse, and more complete data to serve as the basis for decisionmaking.

I served as the fourth panelist and began by pointing out engineers disagree sharply over the relative merits of layering and the end-to-end argument. Moreover, while the policy debate tends to equate layering with ensuring that the lower layers and the core of the network remain relatively "dumb," the engineering community tends to regard the layered stack as following an "hourglass" model that recognizes both that the upper and lower layers of the network are often quite complex and that only the middle layer primarily responsible for addressing that must be kept simple. In addition, contrary to what others suggest, layers do not operate completely independently. Many common protocols cross layers, and interactions across layers have led to the development of active queue management and other core-based solutions to ensure that network resources are allocated fairly. Moreover, because routers operating in the core of the network are able to see what multiple end users are doing, they are often in a better position to implement certain security and congestion management techniques. Lastly, protocol layering can create a design hierarchy that promotes innovations that are consistent with the hierarchy while simultaneously discouraging innovation that is inconsistent with the

hierarchy.[11]

IV. ARCHITECTURE AND NETWORK SECURITY

The engineering community has long recognized that the anonymity and connectionlessness of the Internet's original architecture limits the network's ability to meet end users' growing need for security. The conference's third panel, chaired by Matthew Blaze, associate professor of computer and information science at the University of Pennsylvania, explored ways in which the current architecture can support network security as well as technical changes under consideration that could enhance its ability to do so.

Andrea Matwyshyn, assistant professor of legal studies and business ethics at the Wharton School, emphasized the importance of taking human considerations into account when designing network security. Instead of reflexively regarding failures as the result of user error, exemplified by the oft-used acronym PEBKAC ("problem exists between keyboard and chair"),[12] security systems should take into account the fact that even the best intentioned end user is imperfect and should reflect the way people interact with technology. Network engineers should also assume that every security system can and will be broken, and they should proactively incorporate response plans for when this inevitably occurs. They should also remember that end users are capable of understanding how to respond to problems—if solutions are clearly explained to them. Network security would also be improved by more frequent interactions between engineers and lawyers, and by bearing in mind that security is governed by a wide range of competing legal regimes—including (but not limited to) contract, intellectual property, telecommunications regulation, and consumer protection laws.

Edward Felten, professor of computer science and public affairs and director of the Center for Information Technology Policy at Princeton University, analyzed the security implications of the decision to place functions in the network's endpoints or in the network's core. As an initial matter, Felten emphasized that end users are not the only endpoints and that many functions that end users regard as being in the network (such as cloud computing, email servers, and other third-party intermediaries) are, from

---

11. I expand on these ideas in Christopher S. Yoo, *Protocol Layering: A Study in Incorporating Engineering Insights into Internet Policy*, 60 DUKE L.J. (forthcoming May 2011).

12. Another commonly used acronym is PICNIC ("problem in chair, not in computer").

the standpoint of network architecture, simply other endpoints. Moreover, the most threatening and most visible security problems (including malware such as botnets and spyware, server attacks, and phishing and other attempts to deceive end users) generally arise on the end hosts. Network-oriented security threats exist, such as attacks on the routing infrastructure, and networks can analyze traffic to help detect security problems. That said, end hosts can view traffic after it has been unpacked from any archives, decompressed, decrypted, and reassembled. This often places them in a better position to implement security, especially because they can conduct dynamic analysis of code while it is operating instead of simply static code residing on a hard drive.

Jonathan Smith, Olga and Alberico Pompa Professor of Engineering and Applied Science at the University of Pennsylvania, noted that while the end-to-end argument decentralizes innovation, it also decentralizes responsibility for security enforcement. Moreover, the network's current bias toward allow-by-default facilitates connection, such a default may no longer be the correct architecture in a network that has become a distributed system increasingly populated by security threats. In addition, although layers create opacity that makes programming easier by reducing what a programmer needs to know about how other layers are configured, hiding information about what is going on in lower layers may possibly be problematic from the standpoint of trust. Smith identified a number of solutions that are not working, including passwords, public key infrastructures, software updates, measures to protect the routing infrastructure (such as IPSEC, DNSSEC, and BGPSEC), firewalls, intrusion detection systems, and rate throttling defenses. Instead of implementing these ineffective solutions, network architects should improve the infrastructure for authentication and attribution, build automated trust systems, provide for a degree of cross-layer transparency through structures such as a knowledge plane, shift to deny-by-default, and make both the edge and the core more extensible.

V. KEYNOTE ADDRESS BY PAUL MOCKAPETRIS

The dinner keynote address delivered by Paul Mockapetris, inventor of the domain name system and currently chairman and chief scientist at Nominum, Inc., noted that the success of the network is often attributed to what is often called Metcalfe's law, which holds that a network's utility is proportional to the square of the number endpoints in the network. This

implies that a network's value grows quadratically as it expands.[13] So long as the value grows faster than the cost, networks keep growing wonderfully. The problem is that in the modern world, being part of a larger network does not necessarily confer benefits to the extent that it provides connections to hackers and other security threats. One solution is to use the DNS to begin tracking reputation data about particular actors. Although some industry observers raise concerns about placing critical information that needs to be secure into the DNS, this objection overlooks that fact that critical information is already in the DNS. Although some people argue that smart DNS services deviate from the simplicity of the hourglass model often used to describe the Internet, in reality, we already have multiple hourglasses to deal with different types of transmission technologies.

Mockapetris closed by offering a few observations about network neutrality, arguing that it should be illegal for parties to give users applications that act against their interests without making clear what those applications are doing, wondering if such safeguards are best served by an architecture that does not reveal who is serving as the counterparty and market maker in any particular transaction, as is the case in the current network architecture.

VI. NEW APPLICATIONS, NEW CHALLENGES

Emerging applications, such as Internet protocol television (IPTV) and gaming, are placing demands on networks that are quite different from the flows generated by the applications that dominated the early Internet, such as email and web browsing. This panel, moderated by Saswati Sarkar, associate professor of electrical engineering at the University of Pennsylvania, explored the pressures that these new applications are creating on the network architecture as well as the technological options to adapt to these changes.

I provided an overview of the technical and policy challenges confronted by IPTV. Some IPTV providers employ dedicated or prioritized connections between the central office and the end users' premises. "Over-the-top" services, such as Netflix and YouTube, rely on the public Internet to transport their packets on a best efforts basis. Over-the-top services

13. *See* George Gilder, *Metcalf's Law and Legacy*, FORBES ASAP ARTICLES BY GEORGE GILDER, BASED ON CHAPTERS IN HIS FORTHCOMING BOOK – TELECOSM (Sept. 13, 1993), http://www.seas.upenn.edu/~gaj1/ggindex.html (click on "Metcalfe's [sic] Law and Legacy" hyperlink).

employ a wide variety of techniques to provide the QoS needed to support video, including content delivery networks and adaptive streaming (which adjusts video resolution quality in light of the available bandwidth). In addition, IPTV providers must decide which platforms to support, both in terms of devices (such as PCs, Blu-ray players, gaming consoles, and smart phones) and encoding formats, which often incorporate varying maximum transfer rates. In order to obtain access to content, IPTV providers must also protect content against illegal copying, either through digital rights management (DRM) or filtering and must anticipate likely reactions to these measures, such as encryption, darknets, and greater exploitation of the analog hole. In addition, the growing importance of video has renewed interest in using multicasting to distribute mass media content. IPTV is also limited by legacy regulatory requirements, such as mandates for public, educational, and governmental (PEG) channels.

Paul Mitchell, general manager for regulatory and standards at Microsoft, discussed some the challenges confronted by game consoles such as the Xbox, Microsoft's effort to use high performance computing, home-theater quality graphics and audio, and network connectivity to provide an interactive, immersive game experience. The feature designed to allow users to communicate with each other while gaming drew the attention of regulators interested in determining whether this feature represented a telecommunications service.[14] Microsoft has now combined the Xbox with other products as services (such as Windows Phone 7, Microsoft Communicator, and the Kin smartphone) to allow voice communications and the sharing of video and audio across a wide variety of platforms. In many countries, however, regulatory restrictions prevent end users from taking advantage of the full range of these features. Another challenge is finding ways to make DRM interoperable. Regarding network neutrality, although that all networks are managed, they should be managed in predictable ways. Mitchell also provided a demonstration of adaptive streaming and described the challenges of supporting features such as closed captioning on a wide range of devices and encoding formats.

---

14. Xbox has also become a platform for distributing Netflix. In earlier conversations, Mitchell also discussed how regulators also inquired whether Xbox's Party Mode, which allows friends in separate locations to watch the same video at the same time, represented a cable service. Telephone Interview with Paul Mitchell, General Manager for Regulatory and Standards, Microsoft Corp. (Apr. 9, 2010). Microsoft has subsequently taken steps to turn the Xbox into a platform for subscription television service. Nick Eaton, *Microsoft Considering TV Service on Xbox*, MICROSOFT BLOG (Nov. 20, 2010, 11:05 AM), http://blog.seattlepi.com/microsoft/archives/229997.asp.

Joe Weinman, vice president for strategy and business development at AT&T Business Solutions, observed that the future demand for video distribution appears to be effectively insatiable, driven by new technologies such as ultra HD, multiscreen video for immersive virtual environments, 3D video, and the incorporation of video into social networking. At the same time, chip manufacturers are producing new products that make mobile video increasingly feasible. Other technologies that will increase the demand for bandwidth include Javascript and XML (Ajax), which triggers request for data when a mouse is moved or a keystroke is struck, such as popup information when a mouse hovers over a link. Other technologies that will increase the demand for bandwidth include sensor networks, cloud computing, and the emergence of households as de facto data centers in their own right. Solutions such as rate adaptation are useful stopgap measures, but may not work well when multiple users adapt in the same way at the same time. More problematically, rate adaptation addresses congestion by degrading the end users' experience rather than by ensuring that end users have access to the network resources needed to run highly interactive, latency-sensitive, and bandwidth-intensive applications.

Marjory Blumenthal, associate provost for academics at Georgetown University and former executive director of the National Academy of Sciences' Computer Science and Telecommunications Board, commented on all of the presentations. She noted the uncertainty implicit in the wide variety of predictions about the future of video, which range from the wildly optimistic to the severely pessimistic, and raised the possibility that adaptive technologies may represent a reasonably effective compromise that sufficiently preserves the end user experience. Regulatory requirements such as PEG can vary widely across different areas.[15] Others such as the Communications Assistance to Law Enforcement Act (CALEA)[16] can lead to unintended consequences.[17] In addition, the increasing cost effectiveness of filtering technologies, the ability to protect against illegal downloads through man-in-the-middle strategies, and the importance of proprietary DRM standards are changing the role of Internet service providers (ISPs). Lastly, the remote storage of data implicit in cloud computing puts someone other than the end user in charge of determining whether particular data is saved or lost, which can limit end users' control

15. *See* 47 U.S.C. § 531 (2006).

16. *Id.* §§ 1001–10.

17. *See, e.g.*, Daniel F. Spulber & Christopher S. Yoo, *On the Regulation of Networks as Complex Systems: A Graph Theory Approach*, 99 NW. U. L. REV. 1687, 1719 (2005).

over their own identities.

VII. THE FUTURE IS WIRELESS

As the FCC's proceeding on "Preserving the Open Internet" recognizes, wireless network face challenges that are quite different from wireline networks.[18] This panel, moderated by David Farber, distinguished career professor of computer science and public policy at Carnegie Mellon University, moved beyond the traditional focus on spectrum allocation to consider the unique management challenges that wireless networks confront, paying particular attention to how the physics of wave propagation, differences in network reliability, and the dynamic changes in the routing architecture associated with mobility often require wireless networks to employ network management techniques.

Dirk Grunwald, professor of computer science at the University of Colorado, discussed the difficulties inherent in the physics of wave propagation. Every frequency has different characteristics in terms of attenuation, absorption, and diffraction. Moreover, multipath reflections can cause the same signal to arrive at the same location along two different paths. If they arrive out of phase, they can cancel each other out in the same way that Bose headphones and sound dampening systems in cars operate. This causes signal quality to vary across time and space, demonstrated by how moving a car slightly can dramatically affect the quality of a radio signal. Engineers compensate for these variations by using different modulation schemes, which necessarily provide less bandwidth to distant locations. Network operators must decide in an environment that is constantly changing whether to equalize the performance of nearby and distant links rather than maximize total throughput. Differences in loss rates also affect the performance of TCP, because the average throughput rate is inversely proportional to the square root of the packet loss rate.[19] The solution may be to employ multiple solutions simultaneously allowing cognitive radios to maximize spectrum reuse.

Charles Jackson, a consultant who has previously held staff positions with the FCC, the U.S. House of Representatives, and the U.S. Commerce Department, addressed some of the network-based issues associated with

---

18. Preserving the Open Internet, *Report and Order*, 52 Comm. Reg. (P & F) 1, at paras. 86, 94–95, 103 (2010), *available at* http://hraunfoss.fcc.gov/edocs_public/attachmatch/FCC-10-201A1.pdf.

19. Matthew Mathis et al., *The Macroscopic Behavior of the TCP Congestion Avoidance Algorithm*, COMPUTER COMM. REV., July 1997, at 67–68.

wireless networking. As an initial matter, wireless networks typically give voice communications priority over data traffic, which is typically less sensitive to latency. Preventing wireless networks from prioritizing in this manner either holding back reserve capacity that cannot be used for data transmissions or permitting voice service to degrade. The fact that radio links are less reliable than wireline connections has also led wireless networks to deploy smart-link technologies such as Automatic Repeater reQuest (ARQ) to shift responsibility for error recovery from the endpoints to the network. In addition, handset upgrades can often substitute for network investments, since receivers that are more sensitive require less capacity from base stations. Moreover, host-based congestion control depends on an honor system that is breaking down, which is causing networks to take a more active role in allocating bandwidth. Jackson also provided examples where traffic surges from Windows updates or earthquakes led ISPs to throttle certain types of traffic.

Robert Khedouri, chief executive officer of MusicGremlin, Inc., and vice president for services/strategy & planning for mobile network operators at SanDisk, described his experience launching the first MP3 player capable of downloading music directly from WiFi hotspots instead of sideloading it from a PC. MusicGremlin chose to adopt a "closed loop" system in which a single entity guaranteed secure delivery all the way from the content owner to the end user's device, similar to the manner in which Apple's iTunes establishes a closed loop between content owners and PCs. Relying on a closed, integrated system, complete with a vertically integrated music service, allowed MusicGremlin to provide the protection against piracy on which content providers insist. It also allowed the system to offer the value proposition to end users of ensuring seamless transfer with low latency. The company also deployed other bandwidth saving technologies, such as pushing content overnight to users who signed up for playlists, using burstable downloads to conserve on battery life, and caching the entire catalog of songs on every device to reduce search latency. MusicGremlin was acquired by SanDisk in 2008.

Christian Sandvig, associate professor of communication at the University of Illinois at Urbana-Champaign, noted that previous metaphors used to describe wireless technologies provide little insight into emerging aspects of spectrum, such as cognitive radios, smart antennas, and innovative forms of spectrum reuse. In addition, these metaphors fail to capture the variability and sensitivity to local conditions that make the performance of wireless networks so unpredictable, as illustrated by the

following example. While living in London, Sandvig deployed a directional antenna to provide WiFi service to the famous Speakers' Corner in Hyde Park,[20] only to find his signal intermittently negated despite the absence of any direct obstructions. The cause was double-decker buses stopped at a nearby traffic light, which periodically created a multipath reflection that cancelled out the direct signal. In addition, wireless networks face a tradeoff between making wireless devices easier to operate by hiding complexity and increasing wireless networks' configurability. On the one hand, the proliferation of wireless devices has turned consumers into overburdened band managers for their own houses. On the other hand, the advent of sensor networks and other technologies have made it easier than ever for them to adapt to local conditions.

* * *

The presentations and discussions at the conference represented a remarkable exploration of the issues that yielded fresh insights into issues of broadband policy. Indeed, former FCC Chief Economist Gerald Faulhaber congratulated the program for accomplishing something new in telecommunications policy, which he regarded as no mean feat.

The pages that follow contain articles by selected speakers exploring many of the themes raised during the conference. The conference proceedings and this special conference issue represent the first step in what we hope will be a new CTIC-led research initiative designed to better integrate the principles of network engineering into Internet policy debates.

20. For a description of this experiment, see PHILIP N. HOWARD, NEW MEDIA CAMPAIGNS AND THE MANAGED CITIZEN xi–xii (2006).

The End-to-End Argument and Application Design: The Role of Trust

David D. Clark*

Marjory S. Blumenthal**

---

* David Clark is a senior research scientist at the MIT Computer Science and Artificial Intelligence Laboratory, where he has worked since receiving his Ph.D. there in 1973. Since the mid 70s, Dr. Clark has been leading the development of the Internet; from 1981–1989 he acted as Chief Protocol Architect in this development and chaired the Internet Activities Board. His current research looks at redefinition of the architectural underpinnings of the Internet and the relation of technology and architecture to economic, societal, and policy considerations. Dr. Clark is past chairman of the Computer Science and Telecommunications Board of the National Academies, and he has contributed to a number of studies on the societal and policy impact of computer communications. Support for Dr. Clark's effort on this research was provided by the U.S. Office of Naval Research grant number N00014-08-1-0898.

** Marjory S. Blumenthal is associate provost, Academic at Georgetown University. Between July 1987 and August 2003, she served as founding Executive Director of the National Academies Computer Science and Telecommunications Board (CSTB; http://cstb.org). She is a member of the Advisory Board of the Pew Internet & American Life Project and the Center for Strategic and International Studies Commission on Cybersecurity; she is a fellow of the National Academy of Public Administration; she chairs the External Advisory Board of the Center for Embedded Networked Sensing at UCLA; and she is a RAND adjunct and an Office of Naval Research grantee. This work was supported by a grant from the U.S. Office of Naval Research grant number N00014-09-1-0037.

I. INTRODUCTION

Applications are the *raison d'être* of the Internet. Without e-mail, the Web, social media, VoIP and so on, the Internet would be (literally) useless. This fact suggests that the structure of applications, as well as the structure of the Internet itself, should be a subject of study, both to technologists and those who are concerned with the embedding of the Internet in its larger context. However, the Internet, as the platform, may have received more attention and analysis than the applications that run on it.

The original end-to-end argument[1] was put forward in the early 1980s as a central design principle of the Internet, and it has remained relevant and powerful as a design principle, even as the Internet has evolved.[2] However, as we will argue, it does not directly speak to the design of applications. The original end-to-end paper poses its argument in the context of a system with two parts, the communications subsystem and "the rest."[3] That paper says: "In a system that includes communications, one usually draws a modular boundary around the communication subsystem and defines a firm interface between it and the rest of the system."[4] Speaking generally, what the end-to-end argument asserts is that application-specific functions should be moved up out of the communications subsystem and into "the rest" of the system. But the argument, as stated, does not offer advice about how "the rest" should be structured. That paper equates the "rest of the system" with the application, and the application with the end points. It says: "The function in question

1. J.H. Saltzer, D.P. Reed & D.D. Clark, *End-to-End Arguments in System Design*, 2 ACM TRANSACTIONS ON COMPUTER SYS. 277 (1984).

2. This may reflect path dependence—the Internet remains young enough that it should not be surprising to see a common set of underlying uses persist.

3. Saltzer et al., *supra* note 1, at 278.

4. *Id.*

can completely and correctly be implemented only with the knowledge and help of the application standing at the end points of the communication system. Therefore, providing that questioned function as a feature of the communication system itself is not possible."[5]

Applications and services on the Internet today do not just reside at the "end points"; they have become more complex, with intermediate servers and services provided by third parties interposed between the communicating end points. Some applications such as e-mail have exploited intermediate servers from their first design. E-mail is not delivered in one transfer from original sender to ultimate receiver. It is sent first to a server associated with the sender, then to a server associated with the receiver, and then finally to the receiver. By one interpretation, all of these intermediate agents seem totally at odds with the idea that function should be moved out of the network and off to the end points. In fact, the end-to-end argument, as described in the original paper, admits there are interpretations that are diametrically opposed. When we consider applications that are constructed using intermediate servers, we can view these servers in two ways. An Internet purist might say that the "communications subsystem" of the Internet is the set of connected routers; servers are not routers, but are connected to routers; as such, servers are outside the "communications subsystem." This reasoning is compatible with the end-to-end argument of placing servers anywhere in "the rest" of the system. On the other hand, these servers do not seem like "ends," and thus they seem to violate the idea of moving functions to the ends. These issues are prominent today, thanks to the emergence of cloud computing—which involves specific sorts of servers—and the tendency of some popular discourse to treat "the cloud" as a new incarnation of the Internet itself.[6]

The original end-to-end paper, because it uses a simple two-part model of the communications subsystem and "the rest," does not directly speak to the situation where "the rest" has structure. The purpose of this Article is to offer an interpretation of the end-to-end argument, drawing on the original motivation and reasoning, that is applicable to today's application design and today's more complex world of services and service providers.

A. What Is an End Point?

Part of the definitional problem, of course, is to define the end point.

---

5. *Id.* (emphasis omitted).

6. *See, e.g.*, Phil Dotree, *Cloud Computing: The Most Important Technology of 2010*, ASSOCIATED CONTENT FROM YAHOO! (Jan. 13, 2010), http://www.associatedcontent.com/article/2585171/cloud_computing_the_most_important.html.

There is an intuitive model that is often adequate: if computer A is sending a file to computer B (to use the example of "careful file transfer" from the original paper[7]), then A and B are end points. However, they are end points in two ways that are subtly different. In the original example, the end points are the literal source and destination of the data being sent across the communications subsystem. They are also the end points in that they are the prime movers in the activity—they are directly associated with the principals that actually wanted to accomplish the action. Intermediate nodes, whether at the packet level or application service level, seem to play a supporting role, but they are not the instigators of the action, or the nodes that wanted to see it accomplished.

The original paper provides a hint as to the importance of this distinction. Using a telephone call as an example, it points out that the ultimate end points are not the computers, but the humans they serve.[8] As an illustration of human-level end-to-end error recovery, one person might say to another: "[E]xcuse me, someone dropped a glass. Would you please say that again?"[9] The humans are the prime movers in the activity, the ultimate end points. The computers are just their agents in carrying out this objective.

In the case of a phone call, the humans and the computers are colocated. It makes no sense to talk about making a phone call unless the person is next to the phone. So one can gloss over the question of where the human principal is. But in the case of careful file transfer, the location of the person or persons instigating the action and the location of the computer end points may have nothing to do with each other. As an example, there might be one person, in (say) St. Louis, trying to do a careful file transfer from a computer in San Francisco to a computer in Boston. Now, what and where are the end points?

The person in St. Louis might undertake a careful file transfer in three stages. First, she might instruct the computer in San Francisco to compute a strong checksum of the file (i.e., a measure of the bits it contains) and send it to her in St. Louis. Then she might instruct the two computers to carry out the transfer. Third, the person might instruct the computer in Boston to compute the same strong checksum and send it to St. Louis, where she can compare the two values to confirm that they are the same. In this case, the computers in San Francisco and Boston are the end points of the *transfer*, but they seem just to be agents (intermediaries) with respect to the person in St. Louis. With respect to the instigation of the transfer, there seems to be one principal (one end point) located in St. Louis.

---

7. Saltzer et al., *supra* note 1, at 278.
8. *See id.* at 284–85.
9. *Id.* at 285.

It might seem that this example serves to further confuse the story, rather than clarify it. But if we explore one step deeper, we can begin to find some clarity. The example above, building on the example in the original paper, referred to the overall activity as "careful file transfer." It is important to ask, why is that sequence of steps being careful? It is careful only in the context of an assumed failure mode—that is, loss or corruption of information during transfer. But why does the end user assume that the computation of the checksum will not fail? Why does the end user assume that the checksum returned by the computer is actually the checksum of the file, as opposed to some other value? Why does the end user assume that the file transferred today is the same as the file stored earlier? Why does the end user assume that the file will still be there at all? A prudent end user would be careful about these concerns as well. Perhaps the file was copied to Boston because the computer in San Francisco is crash prone or vulnerable to malicious attack. Perhaps this move was part of a larger pattern of "being careful." Perhaps, in a different part of the story, the end user in St. Louis has the computer in San Francisco compute the strong checksum on multiple days and compares them to see if they have changed. All of these actions would represent "being careful" in the context of some set of assumed failures.

But if there is *no* part of the system that is reliable, being careful is either extremely complex and costly, or essentially impossible. For example, the end user cannot protect against all forms of failure or malice using the comparison of strong checksums, because it may not be possible to detect if one of the computers deliberately corrupts the file but returns the checksum of the correct version. Ultimately, being careful has to involve building up a process out of component actions, some of which have to be trustworthy and trusted.

II. RELIABILITY AND FUNCTION PLACEMENT

The example of careful file transfer in the original paper can help us to explore the relevance of the end-to-end argument to today's world. It points to the need to define what it means to be careful in a more general sense. Being careful implies making a considered and defensible judgment about which parts of the system are reliable and which parts are failure prone or open to malicious attack—being careful today implies a degree of risk management. Using careful design implies constructing a set of checks and recovery modes that can compensate for the unreliable parts. The end user in St. Louis, moving a file from San Francisco to Boston, presumably has decided to place some level of trust in those two computers. She has also designed the pattern of information movement and storage to make the overall outcome reliable, based on the assumed level of reliability and trust

of the component parts, including the computers and the communications subsystem that connect them. The trust assumptions are made by the end user (who is, at one level, the end point), and the computers are trusted agents that act on behalf of the end user.

Why does the above view of "being careful" motivate us, in the context of the original end-to-end argument, to move functions out of the communications subsystem and into the end nodes? The original paper lists several reasons:

- In some respects, it is technically very hard to make a communications subsystem fully reliable. In a system with statistical sharing, for example, there is a probability of packet loss. Such imperfections are technical consequences of rational technical design.
- Adding mechanisms to the communications subsystem adds to its complexity, and complexity seems to make systems less reliable, as well as more costly.[10]
- The communications system may not be fully trustworthy. The original paper recognizes this issue—it talks about the peril of having the communications subsystem do encryption on behalf of the end node: "[I]f the data transmission system performs encryption and decryption, it must be *trusted* to securely manage the required encryption keys."[11]
- The providers of the communications subsystem may not be motivated to provide service with the level of reliability the end user desires and can depend on.[12]

There is an explicit assumption in the original paper that the communications subsystem is unreliable.[13] This assumption is justified (both then and now) for the reasons listed above. But there is an *implicit* assumption that the end node *is* reliable and trustworthy. The example of "careful file transfer" in the original paper[14] assumes that the end node can compute a checksum reliably and perform other actions designed to compensate for the unreliability of the communications. It also assumes, implicitly, that the two ends trust each other. One end wants to send the file to the other, and the other wants to receive it. Presumably, the interests of

---

10. Technical advances and a more mature understanding of the system, as well as a desire to add new features, have led to increasing complexity of the communications substrate of the Internet. It is an interesting question as to whether that has reduced the overall reliability of the Internet, but this Article does not focus on issues of this sort of complexity.

11. Saltzer et al., *supra* note 1, at 282 (emphasis added).

12. *Id.* at 287.

13. *See generally id.*

14. *See id.* at 278–82.

the two ends are aligned in this respect. But let us challenge these assumptions and see what happens.

What if the two ends do not trust each other? This situation is common today. People receive e-mail but worry that it is spam or contains a virus. They are willing to receive it (because it is worth the risk), but they do not trust the sender. Now what does it mean to be careful? This is a real-world situation, so we can see what the real-world answer is. People deploy spam filters, virus checkers, and so on. And where is that done? Sometimes it is done at the receiving end point of the mail transfer, and sometimes it is done "in the middle," at one of the mail relay points. Is this a violation of the end-to-end argument?

- As a practical matter, performing these functions at an intermediate point makes sense, because, assuming that the end user trusts the intermediary, it may be more reliable and more convenient.
- The operator of the end node (the end user) may not want to go to the effort of providing the service with the desired level of reliability.
- By performing the function at an intermediate point, the service may have access to more information; for example, a mail filter may be better able to detect spam if it can compare mail going to many recipients.
- By performing the function at an intermediate point, the end user can avoid the cost and overhead of at least temporarily storing and then transferring unwelcome traffic across the communications subsystem to the ultimate end point.
- The end node might have a vulnerability that would allow a virus to attack it before a virus checker on that machine could detect it. Doing the check at an intermediate point can protect the end node from a vulnerability the end user cannot rectify.
- Pre-positioning information at an intermediate point can make the subsequent delivery more responsive as well as more reliable. Replicated intermediate points can specifically improve reliability.

What we see is that function is migrating to the point where it can be done most reliably and efficiently. In some cases, this migration is "naturally" toward the ultimate end points (because of "natural" limits to the reliability of the communications subsystem), but in other cases function may migrate away from the end point to a service point somewhere else in the network.

When we look at the design of applications, we can see different approaches to structure based on different views of those functions that are

reliable and trustworthy and those that are not. Here are two examples.

"Careless" mail transfer. E-mail, an early application for the Internet, has no end-to-end assurance of delivery or data integrity.[15] The mail is sent via a series of servers, any of which might lose the mail. Yet there is no end-to-end confirmation. E-mail seems almost an "anti-careful" file transfer, in contrast to the first example of the original paper. What was the reasoning that made the original design for Internet e-mail come out that way? The original motivation for designing e-mail systems to use forwarding servers was that the sender and the receiver might not be connected to the Internet at the same time, and if the transfer had to be done in one step, it might never succeed. Using an intermediate server is an obvious solution. But for this approach to work with reasonable overall reliability, the servers that relay mail have to be built to a very high standard of availability, reliability, and trustworthy operation. And indeed, each stage of the mail transfer is expected to be "very careful." Given this level of attention to reliability of the intermediate nodes, no end-to-end confirmation was considered necessary. So the overall reliability is built out of a cascade of these steps, rather than an end-to-end confirmation. E-mail is not "careless"; it is just based on a different set of assumptions about which parts of the system are reliable.[16]

What happens if this assumption of reliable delivery is violated? Here is a story passed on by someone who spent two years as a volunteer in Africa, where she was forced to use an e-mail server that often crashed or otherwise lost mail.[17] The end users created a manual reliability mechanism, which was to put sequence numbers in the subject line of each piece of e-mail, and send human-to-human acknowledgements of the sequence numbers by return e-mail. In other words, they added an end-to-end confirmation to deal with the unreliable servers.[18]

Content distribution. Today, much Web content is not delivered to the ultimate recipient directly from the Web server belonging to the original creator, but via a content delivery network (CDN)—a collection of

---

15. Later enhancements to Internet e-mail have provided the option of end-to-end integrity and authenticity checks, often using digital signatures. *See, e.g.*, *Understanding Digital Signatures*, U.S. COMPUTER EMERGENCY READINESS TEAM, http://www.us-cert.gov/cas/tips/ST04-018.html (last visited Feb. 20, 2011). These checks are seldom used today, perhaps because they do not address delivery assurance, something for which tools are lacking. Return-receipt features are used sometimes, but can be ignored by recipients, thereby undermining their value.

16. The same logic can be seen in the recent development of delay- or disruption-tolerant networking; different circumstances give rise to different assumptions about which parts of a system are reliable. *See, e.g.*, *Home*, DELAY TOLERANT NETWORKING RES. GROUP, http://www.dtnrg.org/wiki (last visited Feb. 20, 2011).

17. Interview with Libby Levison in Cambridge, Mass. (2001).

18. *Id.*

servers that cache the content and deliver it on demand. This, like e-mail, has no end-to-end confirmation of correct delivery. Is this design being careful? Is it trustworthy? Commercial CDNs such as Akamai[19] depend on their reputation as a reliable and trustworthy provider. There are no features built into the web standards that assure that they are reliable; there is only the discipline of the competitive marketplace. If they were not reliable and trustworthy, they would go out of business. So they build highly reliable systems, the content creators trust them, and the result is a more efficient overall system.

A. Application-Specific Semantics

There is another aspect to the end-to-end argument, which is that different applications have different semantics—different definitions of what it means to be "reliable" or "correct." In the context of network data transfers, for example, some applications may define "correct" operation as perfect delivery of every byte as sent, while another application may define "correct" as delivery within some time limit, with as few errors and omissions as possible. Putting some mechanism to enhance reliability into the communications subsystem runs the risk of adding a mechanism that does not meet the needs of the application. However, when we look at the placement of application-level function inside "the rest," this argument has less relevance. Wherever application-level components are placed, they can be designed so that they are aware of the application-level semantics. This line of reasoning has been used to argue explicitly for the placement of application-aware components throughout the network, because these components can then be aware of *both* local conditions in the network and application-level requirements.[20]

III. THE CENTRALITY OF TRUST

The previous discussion has used the words "reliable" and "trustworthy" in loose equivalence. However, the distinction is very important. Reliability is a technical concept, and relates to the correct operation of a component or system under specific circumstances. The concept of trust is a broader concept. A component may not be trustworthy

19. *See* AKAMAI, http://www.akamai.com (last visited Feb. 20, 2011).

20. *See, e.g.*, Samrat Bhattacharjee et al., Commentary, *Commentaries on "Active Networking and End-to-End Arguments,"* IEEE NETWORK, May/June 1998, at 66–67. Similar reasoning has also informed planning for the so-called Next Generation Networks by the International Telecommunications Union, where desires by some to support priority access and such applications as telephony have focused attention on in-network mechanisms. *See, e.g.*, *ITU-T Study Group 13 – Future Networks Including Mobile and NGN*, INT'L TELECOMM. UNION, http://www.itu.int/ITU-T/studygroups/com13/questions.html (last visited Feb. 20, 2011).

even though it is technically reliable, because it is operated by an agent with interests and motivations that are not aligned with the end user—the principal who wants to undertake the action. Early experience with public cloud services, including social media, illustrate this concern.[21] Trust or trustworthiness thus includes some of the issues associated with security, and security is recognized as something that can and often should be addressed at multiple points in a system.[22]

A. Multiple Stakeholders

Why would one agent or server be more trustworthy than another? In many applications today, different parts of the application belong to different actors. An ISP may provide a mail server, a third party may provide a web cache or a component of what is displayed on a web page, or a peer system may provide a music-sharing server. The difference in the degree of trustworthiness relates to the motivation and roles of the different actors, and their external influences, which range from economic incentives[23] to legal requirements or constraints.

In many cases, the interests of the different actors are nominally aligned, notwithstanding differences in status or role. End users want to send and receive mail, and ISPs attract customers by providing this service, so both the end user and the ISP want the same thing to happen. The ISP may not want to perform the function exactly as the end user would prefer, and this misalignment is either tolerated or corrected via economic means (competition to provide the service) or through the technical design of the protocol, which allows the trusted elements at each end to compensate for and recover from the failures of the other agents. Recent controversy over privacy on Facebook, a provider of social media services, reflects conflicting incentives facing service providers, who seek to attract and retain both users and advertisers (which want access to users).[24]

---

21. One of the Authors has been examining the potential for the cloud to be a platform for malice from either providers or other users. *See, e.g.*, Marjory S. Blumenthal, *Is Security Lost in the Clouds?*, CONFERENCE ON COMMUNICATION, INFORMATION AND INTERNET POLICY (2011), http://www.tprcweb.com/images/stories/2010%20papers/Blumenthal_TPRC2010.pdf.

22. For example, two mutually trusting end nodes can use encryption to preserve integrity and prevent unwanted disclosure, but preventing attacks that flood the network or disrupt availability by harming network control mechanisms can only be accomplished inside the network.

23. *See* Jonathan Anderson & Frank Stajano, *Not That Kind of Friend: Misleading Divergences Between Online Social Networks and Real-World Social Protocols (Extended Abstract)* (forthcoming in Springer LNCS), http://www.cl.cam.ac.uk/~jra40/publications/2009-SPW-misleading-divergences.pdf (discussing economic incentive weakness) (last visited Feb. 20, 2011).

24. *See, e.g.*, Emily Steel & Geoffrey A. Fowler, *Facebook in Privacy Breach*, WALL ST. J., Oct. 18, 2010, at A1, *available at*

But sometimes, there are actors in the system with motivations that are adverse, rather than aligned. Music lovers of a certain disposition choose to share copyrighted material; the rights-holders try to prevent this. Some end users may prefer to have private conversations; law enforcement (and, in some countries, other governmental elements) wants the ability to intercept conversations.

To understand this situation, one must do an analysis from the perspective of all the actors. Each actor, from its own perspective, has the same ambition about reliable and trustworthy execution of its requirements—but they have different requirements. Performing this analysis will reveal that sometimes one actor's end is another actor's middle, and sometimes the actors fight over the ends. From the perspective of trust, different actors will have different views about which servers and services they can trust, and in this respect, these different servers and services represent different "ends" of the application.

Lawful intercept. Lawful intercept, or government-ordered "wiretapping," is usually conceived as being implemented in the "middle" of the network. One approach is to carry out lawful intercept within the communications subsystem (e.g., the routers of the Internet). This would imply finding a router (perhaps one very close to the end node) that the traffic of interest is likely to pass through. Another idea is to identify some service at a higher layer (an "application layer" service) that is involved in the communication, and implement the intercept there. In the e-mail system, the mail servers are a natural point of intercept. For instant messaging, the IM server would be the target.

In order for an interceptor (lawful or otherwise) to locate a node or server through which the content is flowing, it may be necessary (or at least helpful) if this actor can constrain the set of choices, both technical and commercial, that the end user can exploit. If, because of technical design or economic or policy reasons, the end node is forced to use a particular server that can be easily identified, this makes the intercept much easier to carry out. If the end user can be prevented from using encryption (an obvious "end-to-end" reliability enhancement from the perspective of the communicating end users), the effectiveness of the intercept improves. Accordingly, the legal authorities might try to limit the use of encryption, either by influencing the development of standards, legal restrictions, making encryption hard to use and understand, and so on.[25]

---

http://online.wsj.com/article/SB10001424052702304772804575558484075236968.html.

25. The Internet Engineering Task Force has addressed these concerns for over a decade, declining to accept the task of designing corresponding protocols. *See* Brian E. Carpenter & Fred Baker, *IAB and IESG Statement on Cryptographic Technology and the Internet*, IETF RFC 1984 (rel. Aug. 1996), http://www.ietf.org/rfc/rfc1984.txt; Brian E. Carpenter & Fred Baker, *IETF Policy on Wiretapping*, IETF RFC 2804 (rel. May 2000),

In several countries, as government sophistication about the Internet has grown, so, too, have efforts to monitor and control use, both of which can involve forms of interception. Attempts to visit certain websites, to search the web for certain words, to blog using certain words, to send e-mail to certain recipients, or to send e-mail using certain words have been affected by such government efforts. Even use of anonymizing services can be affected if it constitutes a pattern that can be recognized and constrained.[26] The year 2010 saw a number of countries attempt to prevent use of BlackBerry communication because of its strong encryption, forcing adaptation by BlackBerry as it sought to balance demands from governments and from end users.[27] In some of these countries, regulation of speech and other conduct serves to control Internet access and use, making it, from the perspective of many end users, less trustworthy regardless of ISP or other service provider. An international perspective makes clear that reliability is only one element of trustworthiness and that a well-functioning market is only one kind of force influencing a provider's behavior. Moreover, growth in intergovernmental discussion and cooperation in dealing with cybercrime, spam, and malware—notwithstanding different national stances about such individual rights as privacy and freedom of expression—suggests that pressures for systems to inspect and filter will continue to grow.[28]

Music sharing. The copyright holders for music and other content have taken a more direct approach to achieving their rights-protection aims—they are taking the fight to the end points themselves. They do this in a number of ways. For example, they have tried introducing their own (untrustworthy, from the end user's point of view) end nodes into some peer-to-peer systems to disrupt the delivery of illicitly shared content, and they attempt to identify sources of that content and take nontechnical (e.g.,

---

http://www.ietf.org/rfc/rfc2804.txt.

26. *See* Julien Pain, *Bloggers, the New Heralds of Free Expression*, *in* HANDBOOK FOR BLOGGERS AND CYBER-DISSIDENTS 5, 6 (Reporters Without Borders Sept. 2005), http://www.rsf.org/IMG/pdf/handbook_bloggers_cyberdissidents-GB.pdf [hereinafter HANDBOOK].

27. *See, e.g.*, Margaret Coker et al., *U.A.E. Puts the Squeeze on BlackBerry*, WALL ST. J., Aug. 2, 2010, at B1, *available at* http://online.wsj.com/article/SB10001424052748704702304575402493300698912.html?KEYWORDS=uae+puts+the+squeeze+on+blackberry; Bibhudatta Pradhan & Mark Lee, *India Seeks Permanent BlackBerry Solution from RIM, Pillai Says*, BLOOMBERG BUSINESSSWEEK (Sept. 16, 2010), http://www.businessweek.com/news/2010-09-16/india-seeks-permanent-blackberry-solution-from-rim-pillai-says.html.

28. The U.S. is not immune: A defense contractor announced a product aimed at monitoring social media use by client enterprise personnel in late 2010. *See Raytheon Unveils Cybersecurity Product*, UNITED PRESS INT'L (Sept. 17, 2010), http://www.upi.com/Business_News/Security-Industry/2010/09/17/Raytheon-unveils-cybersecurity-product/UPI-15531284735793/.

legal) action against them.[29] This is a classic example of end nodes that communicate even though they have no mutual trust and adverse interests. The long-term strategy of the rights-holders is to influence the hardware manufacturers to build what they call "trusted systems," which prevent the end users from performing certain actions on data that the rights-holders deem unacceptable. The term for this may be "trusted system," but it begs the question of "trusted by whom?"

B. "Good Guys" and "Bad Guys"

As we have noted in several places in this Article, while the original end-to-end paper used examples in which the two end points had a common interest in communicating, today more and more users who choose to communicate do not trust each other. Whether it is e-mail designed to defraud as in the case of phishing, a node in a peer-to-peer content distribution system that is designed to nab copyright violations, or a website that attempts to download malware or third-party tracking software onto an unsuspecting client, the Internet is full of examples where there is good reason for the ends of a communication not to trust each other.

In this context, the end-to-end argument is a two-edged sword. Since the end-to-end argument leads to a general-purpose network in which end users can run the application of their choice, without constraint from the network, it empowers both the "good guys" and the "bad guys." As the Internet seems to be increasingly overrun with bad guys, some security advocates deem the end-to-end argument itself as too dangerous to tolerate, since it is an enabler for bad guys. Further, the proliferation of malware transmitted by e-mail and the web provides some with an argument against end-to-end encryption, on the grounds that it makes filtering such material by service providers harder and therefore facilitates its circulation. On the other hand, the Internet Engineering Task Force has emphasized the value of end-to-end security, taking what some might call a "good guy"-centric position that, because in part of rampant exploitation of compromised end systems, development and use of secure protocols by end systems is critical for the Internet to serve the purpose of an international infrastructure.[30]

---

29. Peer Media Technologies offers "noninvasive" techniques (such as posting of false files and propagation of false signals) aimed at limiting illicit transfers of copyrighted materials on peer-to-peer networks. *See* PEER MEDIA TECH., http://www.peermediatech.com/services.html (last visited Feb. 20, 2011). A discussion of what have been called pollution and poisoning can be found in Nicolas Christin et al., *Content Availability, Pollution and Poisoning in File Sharing Peer-to-Peer Networks*, *in* EC '05 PROCEEDINGS OF THE 6TH ACM CONFERENCE ON ELECTRONIC COMMERCE, 68, 68, 75–77 (2005).

30. *See* Jeffrey Schiller, *Strong Security Requirements for Internet Engineering Task Force Standard Protocols*, IETF RFC 3365 (rel. Aug. 2002), http://www.ietf.org/rfc/rfc3365.txt. Schiller's 2002 RFC reiterates and amplifies the

We will revisit this point at several points in this Article. However, our overall approach is to reframe the end-to-end argument in terms of trust (where trust exists, and between which parties), rather than in terms of physical location (e.g., an "end point"). In this approach, adding protection to keep the bad guys from harming the good guys is consistent with (and integral to) the end-to-end argument, rather than being at odds with it.

IV. THE NEW END-TO-END

The discussion of what it means to be careful provides a framework for proposing a reformulation of the end-to-end argument for today's context: we can replace the end-to-end argument with a "trust-to-trust argument." The original paper said: "The function in question can completely and correctly be implemented only with the knowledge and help of the application standing at the endpoints of the communication system."[31] The generalization would be to say: The function in question can completely and correctly be implemented only with the knowledge and help of the application *standing at a point where it can be trusted to do its job in a reliable and trustworthy fashion*. Trust, in this context, should be determined by the ultimate end points—the principals that use the application to fulfill their purposes. Because the locus of trust is naturally at the ends, where the various principals are found, "trust-to-trust" is preferable to "end-to-end" from the point of view of the principals, because it more directly invites the important question of "trusted by whom?" That question, in turn, relates to questions that implicate application design, notably "who gets to choose which service is used?" or "which parts of an application are in which service modules?" Answers to these questions illuminate who controls what aspects of an application.

To reconstruct the end-to-end argument in the context of trust, we proceed in two steps. We first look at the range of options that each participant in the communication can take, based on their individual choices about trust, and then we look at the range of options that arise *jointly*, depending on the degree to which the various communicants trust each other. Trust-to-trust acknowledges that, unlike when the original paper was written, there is more reason for one end to question the trustworthiness of another and therefore more reason to seek something beyond simple end-to-end communication. As we noted in our earlier paper, the population of end users has become more diverse, and this raises questions for the end-to-end argument. [32]

"Danvers Doctrine" agreed to in 1995. *Id.* at 3.

31. Saltzer et al., *supra* note 1, at 278 (emphasis omitted).

32. Marjory S. Blumenthal & David D. Clark, *Rethinking the Design of the Internet: The End-to-End Arguments vs. The Brave New World*, 1 ACM TRANSACTIONS ON INTERNET

A. Trust Options for the Individual End Node

Each end user, to the extent the option is available, must make decisions about where services should be positioned so that they can be performed in a trustworthy manner. They can be positioned on a computer that is directly associated with the end user (the classic "end node" of the original paper), or they can be delegated to a service provider elsewhere in the network. A marketplace of providers and subscribers gives the end user control over which provider is selected to perform the service. Given choice, users can be expected to select services and service providers that they deem trustworthy. Only if the principals at the edge of the network, where they connect to it, are constrained from making choices about what agents to use, and are thus constrained to depend on agents that are not trustworthy, is this natural pattern of edge-determined trust broken. The above anecdote about problematic e-mail in Africa illustrates this point.[33] First, of course, the mail relay was unreliable. But second, the end users had no reasonable alternative but to use the mail relay of their ISP—they could not choose to move to another one, for reasons of ISP policy and pricing. There was thus no market incentive to motivate the provider to be reliable or trustworthy. This story also shows how end users may respond to untrustworthy agents by adding a new layer that they believe they can trust, in that case by trusting each other to use sequence numbers properly.

There are many reasons why the end user might be constrained in some way from making a choice to select trustworthy services and forced to use a service, whether or not she trusts it. An ISP can try to force its customers to use its own e-mail servers (most end users today depend on the DNS servers of the ISP, which influence where traffic is directed, without even thinking about whether it is wise to do so); and some ISPs try to force the end user to use an ISP-provided web proxy. Certain applications may be designed so there are few (or no) choices available to the prospective users as to the provider of the service. For example, a dominant social media service provider, such as Facebook, defines both hidden and visible aspects of its service; the user has no view into and no control over the hidden aspects. More generally, there are countries where all local agents may not be trustworthy, for reasons other than their use of inadequate or unreliable technology. For example, government interception may diminish the trustworthiness of all services available locally.[34] And in

---

TECHNOLOGY 70, 74 (2001).

33. *See supra* note 18 and accompanying text.

34. The OpenNet Initiative tracks such government-based interception. *See* OPENNET INITIATIVE, http://opennet.net/ (last visited Feb. 20, 2011). China now requires cell phone users to register for new accounts with their names to facilitate monitoring of the increasingly mobile Internet. *See* Loretta Chao, *China Starts Asking New Cellphone Users for ID*, WALL ST. J., Sept. 1, 2010, *available at*

developed, as well as developing countries, there is a growing number of reasons, including private sector monitoring for commercial purposes and selective blocking and filtering of communication, for end users to question the trustworthiness of available agents in at least some regards.

Constraint also comes from the operating system (and the browser) of the end user's computer. As we will discuss, the end user is more or less forced today to use one of a very small set of operating systems (and browsers). Whether or not the end user trusts those operating systems, convenience drives end users to use them. The power of convenience as a driver is manifest in the rapid growth in use of smartphones and other mobile devices, which support mobile Internet use.

In most of the examples we have listed of this sort, and in most countries today, the provider of the service has some motivation to provide services in a reasonably reliable and trustworthy manner. There are enough checks and balances in the system (through market or legal/regulatory mechanisms) to discipline a provider. But the match of expectations is often not perfect, as illustrated by the surge in concerns about privacy motivated by social media and other public cloud applications, and the end user is often forced into various sorts of compromises.

One of the most problematic situations is where a user is constrained to use an ISP that is not trustworthy. The ISP may indeed forward traffic correctly, but may monitor or log it. In this case, users with sufficient skills and knowledge invoke services (such as encryption) that disguise what is being sent. In other cases, the ISP may block the sending of certain traffic, or to certain destinations. Here, sophisticated users may invoke some sort of "higher-level" forwarding service, so that the ultimate destination of the communication is not visible to the ISP. Some dissidents in censorship-prone regimes resort to third parties in different countries to get their messages out on their behalf, perhaps without attribution.[35] Tools such as onion routing[36] can be used to disguise both the content and the destination of a transmission; it essentially overlays the routing algorithm of the ISP with a separate routing scheme carried out by (presumably) more trustworthy nodes.

---

http://online.wsj.com/article/SB10001424052748704791004575465190777886192.html?KEYWORDS=china+requires+id+cellphone+customers.

35. *See* Nart Villeneuve, *Technical Ways to Get Round Censorship*, *in* HANDBOOK, *supra* note 26, at 63, 75. The U.S. government has funded the development of software for this purpose. *See, e.g.*, *Freegate*, DYNAMIC INTERNET TECH., http://www.dit-inc.us/freegate (last visited Feb. 20, 2011).

36. For a description of onion routing, see TOR PROJECT: ANONYMITY ONLINE, http://www.torproject.org (last visited Feb. 20, 2011).

B. *Delegation of Function*

E-mail and content distribution as described above, as well as the example of checking for viruses and spam, illustrate what we might call *delegation* of function to a trusted agent. The content producers trust the CDN, and they delegate the delivery function to it. In most cases, end users trust their mail agents (in contrast to the story about the African service), and they delegate the transfer of mail to these services. We could draw a circle around each end point and the servers (including supporting services such as the DNS) the user has chosen to trust, and (at the application layer) we could call this an *application* end point.

Figure 1 illustrates how the e-mail system might be drawn. To get between parts of this "higher-level" end point it will be necessary to make use of the lower-layer communications subsystem, and there will be reliability mechanisms designed and used at that level. At this lower level, the end-to-end argument will apply as each part of the service communicates with the other parts. At a higher level, there is a different interpretation of the end-to-end argument, as one application end point talks to the other application end point.

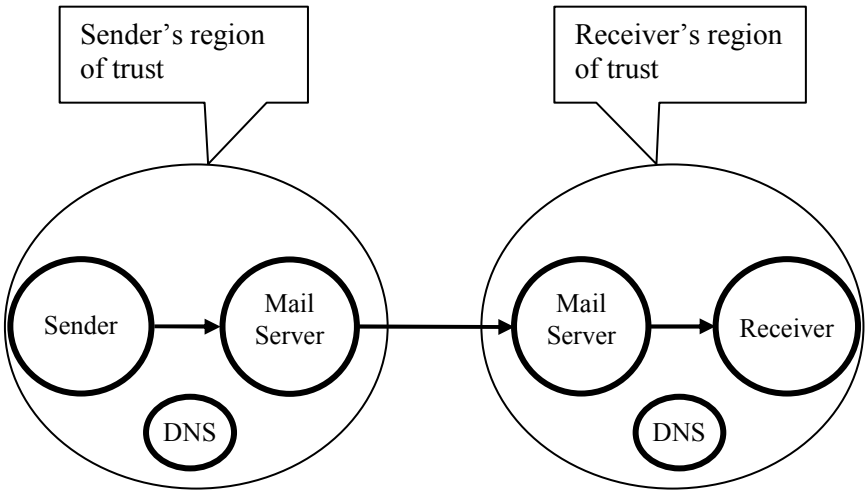


Figure 1: Regions of trust in email forwarding system

C. *Mandatory Delegation*

In the real world, there are many circumstances where an individual user does not have control over which services and servers to use. Perhaps the most obvious example is the context of employment, where individual

employees are constrained by corporate policy to use their computers in certain ways, to use only certain servers (e.g., their corporate e-mail or instant message servers), and so on. The fact that the user is forced to use these services does not automatically make them untrustworthy. Some employees may be entirely comfortable with the way their employers operate their IT infrastructure; others may have fears about surveillance, logging, or other practices. But whatever judgment the employee makes about the trust to place in his or her circumstances, he or she has no realistic control over the situation (although alternative platforms and applications may be chosen for personal use).

There is a useful analog between this situation and a duality that arises in the security community. Certain security controls are cataloged as "discretionary access controls," or DACs, and "mandatory access controls," or MACs.[37] MACs originated from the need to enforce rules for the proper handling of classified information, and access decisions were taken away from the individual at his computer and given to a system security administrator, who would impose access controls based on corporate or institutional security policies.[38] Because the individual user had no control over these mechanisms, they were called *mandatory*, which is a word that signals that somebody other than the end user has the discretion to control them.[39]

One way to capture the range of situations that apply to the individual end user is illustrated in Figure 2.

37. *See, e.g.*, SYS. SEC. STUDY COMM. ET AL., COMPUTERS AT RISK: SAFE COMPUTING IN THE INFORMATION AGE 251 (1991).

38. *See id.*

39. An illustrative example of such a control in the network context is an intrusion detection system. Such systems look at incoming and outgoing traffic to attempt to detect patterns that suggest an ongoing attack. They can benefit from seeing the traffic to and from many nodes, not just one. They are often installed by network managers (so they are mandatory from the perspective of the end user), and they are generally viewed as benign.

	Likely degree of end user trust: Lower	Likely degree of end user trust: Higher
Mandatory selection of mechanism	Monopoly provider Government intervention	Employee
User choice or discretion over mechanism selection	Uncommon: given choice and knowledge, users select trusted option	Competitive market with consumer choice

Figure 2: Range of options for choice and trust

D. When End Users Do Not Trust Each Other

In the analysis above, we defined "trust end points" by drawing circles around end points and the various trusted services to which they have chosen to delegate functions. But once we have drawn these circles, an important question remains—do the different trust end points trust each other?

For most of the examples in the original paper, the answer is yes. In the example of careful file transfer, the two end points are collaborating to make the transfer reliable. But as we hinted above (using the example of viruses), we often communicate with other end points that are not trustworthy and/or that we do not choose to trust. How do we deal with this situation?

One class of response is to try to devise and deploy defenses inside each circle of trust that are robust enough that the other end point cannot inflict any material damage. We deploy virus checkers, spam filters, and so on, and then we cautiously try to exchange e-mail.

But the other class of response is to invoke the services of a mutually trusted third party to remove some of the risk of the interaction. I do not trust you, you do not trust me, but we both trust this other party— perhaps that other party can help us interact. The real world is full of examples of this sort—trusted institutions of all sorts are what make contemporary, economically developed society function, from banks to contract law and courts to credit card companies and various sorts of negotiators. In the real world, when two parties view each other with suspicion, they seldom try to

resolve the problem on their own.

And we see more and more the emergence of online analogs. For example, credit card companies, because they can verify the identity of all parties and because they protect against fraudulent actions, act to add trust so that transactions can be completed between untrusting end users and merchants. Providers of digital certificates assist in the authentication of communicants that may not trust each other. And today, a variety of projects aim to provide identity-management services in ways that suggest new categories of third-party actors facilitating trust.[40] By providing assurance-supporting services such as identity management and insurance against specific risks, such third parties permit untrusting parties to decide to take the risk of interacting. More directly, many applications are designed so that services developed and operated by the designer of the application are interposed between the end users. When two end users communicate using popular instant messaging applications today, they do not directly connect across the communications subsystem. Instead, the IM communications are relayed through an IM server run by the service itself. This service enhances many aspects of the overall function. For example, the centralized management of identities provides level of confidence to the users about the identities.. Second, the service provides isolation between the end users. Since end users do not communicate directly, they need not reveal low-level information such as IP addresses to each other, which prevents them from attacking each other directly across the communications subsystem.

Similarly, eBay, interposed between buyer and seller, provides a neutral meeting ground (more specifically, a neutral place where markets can be made). eBay also illustrates the role of reputation in assessing trustworthiness: eBay is a third party that facilitates communication about reputation and implied trustworthiness. This is one way that identity can be invoked for trust.

For some applications, for example multi-player games, it is fairly obvious that much of the implementation of the game resides on servers rather than on the end nodes of the players. This structure arises both because there is a great deal of shared information (about the game and its state of play) among the players that must be coordinated, and also because the players must be kept from cheating. The players certainly want to communicate, but they just as certainly do not trust each other.

Here is a partial list of functions that a trusted third party might

---

40. *See, e.g.*, HIGGINS: OPEN SOURCE IDENTITY FRAMEWORK, http://www.eclipse.org/higgins/ (last visited Feb. 20, 2011); SHIBBOLETH, http://shibboleth.internet2.edu/ (last visited Feb. 20, 2011) (explaining the single sign-on approach); OPENID, http://openid.net/ (last visited Feb. 20, 2011).

perform:

- Manage identity, in many ways
- Facilitate appraisal of reputation
- Provide robust authentication (prevent identity theft, prevent fraud)
- Control the release of attributes (limit what one party can see about others, e.g., IP addresses)
- Preserve anonymity (extreme form of controlled release—sender wants to hide all aspects of his identity from receiver)
- Protect end users from each other
- Prevent attacks
- Regulate and filter content
- Prevent cheating (e.g., in games)
- Provide mutual assurance and guarantees (escrow, fraud insurance, nonrepudiation)

Sometimes the third party software is directly interposed in the communication path between the end nodes, as with instant messaging, games, eBay, and the like. In other cases, the third party is not literally in the communication path between the two untrusting users but is invoked by one or both of those parties to augment the trustworthy nature of the overall transaction. It is tempting to try to analyze the implications of trusted third parties for the end-to-end argument by looking to see if the third party is literally in the path of communication. If we allow ourselves to fall back to a lower-level view of end-to-end, looking at the role of the communications subsystem, models where the third party is "off to the side" (invoked by one of the end nodes) might seem more "end-to-end." But we would argue that this lower-level detail is irrelevant in an analysis of trust, which is the basis for our higher-level model. If two parties decide to involve a trusted third party, then that party is in the middle of the "path of trust," regardless of whether that party is in the middle of the packet flow. We should not be concerned with how the packets flow, but instead look at which aspects of the trust depend on our mutual dependence on that third party, and which aspects we can determine for ourselves.

The choice as to whether to invoke a third party to enhance trust in a particular application is usually not up to the individual user. It will usually be embedded into the design of the specific application at hand; in other words, the designer of the application has control over the patterns of communication and thus the "architecture of trust." Whether or not a buyer and a seller on eBay have reason to trust each other, they must interact in the context of the marketplace defined by eBay. This fact begs the obvious question as to whether it is any more reasonable for end users to trust the third-party service provider than to trust each other. One way to try to

answer this question would be by analogy to the original end-to-end argument, where one might argue that it is better for the end nodes to solve what problems they can by themselves, because involving a third party can only add to the complexity, and perhaps to the lack of certainty about trust.[41] An issue for the design and operation of such third parties, as recently publicized identity-theft cases illustrate, is to avoid having them emerge as a bigger, let alone just another, source of vulnerability. To some observers who are concerned about the loss of personal control, the use of certain kinds of remotely provided services (services "in the cloud") is a major source of risk.[42] But the outcome of the analysis, in this case as in the original paper, is not a dogmatic stricture but a preference to be validated by the facts of the situation. And this construction by analogy may be nonsense. While there are specific reasons to assume that the communications system will be unreliable, there is no similar reason to assume that third-party services are intrinsically unreliable. The decision will be based on a trust assessment, as well as considerations of convenience and utility. So perhaps at this level there should not be a preference for end-to-end patterns of communication, but a preference for the use of third-party services and multiway patterns of communication—that is the kind of thinking that has contributed to growth in demand for cloud services.

In the marketplace of the 2000s, a number of developments shift activities away from end nodes. "Service oriented architecture" (SOA) is a buzzphrase for accessing a variety of applications and data over a network. It is linked to a model in which end users, within some enterprises, access what they need from servers as they need it, rather than investing in capabilities at their individual end nodes. It is also a concept fundamental to social media and various public cloud applications. For example, Google's move to provide office-productivity capabilities aims to motivate end users, as individuals and as members of enterprises, to use capabilities hosted on its servers rather than at the end nodes (or servers controlled by the enterprise).[43] This mode of operation, combined with a style of operating the end node in which no new software or functions can be downloaded or installed, tries to accomplish stable operation through delegation and outsourcing.

---

41. This is a big question for cloud computing, at least for the public cloud services. *See* Blumenthal, *supra* note 21.

42. *See, e.g.*, Richard Stallman, *What Does That Server Really Serve?*, BOS. REV. (Mar. 18, 2010), http://bostonreview.net/BR35.2/stallman.php (revised version available at http://www.gnu.org/philosophy/who-does-that-server-really-serve.html).

43. *Stay Connected and Be More Productive*, GOOGLE APPS, http://www.google.com/apps/ (last visited Feb. 20, 2011).

V. THE ULTIMATE INSULT

The erosion of the end-to-end argument is often equated to the emergence of intermediate servers and services not located at the end points. As we have argued, this is not necessarily so. If the end user has a choice and can pick services that he trusts, this can be seen as delegation and the creation of a distributed end point. The more fundamental erosion of the end-to-end argument is that the end user can no longer trust his own end node—his own computer. There are forces, both lawful and unlawful, that try to shift the balance of control and trust away from the end user toward other parties such as rights holders. Malicious software such as spyware and key loggers—sent by malicious end systems—try to attack the reliability and trustworthy nature of typical end user activities by penetrating the end node computer and turning it against the end user or against other end users. Criminal elements make surreptitious use of large numbers of end nodes owned or used by others via botnets that attack, send spam, and otherwise make mischief for yet other end users. Legitimate businesses seeking advertiser support tolerate tracking software that can compromise end user privacy.[44]

Whatever the cause for distrust, what is the future of the end-to-end argument if the end user cannot trust his own computer to behave reliably? This trend could signal the end of end-to-end, and more catastrophically, the end of any ability to make rational trust assumptions at all. If the end user cannot trust her own computer, what can she trust?

A. Can We Take Back the End Node?

One response to end users' diminishing ability to trust their own end nodes might be further delegation, as mentioned above: to move away from using the computer as a platform for trustworthy activities, and to move those activities to servers provided by operators who seem to be able to offer them reliably. This approach would signal the return (yet again) of the thin client and a "services architecture" for applications. Using our analysis, what would be required to make this work? First, this scheme would still require a trustworthy path of communication from the end user to the service. This path has to reach all the way to the human user—this implies that what the end user sees on the screen is what the service wanted

---

44. *See* Nick Wingfield, *Microsoft Quashed Effort to Boost Online Privacy*, WALL ST. J., Aug. 2, 2010, at A1, *available at* http://online.wsj.com/article/SB10001424052748703467304575383530439838568.html; Steve Stecklow, *On the Web, Children Face Intensive Tracking*, WALL ST. J., Sept. 17, 2010, at A1, *available at* http://online.wsj.com/article/SB10001424052748703904304575497903523187146.html?mod=WSJ_article_RecentColumns_WhatTheyKnow.

to put there.[45] The potential for a key logger on the client, no matter how thin the client, destroys the trustworthy nature of the scheme. The need for a trusted path might lead to a model of end node software where the machine has a fixed set of software and no ability to download any active code or new applications. Second, to make this scheme viable, service providers who declare that they are going to offer a trustworthy service must be able to do so. If their servers are susceptible to being infested with spyware or are readily intercepted for censorship or surveillance purposes, we are no better off.

Another approach is to try to reclaim control of the end node, both by reducing vulnerability (bugs) and by allowing the end user to know what is in the system. Part of the appeal of Linux is that since the code is open, skilled programmers can read it and try to verify that there are not any intentionally installed controls and features that make the machines using it less trustworthy and less suited to the needs of the end user.

VI. DESIGN FOR DELEGATION

If we agree that it is useful in certain cases for end nodes to delegate functions to servers and services within the network, then applications have to be designed to make this both possible and easy. The application has to be broken up into parts connected by well-specified protocols that seem to represent useful functional building blocks. This act of modularization, of course, takes a lot of judgment, and is probably best suited to be the subject of a book, rather than an article. Assuming that the application has been properly modularized, there are then some further points that arise from the discussion of trust and the reality of both trusted and untrusted third parties.

First, one can ask whether the modularization of the application allows the trust assumptions to be violated in unexpected ways. For example, one of the ways that untrusted third parties can insert themselves into an application is by interjecting themselves into the path of a well-specified protocol—the sort that is designed to allow functional decentralization—and playing the part of the other communicant. One of the implications of an open and documented protocol is that since any actor can "speak the language," it may be possible for a third party to insert itself into the middle of a path and pretend that it is the intended destination of

45. This idea is not new, of course. It relates to the idea of a "Trusted Path" in secure computer systems, as articulated in the Trusted Computer System Evaluation Criteria. DEP'T OF DEF. STANDARD, Trusted Computer System Evaluation Criteria No. DoD 5200.28/STD (1985), *available at* http://csrc.nist.gov/publications/history/dod85.pdf. This reference defines a Trusted Path as "[a] mechanism by which a person at a terminal can communicate directly with the Trusted Computing Base," which emphasizes that the trusted path must reach all the way to the human user to be effective. *Id.* at 113.

the conversation.[46] A (mostly) harmless example of this occurs quite often when an Internet user at a hotel or WiFi hot-spot tries to send mail. It is often the case that the connection back to the Simple Mail Transfer Protocol (SMTP) server chosen by the end user is redirected to a different SMTP server operated by the local provider. The hotel intends this to be a helpful feature (it solves the problem that not all SMTP servers will accept connections from distant parts of the network), but at a philosophical level, it represents a complete overriding of the end user's right to choose which service to use. Protocols should be designed so that the end user who makes the choice of which service and servers to use maintains control over that choice. Distributed elements should always be able to tell which other elements they are talking to, and it should not be possible to subvert the protocol so that untrusted parties can exploit them to insert themselves. Tools (often based on encryption) that provide assurance about identity and nondisclosure can ensure that only the services chosen by the end nodes are the ones being used.

Second, trust is with respect to a given role. I may be willing to trust a router to forward my packets—or, putting this differently, there may be enough constraints that I can count on the router to forward my packets even if I do not fully trust it—but I may not trust it to protect my packets from disclosure. If the protocols that are designed to allow functional decentralization and delegation are designed so that the capabilities of the servers and services are limited to the intended functions, then we need not make as strong a trust assumption about these devices, which will provide more flexibility regarding which services we are prepared to choose. For example, if different parts of the application payload are encrypted and/or signed (so an intermediate cannot see or change them) and other parts are revealed, this can allow servers to be employed without having to trust them to preserve all aspects of the information.[47]

An important aspect of application design applies to protocols and mechanisms that can operate both in the context where the end users trust each other and where they do not. If the end users have the choice among

46. In security parlance, when a malicious node manages to insert itself into the middle of a conversation, pretending to each of the communicants to be the other communicant, this is called a "man in the middle" attack. It may give the attacker (or more generally the third party with adverse interests) the ability to see and modify anything that is being transmitted.

47. Of course, if the design process for the application included an explicit discussion about which parts of the payload should be encrypted or revealed, this might trigger vigorous advocacy among the different stakeholders as to how the application should be designed. There is a parallel with the debates that occurred during the design of IPsec—the IP level encryption standard—where there were competing views as to which parts of the original packet header should be hidden and the eventual development of two alternatives (Encapsulating Security Payload and Authentication Header) that offer a different answer to this question.

invoking a third party, using mutual checks and constraints, or communicating openly based on mutual trust; and if the application can easily adapt to all of these modes, then it becomes more practical for the end users to operate in each of these modes and to move among them as they deem appropriate.

Research is driving some new approaches to the architecture of social media applications that restore some control to end users. Recent research projects illustrate the impact of different choices about modularizing applications. The Lockr system, for example, decouples information about an end user's social network from distribution of content to members of that network, allowing end users to limit the number of services with which they share their social network information.[48] It also provides for asymmetric relationships among people in a social network and revocation of relationships.[49] Another approach is taken by the authors of the proposed PrPl "person-centric" infrastructure for storing and sharing information with "fine-grained access-control."[50] These specific examples illustrate that application design and modularity can enhance or reduce options for user choice. Different designers will have different motivations to offer or constrain choice, and thus control the degree to which a user can make personal decisions about trust within specific applications. Our earlier example of e-mail illustrated an application based on a design that gives the user choice.

We have taken the view here that if some principal chooses to trust some agent and, for example, delegates function to it, this should lead to a system that is just as trustworthy as a system in which all the functions are carried out on the end node. The IETF has explored this space, and its analysis illustrates the limits of its willingness to depend on trust, as assessed by the user, as a building block of a trustworthy system. Several years ago, an IETF working group was proposed to design what was called Open Pluggable Edge Services, or OPES.[51] The OPES proposal was essentially an architecture for delegation, and it triggered a controversy in the IETF that led to a policy assessment of the OPES concept by the Internet Architecture Board.[52] This assessment reached several of the same

---

48. *See* Amin Tootoonchian et al., *Lockr: Better Privacy for Social Networks*, INTERNATIONAL CONFERENCE ON EMERGING NETWORKING EXPERIMENTS AND TECHNOLOGIES (CONEXT) (2009), http://conferences.sigcomm.org/co-next/2009/papers/Tootoonchian.pdf.

49. *Id.*

50. Seok-Won Seong et al., *PrPl: A Decentralized Social Networking Infrastructure*, ACM WORKSHOP ON MOBILE CLOUD COMPUTING & SERVICES: SOCIAL NETWORKS AND BEYOND (MCS) (2010), http://prpl.stanford.edu/papers/mcs10.pdf.

51. *Description of Working Group*, OPEN PLUGGABLE EDGE SERVICES (OPES), http://datatracker.ietf.org/wg/opes/charter/ (last visited Feb. 20, 2011).

52. Memorandum from Sally Floyd & Leslie Daigle, *IAB Architectural and Policy Considerations for Open Pluggable Edge Services*, IETF RFC 3238 (rel. Jan. 2002),

conclusions that we do:

- Delegation is only acceptable if one end or the other has explicitly put it in place (that is, injection of service elements by unrelated actors should not be permitted by the architecture).[53]
- Messages being sent to the service element should be explicitly addressed to the element, and tools such as encryption should be used to ensure that only the expected elements are participating in the delegation.[54]

However, after reaching these conclusions, its analysis suggests that the IAB had an instinctive reaction that services delegated to a server were somehow intrinsically less trustworthy than services running locally on the host. The assessment called for the addition to the architecture of technical means for an end node (or the principal using the end node) to be able to check or review what the service element had done. It says:

> [W]e recommend that the IESG require that the OPES architecture protect end-to-end data integrity by supporting end-host detection and response to inappropriate behavior by OPES intermediaries. We note that in this case by "supporting end-host detection", we are referring to supporting detection by the humans responsible for the end hosts at the content provider and client.[55]

One could see this recommendation as arising from the traditional roots of the Internet, where the users are technically sophisticated and able to fall back on technical intervention to validate what a server is doing. In today's Internet, most users do not have the skills to verify (technically) what a program is doing, whether it is running on their own machine or on a server. Today, most users select and use a program based on some assessment of its suitability and trustworthy nature, no matter where it runs.

VII. REINTERPRETING THE END-TO-END ARGUMENT

If this Article represents a significant (re)interpretation of the original end-to-end argument, it is part of a larger tradition of reinterpretation. Perhaps because the argument is described in the original paper as much by example as by definition, there has been a rich history of assertion and speculation about how to interpret the end-to-end argument, and what it really means. This section surveys some of that history to put our Article into a larger context.

The original paper states the end-to-end argument in terms of how function must be placed to achieve correct operation and to align with

---

http://www.ietf.org/rfc/rfc3238.txt.

53. *See id.* at 13.

54. *See id.*

55. *Id.* at 1.

application-level semantics. There is an implication that a system built according to this approach is more general, in that it is not designed to support a specific, known set of applications. However, the benefit of generality is implicit—it is not directly argued in the paper. This virtue is often associated with the *open* nature of the Internet, although the word "open" hardly appears in the paper.[56]

The importance of openness was spelled out for a broad audience in an interpretive work crafted by a committee involving the authors of this Article and others from networking and other fields. Published and extensively presented in 1994, *Realizing the Information Future: The Internet and Beyond*[57] articulated in plain English the virtues of the Internet and served to educate a wide range of U.S. and foreign policy makers, industry executives, and civil society leaders about the concept of an "Open Data Network," exemplified by the Internet. The Open Data Network is defined as open to users, service providers, network providers, and change,[58] and the book calls for research to further the development of "general and flexible architecture" for networking and the development of security architecture.[59] It also noted that the logic of an Open Data Network implied the unbundling of higher-level applications and services from lower-level networking functions.[60]

The authors of the original paper expanded on the implications of the end-to-end argument for application innovation in a 1998 paper,[61] motivated by a research program called Active Networks.[62] Beginning

---

56. Note that a well-known amplifier of the end-to-end argument, IETF RFC 1958, also does not use the word "open"; it appears that more social and economic experience with the Internet was needed before the concept was broadly appreciated. *See* Brian Carpenter, *Architectural Principles of the Internet*, IETF RFC 1958 (rel. June 1996), http://www.ietf.org/rfc/rfc1958.txt.

57. *See generally* NRENAISSANCE COMMITTEE, COMPUTER SCI. AND TELECOMM. BD., NAT'L RES. COUNCIL, REALIZING THE INFORMATION FUTURE: THE INTERNET AND BEYOND (1994).

58. *Id.* at 44.

59. *Id.* at 93.

60. *Id.* at 51.

61. David P. Reed et al., Commentary, *Commentaries on "Active Networking and End-to-End Arguments,"* IEEE NETWORK, May/June 1998, at 69–70. This states, among other things, that

> [p]art of the context of an end-to-end argument is the idea that a lower layer of a system should support the widest possible variety of services and functions, to permit applications that cannot be anticipated. . . . Higher-level layers, more specific to an application, are free (and thus expected) to organize lower-level network resources to achieve application-specific design goals efficiently (application autonomy).

Id. at 70.

62. *See generally* David L. Tennenhouse & David J. Wetherall, *Towards an Active Network Architecture*, COMPUTER COMM. REV., April 1996. The Active Networks program

shortly thereafter, as Internet virtues became more evident to a wider range of people, other authors championed the open nature of the Internet, focusing on its ability as a platform to support a wide range of unanticipated and unplanned applications. This open nature has economic and social impacts, which, as we noted in our earlier paper cited above, have motivated rhetoric by advocates of various sorts. Most prominently, Larry Lessig has used the end-to-end argument as the basis for a defense of the open nature of the Internet as an enabler of third-party innovation and what has become known as "network neutrality."[63] David Reed, one of the authors of the original paper, has reflected on the roots of the end-to-end argument, the push by telecommunications companies for more centralized control as the broadband market grows, and the chilling effect on innovation associated with in-network chokepoints.[64] Another author of the original paper, Jerry Saltzer, has chronicled "gatekeeping restrictions" arising in cable-company Internet service.[65] He has been quoted as noting that such restrictions are at odds with the end-to-end argument and,

---

was a DARPA-sponsored research project to explore a novel networking approach in which packets carry code that can be executed by routers to modify their operation. While this idea might be seen as the antithesis of the end-to-end approach, as it could move application or service-specific function into every router, the commentary cited below gives a nuanced view. *See infra* note 64.

63. *See, e.g.*, Lawrence Lessig, *It's the Architecture, Mr. Chairman*, BERKMAN CENTER FOR INTERNET AND SOC'Y, HARVARD U. (1996), http://cyber.law.harvard.edu/works/lessig/cable/Cable.html. Lessig observes,

> The Internet has a constitution. Its architecture is this constitution—the way the net is coded, its design, the principles that govern its control. Like any constitution, this architecture embeds certain values. These values have consequences. In the case of the Internet, they have produced the greatest space of innovation that we have seen this century. . . . The value at stake is a design principle called "end-to-end."

Id. at 1. Similar ideas are expressed at greater length in a variety of Lessig's writings around the turn of the century. *See, e.g.*, Lawrence Lessig, *The Internet Under Siege*, FOREIGN POL'Y, Nov. 1, 2001, *available at* http://www.foreignpolicy.com/articles/2001/11/01/the_internet_under_siege.

64. David P. Reed, *The End of the End-to-End Argument*, REED'S LOCUS (Apr. 2000), http://www.cs.sfu.ca/~vaughan/teaching/431/papers/ReedEndOfTheEndToEnd.pdf ("Today's applications (eCommerce storefronts, telephone calls routed over IP networks, streaming video broadcast of Hollywood movies, and banner-ad-sponsored web pages) are being used to justify building in idiosyncratic mechanisms into the network's core routers and switches. Though it is clearly not possible to meet the requirements of today's hot applications solely with functionality in the network's core, we are being asked to believe that this is the only possible architecture. Implicitly, we are being told that the impact of building these structures into the network is worth the cost of erecting major barriers to future innovation. . . . In the Internet's end-to-end design, the default situation is that a new service among willing endpoints does not require permission for deployment. But in many areas of the Internet, new chokepoints are being deployed so that anything new not explicitly permitted in advance is systematically blocked.").

65. Jerome H. Saltzer, "Open Access" Is Just the Tip of the Iceberg (Oct. 22, 1999) (unpublished article), http://mit.edu/Saltzer/www/publications/openaccess.html.

therefore, a threat to innovation.[66] He continues to observe shrewdly that people are not passive in the face of such corporate conduct, suggesting that effective responses can arise from consumer behavior and/or government regulation.[67]

Barbara van Schewick, in her dissertation[68] and book,[69] has undertaken an extensive analysis of the economics of the Internet market, which she prefaces with a thorough and careful review of work that interprets and contextualizes the original end-to-end argument in various ways. Van Schewick asks what it means to adhere to the original argument when its own authors varied the wording over time. In the original paper, the authors wrote: "The function in question can completely and correctly be implemented only with the knowledge and help of the application standing at the end points of the communication system."[70] In their 1998 commentary on the end-to-end argument and active networks, they wrote a somewhat different sentence: "[A] function or service should be carried out within a network layer only if it is needed by all clients of that layer . . . , and it can be completely implemented in that layer."[71] Van Schewick calls the earlier version "narrow" and the later version "broad," and then considers how the economics vary with the version.[72] The analysis in this Article is consistent with either version of the end-to-end argument.

In addition to openness and flexibility, simplicity (of the communications subsystem) has also been identified as a benefit of the end-to-end argument. The original authors discuss these benefits of the end-to-end argument in their 1998 commentary, where they argue for the architectural benefit of "moving function from lower layers to more application-specific layers . . . ."[73] They explain that "building complex functions into a network implicitly optimizes the network for one set of uses," arguing that "an end-to-end argument . . . strongly suggests that enthusiasm for the benefits of optimizing current application needs by

---

66. *Id.*

67. *See id.*

68. Barbara van Schewick, Architecture & Innovation: The Role of the End-to-End Arguments in the Original Internet (July 21, 2004) (unpublished Ph.D. dissertation, Technische Universität Berlin), http://www.lessig.org/blog/archives/van%20Schewick%20Dissertation%2012102004.pdf.

69. BARBARA VAN SCHEWICK, INTERNET ARCHITECTURE AND INNOVATION (2010).

70. Saltzer et al., *supra* note 1, at 278 (emphasis omitted).

71. Reed et al., *supra* note 61, at 69.

72. SCHEWICK, *supra* note 69, at 5. This is the most detailed textual and economic analysis to date. Its almost Talmudic character begs the question of how important is the exact wording used by technologists who acknowledge that their own understanding of their subject has grown with time and experience.

73. Reed et al., *supra* note 61, at 70.

making the network more complex may be misplaced."[74] The 1998 paper reflects the broad acceptance of the layered-system architectural paradigm, deeper understanding of the challenges posed by system complexity as a result of technical and economic activity since the original paper, and insight into evolving views of the tension between programmability and flexibility on one hand, and specialization on the other. Specialization, or the adding of function to facilitate specific applications, can privilege specific uses and users by making what they do more efficient.[75]

The idea of trust as a fundamental tool for the analysis and application of the end-to-end argument is not original to this Article. Consistent with our discussion herein, the previously cited *Realizing the Information Future* observed that, "If the [National Information Infrastructure] is to flourish, we must provide solutions so that any end node attached to the network can mitigate its risk to an acceptable level."[76] More recently, Tim Moors examined the influence of responsibility and trust on the end-to-end argument.[77] His emphasis on the role of trust is very similar to our point of view, but his analysis focuses on lower-level functions such as congestion control.[78] He observes that in today's commercial environment (as opposed to the smaller, nonprofit community of the early Internet years) it is naïve to expect end points to behave altruistically (e.g., in terms of refraining from congestion-inducing behavior).[79] He also points out the need to identify the end nodes carefully as part of understanding "what entity is *responsible* for ensuring that service, and the extent to which that entity can *trust* other entities to maintain that service."[80]

Kempf and Austein assert that "the single most important change from the Internet of 15 years ago is the lack of trust between users,"[81] underscored by the rise of "deliberate, active attacks on the network infrastructure and end nodes."[82] They argue that that lack of trust drives

---

74. *Id.*

75. The companion piece by Partridge, et al., suggests that growth in understanding of complexity and programmability shift the balance toward more programmability in network management while preserving simplicity in the Internet layer to assure broad connectivity. *See* Craig Partridge et al., BBN Techs., Commentary, *Commentaries on "Active Networking and End-to-End Arguments,"* IEEE NETWORK, May/June 1998, at 67–69.

76. NRENAISSANCE, *supra* note 57, at 79.

77. Tim Moors, *A Critical Review of "End-to-End Arguments in System Design,"* 5 IEEE INTERNATIONAL CONFERENCE ON COMMUNICATIONS 1214 (2002).

78. *See id.*

79. *Id.*

80. *Id.* at 1219.

81. Memorandum from James Kempf & Rob Austein, *The Rise of the Middle and the Future of End-to-End: Reflections on the Evolution of the Internet Architecture*, IETF RFC 3724, at 5 (rel. Mar. 2004), http://www.ietf.org/rfc/rfc3724.txt.

82. *Id.* at 8.

choices by application and system designers about authentication, and they observe that

> One of the most common examples of network elements interposing between end hosts are those dedicated to security . . . .[they] are designed to protect the network from unimpeded attack or to allow two end nodes whose users may have no inherent reason to trust each other to achieve some level of authentication."[83]

Those users, in turn, need to "determine which third parties they trust."[84] Third parties, such as ISPs, have their own interests (e.g., making profits) to address, and while they can serve as "trust anchors" by acting to protect end users, they can insert mechanisms to support their own policy (e.g., censorship) into the network.[85] Kempf and Austein caution against application design that creates dependencies among protocols and system layers, citing the controversy (discussed above) associated with Open Pluggable Edge Services.[86] They assert that

> the trust relationships between the network elements involved in the protocol must be defined, and boundaries must be drawn between those network elements that share a trust relationship. The trust boundaries should be used to determine what type of communication occurs between the network elements involved in the protocol and which network elements signal each other.[87]

They suggest that the right approach to decomposition allows for the end-to-end argument to apply internally to an application, and while it may not apply to the application as a whole, this approach can assure the benefits that have come to be associated with the end-to-end argument, such as innovation and robustness.[88]

VIII. CONCLUSIONS

We have argued that "trust-to-trust" is an important generalization of end-to-end. The original paper equated the end node with the trusted node, and therefore it did not elaborate on this issue. But we argue that the fundamental basis for placement of function is that it is placed where it can be trusted to carry out its function reliably. Our preference, consistent with the end-to-end argument, is that the end user should have control over the trust decisions. It is the movement of trust to the edge that is consistent with the end-to-end argument, not the placement of all function at the end node.

The inability of the end users to trust their own computers (their end

---

83. *Id.* at 5.
84. *Id.* at 6.
85. *Id.* at 7.
86. *See id.* at 3–5.
87. *Id.* at 8.
88. *Id.* at 10.

nodes), and uncertainty about this, is the most corrosive problem for the end-to-end argument, not the placement of services in the net, per se. Accordingly, we have highlighted the challenge of designing trustworthy end nodes.

The original reasoning about the communication subsystem remains valid. We now have a "two layer" end-to-end argument, and a more complex "the rest," where "the rest" is broken up into regions based on trust.

We have mentioned economics and the discipline of competition. We argue that the "trust architecture" is the most fundamental factor, and the economic architecture can only be understood in the context of the trust architecture. With market power, monopolists can attempt to trump trust; furthermore, governments may erode trust in other ways (but they also have ways to enhance trust). If market power is the only force undermining trust, the applications may be designed to work around this and recreate the desired trust relationship. In countries where governments make "lawful" interception pervasive, application work-arounds may remain limited, and so may the experience of communication that can be described as trust-to-trust. Depending on the regime, the notion of trust may be more or less nuanced—and that variation may be tempered by movement among governments to collaborate in combating cybercrime and related concerns.

We have identified a number of reasons why it might be beneficial to design applications so that parts of the application function are positioned, if not "in the network," then in a more broadly distributed implementation of the application—that is, at intermediate points rather than at the end point computers associated with the end users:

- The operator of an end node (the end user) may not want to go to the effort of providing the service with the desired level of reliability. It may be easier to delegate or out-source it.
- By performing the function at an intermediate point, the service may have access to more information (e.g., the state of many end users, not just one).
- By performing the function at an intermediate point, the end user can avoid the cost and overhead of transferring unwelcome traffic across the communications subsystem to the ultimate end point.
- An end machine might have a vulnerability that would allow a virus to attack it before a virus checker on the machine could detect it. Doing the check at an intermediate point can protect the machine from a vulnerability that its owner cannot rectify.
- Pre-positioning information at an intermediate point can make

> the subsequent delivery more responsive as well as more reliable. Replicated intermediate points can specifically improve reliability.

For each of these reasons, of course, there is a range of further considerations, which, as in the framing of the original end-to-end argument, must be seen as a starting point for the consideration of the inevitable second-order effects, not as dogma.

All of these reasons seem to fit within the paradigm of *delegation*. That is, a service of these sorts would be deployed as part of an application because one of the end points chose to do so, based on a unilateral assessment of trust, function, and reliability. We could refer to the "trust circles" in Figure 1, and in most of the cases above we could include the server for such services unambiguously inside the circle belonging to one specific end point. This was the "trust-to-trust" model with end nodes that were distributed at the application level.

On the other hand, we stress the importance of "trusted third parties," and argue that these are going to be especially important in the context of parties that want to interact but do not trust each other. Again, if the third parties are selected by the end points, we see their presence as consistent with the end-to-end argument (or, as we have reframed it, the trust-to-trust argument).

Finally, we have posed a number of interesting design questions for application designers:

- Identify functional modules that might be usefully delegated or outsourced, and specify protocols that hook these together.
- Design these protocols so that the end node (the point where trust decisions are made) can keep control of the actual delegation.
- Design applications so that they can support several modes of communication, ranging from mutually open and trusting, to suspicious and bilaterally verified, or mediated by a trusted third party.

We have also highlighted the challenge of designing trustworthy end nodes.

Resilience: Building Better Users and Fair Trade Practices in Information

Andrea M. Matwyshyn*

* Andrea M. Matwyshyn is an assistant professor of Legal Studies and Business Ethics at the Wharton School at the University of Pennsylvania; more information about her work is available at http://www.amatwyshyn.com. She thanks Marcia Tiersky for comments on this work and can be reached at amatwysh@wharton.upenn.edu.

A long-running joke about the law asserts that that the practice of law would be more pleasant if it weren't for all those pesky clients. In the world of technology, a more terse version of this same sentiment exists: PEBKAC—Problem Exists Between Keyboard and Chair. Technologists often long for "better" users of their products. Naturally, the logical reaction to this type of statement is to encourage developers of products to engage in better usability testing of their products on actual consumers. However, a deeper question may lurk beneath the superficial flippancy of PEBKAC. Is there in fact a way that we can "build better users?" This Article argues that there is. Despite a long running discourse regarding the resilience of infrastructure and networks themselves, a portion of the discussion that has been neglected relates to human resilience—buttressing the resilience of *users* of technology and the role of law in furthering this goal. Borrowing lessons from developmental psychology and securities regulation, this Article expands the concept of resilience into the software and digital contracting ecosystem. It argues that technology law and policy can be tooled in part to adopt an explicit focus on building users' resilience and sense of self-efficacy, particularly in connection with data privacy and information security. Technology law and policy can help to train consumers to be confident users and bounce back from technology problems. With the assistance of strengthened fair trade practices in privacy, contract law offers one avenue for explicit trust-reinforcing mechanisms to assist consumers in becoming more resilient users.

I. WHAT IS RESILIENCE?

Many of us have found ourselves in a situation where we did not understand how a piece of software worked behind the scenes on our machines. We wondered what exactly we had agreed to when we clicked "yes" on the user agreement, whether we could really trust the code, and whether we understood the extent to which data would be collected about us. For some of us, a mild panic followed. Yet, in these moments of privacy "freakout," we had no one to ask. Reading a privacy policy—to the extent we understood it—likely yielded only more questions. We found ourselves cursing the software product as "creepy" privacy-invasive code. Meanwhile, the technologists who write software frequently feel equally frustrated by the way we, the consumer base that uses their products, interact with these products. In other words, a perception gap exists between the way that builders of technology tools perceive their products and the way that average consumers perceive these same products. The reason for this disconnect can be understood as a deficit of what developmental psychologists might call resilience. This resilience, or ability to recover and flourish in the face of obstacles, is frequently absent on both sides of the software equation—both in the code writing process

itself and in consumers' ability to overcome technology obstacles when using products.

A. Building Resilience in Systems: The Software Ecosystem

The concept of resilience has long been prevalent in systems literature. When applied to technological, human, and ecological systems, resilience refers to the ability of the system to restore and maintain itself in a functional state, providing all services, despite disruptive changes to the system.[1] As such, the concept of resilience springs from complexity theory and its focus on dynamic, emergent change and system evolution in response.[2] "The challenge [to a resilient system] . . . is to conserve the ability to adapt to change, to be able to respond in a flexible way to uncertainty and surprises" and "to identify the properties and processes that shape the future."[3] By definition, resilience involves the ability of a system to evolve in advance of and in response to known vulnerabilities to avoid or minimize their impact. However, this enterprise of anticipation is always limited by human knowledge and other factors.[4]

Resilient systems have been identified to possess three distinct types of properties or processes. First, the system is built with an eye to the future and possesses redundancy, which allows for bouncing back from destabilizing events to come.[5] In other words, they possess the ability to change. Second, the system demonstrates a shifting balance between stable and unstable forces, with internal controls intended to counterbalance external variability.[6] This means that the system is still capable of performing when an external force pushes on it. Third, the system demonstrates a dynamic, changing nature that compensates for vulnerability and persists.[7] In other words, the system possesses the ability to self-correct and return to a normal state.

The concept of resilience has been applied in legal literature to

---

1. *See* C. S. Holling & Lance H. Gunderson, *Resilience and Adaptive Cycles*, *in* PANARCHY: UNDERSTANDING TRANSFORMATIONS IN HUMAN AND NATURAL SYSTEMS 25, 28 (Lance H. Gunderson & C. S. Holling eds., 2002).

2. *See, e.g.*, Barbara A. Cherry, *Institutional Governance for Essential Industries Under Complexity: Providing Resilience Within the Rule of Law*, 17 COMMLAW CONSPECTUS 1, 4–5 (2009). Resilience can be "measured by the magnitude of disturbance that can be absorbed before the system changes its structure by changing the variables and processes that control behavior." Holling & Gunderson, *supra* note 1, at 28.

3. Holling & Gunderson, *supra* note 1, at 32.

4. *See generally* ROBERT ROSEN, LIFE ITSELF: A COMPREHENSIVE INQUIRY INTO THE NATURE, ORIGIN, AND FABRICATION OF LIFE 67–107 (T.F.H. Allen & David W. Roberts eds., 1991).

5. *See* Holling & Gunderson, *supra* note 1, at 32–33.

6. *Id.*

7. *Id.*

various types of connected systems, including the environmental ecosystem,[8] tribal sovereignty,[9] agencies and social trust,[10] human communities[11] (such as families[12]), social decay,[13] disasters,[14] markets and financial systems,[15] technology,[16] and critical infrastructure[17] (such as electrical grids[18] and internet infrastructure[19]). This idea of resilience analysis of the software development lifecycle and ecosystem, however, presents a newer undertaking, and one to date almost entirely unexplored in the legal literature.[20] The software ecosystem, including the processes of

---

8. As applied to ecological systems, "[r]esilience is the capacity of a system to absorb disturbance and reorganize while undergoing change so as to still retain essentially the same function, structure, identity, and feedbacks." Brian Walker et al., *Resilience, Adaptability and Transformability in Social-Ecological Systems*, ECOLOGY & SOC'Y, Dec. 2004, at 2.

9. *See, e.g.*, Patrice H. Kunesh, *Constant Governments: Tribal Resilience and Regeneration in Changing Times*, 19 KAN. J.L. & PUB. POL'Y 8 (2009).

10. *See, e.g.*, Rebecca M. Bratspies, *Regulatory Trust*, 51 ARIZ. L. REV. 575 (2009).

11. *See, e.g.*, Barbara Cosens, *Transboundary River Governance in the Face of Uncertainty: Resilience Theory and the Columbia River Treaty*, 30 J. LAND RESOURCES & ENVTL. L. 229 (2010). When applied to social systems, resilience is the "ability of human communities to withstand and recover from stresses." *Id.* at 237 (citing *Resilience Dictionary*, STOCKHOLM RESILIENCE CENTRE, http://www.stockholmresilience.org/research/whatisresilience/resiliencedictionary.4.aeea46911a3127427980004355.html (last visited Feb. 22, 2011)).

12. A dichotomy exists between the resilience-building and resilience-reducing potential of particular ecologies, such as families. Families can either assist in coping with change or hamper a child's ability to adapt. *See, e.g.*, Alastair Ager, *What Is Family? The Nature and Functions of Families in Times of Conflict*, *in* A WORLD TURNED UPSIDE DOWN: SOCIAL ECOLOGICAL APPROACHES TO CHILDREN IN WAR ZONES 39 (Neil Boothby et al. eds., 2006).

13. *See, e.g.*, Lawrence J. Vale & Thomas J. Campanella, *The Cities Rise Again*, *in* THE RESILIENT CITY: HOW MODERN CITIES RECOVER FROM DISASTER 3, 7 (Lawrence J. Vale & Thomas J. Campanella eds., 2005) (differentiating between "protracted socioeconomic decay" and disasters and noting that it is often more difficult for cities to respond with resilience to the former).

14. *See, e.g.*, W. Neil Adger et al., *Social-Ecological Resilience to Coastal Disasters*, 309 SCI. 1036 (2005).

15. *See, e.g.*, Interagency Paper on Sound Practices to Strengthen the Resilience of the U.S. Financial System, 68 Fed. Reg. 17,809 (Apr. 11, 2003), *available at* http://www.sec.gov/news/studies/34-47638.htm.

16. *See, e.g.*, Elizabeth Burleson, *Climate Change Consensus: Emerging International Law*, 34 WM. & MARY ENVTL. L. & POL'Y REV. 543, 584 (2010).

17. *See, e.g.*, Bennie G. Thompson, *A Legislative Prescription for Confronting 21st-Century Risks to the Homeland*, 47 HARV. J. ON LEGIS. 277, 298 (2010).

18. *See, e.g.*, Kelly A. Gable, *Cyber-Apocalypse Now: Securing the Internet Against Cyberterrorism and Using Universal Jurisdiction as a Deterrent*, 43 VAND. J. TRANSNAT'L L. 57 (2010).

19. *See, e.g.*, Gregory S. McNeal, *Cyber Embargo: Countering the Internet Jihad*, 39 CASE W. RES. J. INT'L L. 789, 802 (2007).

20. Although legal literature hasn't explored resiliency analysis, computer science has. These basic tenets are: protection from disclosure (confidentiality); protection from alteration (integrity); protection from destruction (availability); who is making the request

software development, deployment, and repair, should be viewed as another type of system that warrants a resilience analysis. Why? The reason for this extension is the avoidability of much consumer harm, particularly with respect to privacy and information security concerns. A significant portion of consumer complaints arise because particular digital products cannot withstand the entirely foreseeable wear and tear of consumer use and foreseeable third party attacks.

What users perceive to be unacceptable, privacy-invasive code frequently surprises even sophisticated companies. However, with adequate resiliency analysis beforehand, most consumer privacy and information security freakouts are entirely avoidable. Two recent examples of this underestimation of consumer reactions involve Google and Facebook. In early 2010, Google launched a product called Buzz.[21] By external appearances, Buzz seemed to be a type of crossover product between a Facebook-like interface and a Twitter feed. To assist in its adoption, Google decided to repurpose the data in users' Gmail e-mail account contact lists for their individual starter group of "followers" in Buzz, making these lists public by default.[22] Almost immediately, public outcry ensued.[23] Gmail address books for some users contained contact information for individuals who were unwelcome "followers."[24] In its zealousness to promote Buzz, Google had, according to press accounts, cut short its usual beta testing process and unintentionally triggered the "privacy invasion" sensitivity of some of its users.[25] This product shipping decision was subsequently labeled by a Federal Trade Commission (FTC)

(authentication); what rights and privileges the requestor has (authorization); the ability to build historical evidence (auditing); and the management of configuration, sessions, and exceptions. *See, e.g.*, OFFICIAL (ISC)^2 GUIDE TO THE CISSP CBK (Harold F. Tipton & Kevin Henry eds., 2007); Kristin R. Eschenfelder & Anuj C. Desai, *Software as Protest: The Unexpected Resiliency of U.S.-Based DeCSS Posting and Linking*, 20 INFO. SOC'Y 101 (2004) (demonstrating the proliferation of U.S.-based websites either posting or linking to the DeCSS program over the course of *Universal City Studios v. Corley*, 273 F.3d 429 (2d Cir. 2001)).

21. *Google Buzz—What Is the Purpose?*, CLEAN CUT MEDIA (Feb. 16, 2010), http://www.cleancutmedia.com/internet/google-buzz-what-is-the-purpose; GOOGLE BUZZ, www.google.com/buzz (last visited Feb. 22, 2011).

22. Jonathan Fildes, *Google Admits Buzz Social Network Testing Flaws*, BBC NEWS, http://news.bbc.co.uk/2/hi/technology/8517613.stm (last visited Feb. 22, 2011).

23. *See id.*

24. In one case, an abusive ex-husband was added as a follower to one woman's Buzz feed, much to her dismay. Nick Saint, *Outraged Blogger Is Automatically Being Followed by Her Abusive Ex-Husband on Google Buzz*, BUS. INSIDER (Feb. 12, 2010), http://www.businessinsider.com/outraged-blogger-is-automatically-being-followed-by-her-abusive-ex-husband-on-google-buzz-2010-2.

25. Jonathan Fildes, *Google Admits Buzz Social Network Testing Flaws*, BBC NEWS (Feb. 16, 2010), http://news.bbc.co.uk/2/hi/technology/8517613.stm.

member as "irresponsible conduct"[26] and at least eleven U.S. lawmakers called for an FTC investigation.[27] Along similar lines, Facebook found itself in court because of its Beacon program,[28] which collected data regarding user behaviors on "partner" websites.[29] The Beacon program involved embedded code in partner sites that triggered a post regarding consumer conduct on those partner sites to be posted to some consumers' Facebook feeds.[30] Because some users did not understand how this information was being shared, and they considered the practice an invasion of their privacy.[31] This confusion resulted in what the media has termed a "public relations disaster"[32] and in a class action lawsuit against Facebook that resulted in a settlement in the amount of $9.5 million.[33]

Both companies in question were surprised by the consumer reaction. However, in both cases this surprise was likely avoidable. More extensive usability testing on average consumers likely would have revealed the code's lack of resilience when embedded into the broader software ecosystem.

That said, the lack of resilience of the developers' code in the two cases above was only part of the problem. It was undoubtedly exacerbated by some users' lack of individual resilience. Some consumers poorly adjust to new technology and experience potent emotions of stress and confusion with respect to even small changes in existing software. To understand this parallel consumer resilience side of this dynamic, we now turn to developmental psychology.

---

26. Emily Steel, *Google Buzz Exemplifies Privacy Problems, FTC Commissioner Says*, WALL ST. J. BLOGS (Mar. 17, 2010, 2:37 PM), http://blogs.wsj.com/digits/2010/03/17/google-buzz-exemplifies-privacy-problems-ftc-commissioner-says/.

27. Grant Gross, *Lawmakers Ask for FTC Investigation of Google Buzz*, PCWORLD.COM (Mar. 29, 2010), http://www.pcworld.com/article/192801/lawmakers_ask_for_ftc_investigation_of_google_buzz.html?tk=rss_news.

28. Lane v. Facebook, Inc., No. C 08-3845 RS, 2009 WL 3458198 (N.D. Cal. Oct. 23, 2009).

29. Juan Carlos Perez, *Facebook's Beacon More Intrusive than Previously Thought*, PCWORLD.COM (Nov. 30, 2007), http://www.pcworld.com/article/140182/facebooks_beacon_more_intrusive_than_previously_thought.html.

30. *See* Om Malik, *Is Facebook Beacon a Privacy Nightmare?*, GIGAOM (Nov. 6, 2007), http://gigaom.com/2007/11/06/facebook-beacon-privacy-issues/. For a discussion of the public relations problems for Facebook caused by the "Beacon" technology, see Perez, *supra* note 29.

31. *See, e.g.*, Gil Kaufman, *Facebook Bows to User Complaints About Beacon's Privacy Violation*, MTV.COM (Nov. 30, 2007), http://www.mtv.com/news/articles/1575455/facebook-bows-user-complaints.jhtml.

32. Caroline McCarthy, *Facebook Notifies Members About Beacon Settlement*, CNET NEWS (Dec. 3, 2009), http://news.cnet.com/8301-13577_3-10409034-36.html.

33. *Id.*

B. Building Resilience in Users

In developmental psychology literature, resilience of humans refers to the process through which a person is exposed to adversity and manages to adapt and function successfully despite setbacks.[34] Many factors contribute to the development of resilience, and the process is inherently socially embedded. This means that the resilience of the community and other contexts that the individual experiences can either assist or diminish resilience in the individual. Further, resilience can be learned,[35] and individuals functioning under conditions of stress can indeed rise to the occasion, overcoming challenges and succeeding.[36] Although methodological variation exists, generally resilience studies look for "risk" factors[37] and mitigating "protective" factors that assist with overcoming stressors.[38] In particular, the extent to which individuals participate in decision making tends to correlate positively with improved resilience.[39] What this means for the software ecosystem is that designing products with greater transparency and user participation in mind will likely yield more resilient users over time.

In other words, "building better users" entails, first and foremost,

---

34. *See, e.g.*, Corey L. M. Keyes, *Risk and Resilience in Human Development: An Introduction*, 1 RES. HUM. DEV. 223, 224 (2004), http://www.sociology.emory.edu/ckeyes/rhd14_1.pdf.

35. The American Psychological Association identified four factors in particular shared by individuals who tended to be viewed as "resilient": a) "the capacity to make realistic plans and to carry them out," b) a positive self-image and confidence in one's strengths and abilities, c) the ability to communicate skillfully and solve problems, and d) "the capacity to manage strong feelings and impulses." AM. PSYCHOLOGICAL ASS'N, *Resilience: After a Hurricane*, APA.ORG, http://www.apa.org/helpcenter/hurricane-resilience.aspx/ (last visited Feb. 22, 2011).

36. For example, in the words of one researcher, "'[t]here are kids in families from very adverse situations who really do beautifully, and seem to rise to the top of their potential, even with everything else working against them.'" David Gelman, *The Miracle of Resiliency*, NEWSWEEK, Summer 1991, at 44 (quoting Dr. W. Thomas Boyce, Director of Behavioral and Developmental Pediatrics at the University of California, San Francisco).

37. "'A risk factor is an individual attribute, individual characteristic, situational condition, or environmental context that increases the probability' of an undesirable outcome." Laura Greenberg, *Compensating the Lead Poisoned Child: Proposals for Mitigating Discriminatory Damage Awards*, 28 B.C. ENVTL. AFF. L. REV. 429, 455 (2001) (quoting Howard B. Kaplan, *Toward an Understanding of Resilience: A Critical Review of Definitions and Models*, *in* RESILIENCE AND DEVELOPMENT: POSITIVE LIFE ADAPTATIONS 17, 37 (Meyer D. Glantz & Jeannette L. Johnson eds., 1999)).

38. *See, e.g.*, Michael Rutter, *Psychosocial Resilience and Protective Mechanisms*, *in* RISK AND PROTECTIVE FACTORS IN THE DEVELOPMENT OF PSYCHOPATHOLOGY 181, 181 (Jon Rolf et al. eds., 1993).

39. *See, e.g.*, The Consortium on the School-Based Promotion of Social Competence, *The School-Based Promotion of Social Competence: Theory, Research, Practice, and Policy*, *in* STRESS, RISK, AND RESILIENCE IN CHILDREN AND ADOLESCENTS: PROCESSES, MECHANISMS, AND INTERVENTIONS 268 (Robert J. Haggerty et al. eds., 1994).

convincing consumers that they can master a technology before them and guiding them in doing so. As such, the development of resilience in humans is inherently bound up with the concept of self-efficacy, which refers to an individual's beliefs about his control and ability to successfully perform a given task or behavior.[40] Empirical evidence offers support for the connection between self-efficacy perceptions and resilience; there tends to be a correlation in many contexts, such as in academic performance, between the strength of an individual's beliefs about the capability of success and actual success.[41] Even when controlling for ability levels in the specific task, some research demonstrates that students who do not believe they can achieve a goal are, in fact, less likely to do so than their peers who do believe they can achieve that goal.[42] Unlike the concept of self-esteem, self-efficacy pertains to narrow, specific, and concrete goals and varies within humans from task to task. No one is good at everything. I may be a good photographer, but my tennis abilities leave much to be desired; for another person the two tasks' success levels may be reversed.

The leading theory on self-efficacy is found in the work of Albert Bandura. According to Bandura, when individuals select which tasks to undertake and decide whether to persevere "in the face of obstacles or aversive experiences," they do so based on their perceptions of self-efficacy.[43] People develop self-efficacy for a specific task, such as mastering a new technology product, in four ways:[44]

1. Through personal experience;
2. From physiological and/or emotional reactions to an event;
3. Through vicarious experiences or modeling;
4. From feedback from their social environment.

Through these mechanisms, people either adopt a resilient approach to obstacles, mustering feelings of self-efficacy to learn and work through

---

40. For a discussion of self-efficacy see, for example, ALBERT BANDURA, SELF-EFFICACY: THE EXERCISE OF CONTROL (1997).

41. *See* Barry J. Zimmerman, *A Social Cognitive View of Self-Regulated Academic Learning*, 81 J. EDUC. PSYCHOL. 329, 331 (1989) (offering data suggesting that perceptions of high self-efficacy are positively correlated with persistence and achievement in an academic context).

42. *Id.*

43. Albert Bandura, *Self-Referent Thought: A Developmental Analysis of Self-Efficacy*, *in* SOCIAL COGNITIVE DEVELOPMENT: FRONTIERS AND POSSIBLE FUTURES 200, 201 (John H. Flavell & Lee Ross eds., 1981); *see infra* pp. 9–11; *see also, e.g.*, Albert Bandura, *Social Cognitive Theory of Self-Regulation*, 50 ORGANIZATIONAL BEHAV. & HUM. DECISION PROCESSES 248, 257–58 (1991), http://www.des.emory.edu/mfp/Bandura1991OBHDP.pdf (concluding that confidence in self-efficacy positively influences choices, aspirations, effort, perseverance, and stress levels).

44. For a discussion of self-efficacy determinants, see, for example, Michael Hunter Schwartz, *Teaching Law Students to Be Self-Regulated Learners*, 2003 MICH. ST. L. REV. 447, 456 (2003).

new obstacles or they fail to persevere.[45]

Personal experience plays an important cumulative role in learning resilience. A user's history of technology learning is likely to impact self-efficacy in new technology tasks; it brings a backdrop of success or failure to all new technology situations users enter. For Bandura, "partial mastery experiences" predict "subsequent performance of threatening tasks that [an individual has] never done before."[46] Perhaps more dramatically, "[a]rbitrarily instilled beliefs of inefficacy discourage . . . coping behavior even when the opportunity to exercise personal control exists. In contrast, instilled perceived efficacy largely overrides ostensible external constraints on the exercise of personal control . . . ."[47] In other words, when it comes to technology, peoples' negative prior experiences with code prime their future experiences. Stated another way, cumulative learning episodes can create either a virtual circle of self-reinforcing technology success or a vicious circle of self-priming technology failure.

In a similar vein, as in all things human, emotion plays a role in learning and control. Some consumers reach a point in their interactions with technology where they become overwhelmed with frustration and a feeling of lack of control; they have a negative emotional reaction to code they cannot seem to understand and simply give up on learning more. Research in self-efficacy theory indicates a possible relationship between anxiety of this sort and low self-efficacy.[48] In other words, when consumers experience anger or stress over malfunctioning software, their sense of self-efficacy likely diminishes. People who value a goal but develop low self-efficacy with respect to their ability to achieve it, in turn, can become despondent, depressed, and disengaged. Then, viewing the disengagement as failure, they feel powerless in achieving the goal, creating a self-reinforcing negative cycle. As a consequence, they may shy away from another attempt to master the task.[49] This negative dynamic then further diminishes the likelihood of success with a particular task.

Third, self-efficacy can be bolstered by observational learning from

---

45. *See, e.g.*, Zimmerman, *supra* note 41, at 331 (offering data suggesting that perceptions of high self-efficacy are positively correlated with persistence and achievement in an educational context).

46. Albert Bandura, *Self-Efficacy Mechanism in Human Agency*, 37 AM. PSYCHOLOGIST 122, 128 (1982). Thus, "[e]nactive attainments provide the most influential source of efficacy information because [they] can be based on authentic mastery experiences[;] [s]uccesses heighten perceived self-efficacy[,] repeated failures lower it . . . ." *Id.* at 126 (emphasis omitted).

47. BANDURA, *supra* note 40, at 268.

48. *See generally* S. Lloyd Williams, *Self-Efficacy, Anxiety, and Phobic Disorders*, *in* SELF-EFFICACY, ADAPTATION, AND ADJUSTMENT: THEORY, RESEARCH, AND APPLICATION 69 (James E. Maddux ed., 1995).

49. *Id.*

the groups around the person, modeling on the behaviors of "similar others."[50] In other words, people who work or live in environments with people who demonstrate strong computer skills and efficacy with code are probably more likely to develop strong technology skills themselves. "Seeing similar others perform successfully can raise efficacy expectations in observers who then judge that they too possess the capabilities to master comparable activities."[51] Modeling has three major effects.[52] First, it teaches a learner to acquire and perform new responses or skills from observation.[53] Second, it serves to inhibit fear responses because the learner sees that the model does not suffer negative consequences. To the contrary, the learner is potentially emboldened when she sees that such behavior often results in positive consequences.[54] Third, a "facilitation of responses" happens because the learner can emulate the model's cues.[55] Seeing someone similar engage in a behavior leads a learner to believe that he or she has the ability to engage in the same conduct. Social models demonstrate what is possible, thereby changing what the learner believes she too can accomplish—an instilling of feelings of self-efficacy. In other words, technology modeling and technology mentorship helps consumers learn to help themselves.

Learning self-efficacy and, in turn, becoming resilient are cumulative, meaning that episodes of success and failure and environmental inputs blend to evolve an individual's beliefs of self-efficacy.[56] We construct a belief in our ability to succeed with increasingly challenging tasks based on our ability—and by observing others' ability—to finish similar but less difficult tasks.[57] In essence, this is a form of human self-regulation, which,

---

50. *See* Bandura, *supra* note 46, at 126–27.

51. *Id.*

52. *See* GERALD COREY, THEORY AND PRACTICE OF COUNSELING AND PSYCHOTHERAPY 293–94 (5th ed. 1996) (relying on Bandura's research).

53. *Id.* at 293.

54. *Id.* at 294.

55. *Id.*

56. *See* Bandura, *supra* note 46, at 124. According to Bandura, models can serve to instruct, motivate, disinhibit, inhibit, socially facilitate, and arouse emotion in a process of vicarious reinforcement. *See id.* at 126–27. Essentially, development is viewed as a process of quantitative change, during which learning episodes gradually accumulate over time. *See id.* Although Social Learning Theory does not directly address historical or cultural context, it reflects the tradition of Vygotsky and the contextualist approach by recognizing the dialectical process of a person who is working within and shaped by an environment; a triadic reciprocal determinism occurs among behavior, cognitive factors, and the environment. *See* LIONEL NICHOLAS, INTRODUCTION TO PSYCHOLOGY 136–38 (2009). There is no endpoint to development, and universal behaviors are rare. Thus, children are developmentally malleable but only within constraints of biology and environment, an environment replete with technology. *See id.*

57. *See id.* at 128.

Bandura argues, is contingent on learning.

Finally, feedback loops matter. Learning, argues Bandera, requires extensive feedback loops to correct for problems that ensue from individual interpretations of situations. These feedback loops are necessarily social: to extend the cognitive capabilities of the individual through tools and resources, learners need inputs for correction of misguided conduct. In order for even those who are "good self-regulators" "[t]o enhance their competency, they have to figure out what information they lack, how best to frame their inquiry, from whom to seek assistance, and how to overrule any social hesitancy they feel to do so."[58] This is where law can enter the conversation and offer additional feedback loops.

II. RESILIENCE, CONTRACTS, AND FAIR TRADE PRACTICES IN INFORMATION

As I have argued elsewhere, successfully regulating technology means a primary focus on regulating the humans building and interacting with the technology, rather than the products themselves.[59] Technology specific regulation is doomed to failure as the pace of innovation outstrips the law. Human conduct, on the other hand, particularly when framed in terms of traditional legal approaches, is a finite and regulable universe of possibilities. If we stipulate that both innovation in code and consumer protection are equally important social goals, we can reframe the conversation around regulating conduct of both sets of humans involved in the code ecosystem in their relation to each other—both the humans who *write* the code and the humans who *use* the code. The discussion in Section I above articulated that resilience in systems is characterized by redundancy, a shifting balance between stable and unstable forces limited by internal controls, and a dynamic nature that compensates for change and then persists. The above discussion of the developmental psychology literature leads us to the conclusion that four core elements—experience, emotion, modeling, and feedback loops—are integral to building resilience in consumers. When we consider these four core elements, we can begin to construct a user-centered model for consumer protection in technology spaces.[60] Legal approaches, therefore, should focus on enhancing resilience

58. BANDURA, *supra* note 40, at 231.

59. Andrea M. Matwyshyn, *Technology, Commerce, Development, Identity*, 8 MINN. J. L. SCI. & TECH. 515 (2007).

60. Though we frequently anthropomorphize it, technology does not really have a life of its own at present. It is a creation by humans for humans; humans give technology its animating features. Even emergent unintended technology consequences are, nevertheless, at some point caused in fact by humans authoring code and, potentially, proximately caused by other humans interacting with that code. But, first and foremost, the reason that anyone writes or uses code is developmental—code authorship or use is a type of act of creative

on both sides of commercial relationships between the imperfect humans creating technology and the imperfect humans using technology. The natural starting point for such a legal undertaking is contract law.

A. Resilience and Contracts in Technology-Mediated Spaces

The primary law of the code ecosystem since its inception has always been contract law. Despite a greater volume of litigation with respect to high profile intellectual property in technology spaces, ultimately, contract law is currently a more potent framework for legal ordering than is intellectual property law in such spaces. But for very limited circumstances, contract law is not preempted even by copyright law when an agreement exists between the parties.[61] As *ProCD, Inc. v. Zeidenberg* explained,[62] where a contract between the parties exists, regardless of whether the subject matter is copyrightable, contract law is not preempted.[63]

---

expression that intends to expand the capabilities of the author or user with a technological appendage to his or her being. Though perhaps this reflects a melodramatic framing of the deeper social meaning of, for example, a flying pig screensaver, even the creation of this code with arguably limited social impact still reflects an act of self-realization for the coder. It reflects an act of human generativity. Generativity—a developmental psychology concept arising from the work of Erik Erikson—refers to the human desire to create something greater than yourself that survives your own lifetime. *See* ERIK H. ERIKSON, CHILDHOOD AND SOCIETY 231 (1950). Professor Zittrain has eloquently argued that devices and code are inherently generative. *See* JONATHAN ZITTRAIN, THE FUTURE OF THE INTERNET AND HOW TO STOP IT 69–70 (2008). I respectfully submit they are not truly generative in the traditional meaning of the term. Driven by the current limitations of artificial intelligence research, only humans can be generative at present—code is merely a line of symbols in the absence of a human to author it, animate it with values, or give it derivative life.

61. In *Rano v. Sipa Press, Inc.*, the 9th Circuit held that copyright preempted state law relating to the termination at will of a license with an indefinite duration because when "California law and federal law are in direct conflict, federal law must control." Rano v. Sipa Press, Inc., 987 F.2d 580, 585 (9th Cir. 1993). Assignability of a licensee's rights would provide another preemption basis because under federal law such rights cannot be assigned in a nonexclusive license without the consent of the licensor. *See* CFLC, Inc. v. Cadtrak Corp., 89 F.3d 673, 679 (9th Cir. 1996). *Cf.* Chamberlain v. Cocola Assocs., 958 F.2d 282, 285 (9th Cir. 1992) (applying a California statute regarding transfer of a tangible object in the case of a transfer of the intangible rights to use an object).

62. 86 F.3d 1447 (7th Cir. 1996). *ProCD, Inc. v. Zeidenberg* was the first appellate ruling dealing with the enforceability of shrinkwrap licenses, and it held that the contract restrictions ProCD placed on the use of a noncopyrightable database were not preempted by copyright law. *See id* at 1454–55; *see also* DaimlerChrysler Servs. N. Am., LLC v. Summit Nat'l, Inc., 144 F. App'x. 542 (6th Cir. 2005) (holding that copyright defenses are irrelevant to contract enforcement); Davidson & Assocs. v. Jung, 422 F.3d 630, 632, 639 (8th Cir. 2005) (holding that a license is not preempted by fair use); Altera Corp. v. Clear Logic, Inc., 424 F.3d 1079, 1089–90 (9th Cir. 2005) (finding that copyright law does not preempt contract enforcement); Bowers v. Baystate Techs, Inc., 320 F.3d 1317, 1323–26 (Fed. Cir. 2003) (holding that copyright law did not preempt the plaintiff's contractual claims).

63. The court opined that

> enforcement . . . would not withdraw any information from the public domain. . . .

More importantly, however, contract law is critical because it is arguably the field of law most aimed at fostering resilience in the marketplace: it is intended to create a safety net of commercial trust and to assist parties in bouncing back from relationship failures. At its most basic level, contract law involves one set of imperfect persons successfully interacting with another set of imperfect persons to generate a sense of control over the exchange.

A concern for the four core elements of developing resilience in humans are also represented in contract law—the same four elements I have argued should be fostered in users of technology. For example, contract law reflects a concern over imbalanced cumulative learning between the parties in its disparate treatment of sophisticated contracting parties and unsophisticated parties in the Uniform Commercial Code.[64] Emotions of bargaining parties are considered through doctrines such as duress and coercion, where one party can exert psychological influence unfairly over another. Modeling issues arise, rather obviously, in the perennial debate over form contracting. Companies frequently use industry-wide contracts, and their lawyers "borrow" forms from each other or reuse the same form contract with numerous clients. Finally, contract is heavily driven through crafting feedback loops though various doctrines related to breach, remedies, and warranties on a going forward basis, seeking to preserve the relationship whenever possible.

So, is the resilience problem in information contracting solved because of the resilience of contract law itself? No. The existing inherent resilience-fostering nature of contract law is being undercut in new technology contexts, particularly with respect to privacy and information security. Due to certain unique characteristics, rather than bolstering both systemic and individual resilience, technology-mediated contracting instead damages resilience on both sides of the relationship between the code creator and the consumer user. In previous work, I empirically demonstrated that terms of use and end user license agreements online—the contracts that shift risk from the authors of code to users—were becoming progressively more draconian in favor of drafters.[65] I argued that the results indicate that current Internet contracting constructions do not

---

> Everyone remains free to copy and disseminate all 3,000 telephone books that have been incorporated into ProCD's database. Anyone can add SIC codes and zip codes. ProCD's rivals have done so. Enforcement of the shrinkwrap license may even make information more readily available, by reducing the price ProCD charges to consumer buyers.

ProCD, Inc., 86 F.3d at 1455.

64. *See, e.g.*, U.C.C. § 2-207(2) (treating merchants differently).

66. Andrea M. Matwyshyn, *Mutually Assured Protection: Toward Development of Relational Internet Data Security and Privacy Contracting Norms*, *in* SECURING PRIVACY IN THE INTERNET AGE 73 (Anupam Chander et al. eds., 2008).

successfully reconcile the needs of code creators and consumers in a way that is likely to lead to improved trust and growth in the digital marketplace.[66] My predictions in that work appear to have been correct, at least with respect to privacy and information security.

How do we assess the legal implications of this dynamic? Although it is tempting to simply argue in favor of technology contract essentialism, technology-mediated contracts are not really special contracts; instead, they should be analyzed as contracts executed under special circumstances that diminish party resilience, particularly when a bargaining power imbalance already exists. The next question, therefore, is how can we shift the dynamics of technology-mediated contracting back in favor of fostering resilience? As a thought exercise, using the four core elements of building resilience identified previously, let us analyze four common consumer laments regarding understanding data privacy and information security and its relationship to the traditional resilience of contract law. This in turn may help identify a set of guidelines for "fair trade practices" in information that bolsters resilience. Such guidelines, if authored by the FTC, would provide meaningful guidance for code creators on avoiding an unfair trade practices inquiry from the FTC with respect to data privacy and information security practices.

B. Fair Trade Practices, Privacy, and Technology Contracts

As the examples of Google Buzz and Facebook Beacon demonstrated, consumer privacy freakouts can be swift and brutal. Why? As the FTC has correctly identified, the core deficit for consumers is a missing sense of control.[67] This feeling of lack of control and, correspondingly, diminished resilience, is driven by two dynamics: weakened communication and

---

66. *Id.* In particular, my sample did not reflect

> a balance being struck between predictable mitigation of liability for content providers and assumption of obligations to securely treat user data. Instead, the content of the terms of use and privacy policies analyzed reflected an inherently irreconcilable tension in legal strategy adopted in the two constructions: the terms of use tended to reflect a nonrelational approach best suited to a one-shot game of adversaries, while the privacy policies tended to reflect a more relational approach with a continuing obligation to maintain data in accordance with security promises, reflecting an iterated game of commercial partners.

Id. at 81. Another developing tension that was noted was one of contractual interpretation. Browsewrap terms of use are usually not deemed enforceable, but privacy policies in the same browsewrap construction are being enforced by the FTC and private actors as contracts (at least in legal approach, if not explicitly). *See id.* at 77–80.

67. FED. TRADE COMM'N, PROTECTING CONSUMER PRIVACY IN AN ERA OF RAPID CHANGE: A PROPOSED FRAMEWORK FOR BUSINESSES AND POLICYMAKERS (2010), http://www.ftc.gov/os/2010/12/101201privacyreport.pdf; *see also* Press Release, Fed. Trade Comm'n, FTC Staff Issues Privacy Report, Offers Framework for Consumers, Businesses, and Policymakers (Dec. 1, 2010), *available at* http://www.ftc.gov/opa/2010/12/privacyreport.shtm.

bolstered data mining. Perhaps counterintuitively, technology mediated spaces present an impoverished contracting medium when compared to real space. Second, technology-mediated spaces involve far greater data collection medium capabilities when compared with real space; this collection may frequently exceed the scope of information a consumer believes herself to be knowingly volunteering. In order to restore consumers' sense of control and foster resilience, a feedback loop can be implemented by the FTC through articulating additional fair trade practices in information. Such a set of guidelines might include four elements: a single, plain English user agreement that embodies all relevant terms, a summary label, contractual enforcement using a "reasonable digital consumer" standard, and a transparency requirement to reasonably answer all consumer privacy and security inquiries.

1. Experience in Digital Contract: Creating a Plain English "Information License and Security Agreement"

The experience of an average consumer with respect to digital contracting today goes something like this: "I tried to read a EULA once, it was really long and I couldn't understand anything in it. It included references to a bunch of other agreements too. I gave up and just clicked 'yes' because I needed to use the product. Now I just click 'yes' on every contract that pops up. Besides, although I care about privacy, all these companies are just going to follow me around and abuse my data anyway. There's no point to even reading a privacy policy."

Fatalistic default acceptance of terms presented to consumers is the norm in digital contracting. Even consumers who wish to invest the time to understand the contract before them are unlikely to be able to do so. Coupled with the inability to ask questions, this dynamic leaves consumers feeling helpless, without meaningful control and choice, and clicking "yes" on every agreement that appears before them on a screen in a Pavlovian clicking behavior.

Particularly because of the difficulty in understanding companies' data privacy and security practices, consumers require a single point of information regarding companies' practices. As I have argued elsewhere,[68] even assuming for the sake of argument that a consumer can understand the plain face meaning of the terms of the contract, the consumer cannot necessarily verify what particular code is in fact doing on her system. Code can hide itself and its functionalities in elaborate ways. Without full clear disclosure to eliminate this information imbalance, a fair meeting of the

68. *See* Andrea M. Matwyshyn, *Hidden Engines of Destruction: The Reasonable Expectation of Code Safety and the Duty to Warn in Digital Products*, 62 FLA. L. REV. 109, 112–13 (2010).

minds on information collection terms in a contract is not feasible.

Terms of use agreements and privacy policies as separate agreements is merely an unfortunate artifact of early Internet law.[69] These two antiquated contract constructions should be replaced with a single contract form where privacy and security promises are conspicuous, material terms of the user license agreement. Consequently, a breach of these terms obviously presents a material breach of the agreement in its entirety and offers a consumer recourse in the relationship for contract breach.[70]

In addition to creating a single contract, a "Plain English" requirement in digital contract language would greatly assist consumers' sense of control over information exchanges and, consequently, foster resiliency. Plain English requirements have been instituted in situations where a disclosure need was pressing, but the subject matter at hand was inherently complicated. For example, in securities regulation, the Securities and

---

69. *See, e.g.*, Matwyshyn, *supra* note 66, at 77–79.

70. In fact, a business benefit may also arise for the code creator from such a simplified construction. Particularly in large organizations, it is common to find a "lack of cooperation among attorneys, businesspeople, and technologists . . . . The lawyers drafting terms of use may be inadequately sensitive" to the technology in question. *Id.* at 83. "Meanwhile, privacy policies are sometimes written by marketing departments or technologists who may be unaware of the legal implications of particular contract presentation on the user interface[,]" for example. *Id.* Therefore, when these two contracts are analyzed together, they may, at present, not effectively accomplish either liability limitation or user disclosure in their current dual presentation. *Id.* For example, in Internet contracting contexts, terms of use are generally written by attorneys who zealously attempt to limit their clients' liability to the greatest extent possible but may not really understand the website. However, because no negotiation of these terms occurs, they remain in their original, unnegotiated format when the website goes live. These terms of use, meanwhile, are considered unsightly legal verbiage by the designers of websites and are tucked away in inconspicuous places. The effect of these actions on legal enforceability generally goes uncontemplated: the lawyers have been excluded from the business decision loop. Privacy policies, on the other hand, are generally written at least in part by the public relations department of business enterprises. As such, the legally binding effect of these privacy promises is frequently not understood by the businesspeople involved in their creation. Thus, terms of use and privacy policies are not necessarily thought about as being inherently interrelated by businesspeople and attorneys. The standard content of terms of use, such as user indemnification provisions, may be set aside by some U.S. courts. In the United States, challenges could be brought on the basis of substantive unconscionability (for example, user indemnification provisions), embodying offline problems of form contracts of adhesion, procedural unconscionability with regard to formation uncertainty, as well as other formation issues arising from inadequate user notice and consent and the absence of negotiation. *See id.* at 80–81, 83. Most terms of use would almost certainly be set aside in their entirety or at least in substantial part if challenged in the European Union. The European Union's grounds for invalidation of terms-of-use content include violation of, among other directives, the European Union Directive on Distance Contracts and the Directive on Unfair Terms. *See, e.g.*, James R. Maxeiner, *Standard-Terms Contracting in the Global Electronic Age: European Alternatives*, 28 YALE J. INT'L L. 109, 111–13 (2003). Clearly, multijurisdictional unenforceability of terms of use is a suboptimal outcome from the perspective of both technologists and lawyers within an entity attempting to limit liability on a global basis.

Exchange Commission (SEC) promulgated a Plain English Rule with respect to prospectuses.[71] The SEC believed that "using plain English . . . will lead to a better informed [ ] market . . . in which" consumers "can more easily understand . . . disclosure . . . ."[72] Parallel improvements should happen in the data privacy and information security contracting context.

2. Emotion in Digital Contract: Creating a Sense of Transparency in Formation with Summary Labeling

It is not uncommon to hear a consumer say with frustration: "I didn't even see that there were terms of use linked on the bottom of that website. How was I supposed to know I was bound by them? And what are all these links to other contracts? I can't possibly read forty screens of ten-point font on a slow-loading smartphone."

Although obscure presentations of terms without an affirmative act of assent are unlikely to be enforced, these same terms, if merely incorporated by reference in another more obvious set of terms, are likely to be deemed enforceable. The task of reading multiple cross-referenced linked documents, potentially on a small mobile device, is limiting, at best. At worst, it is taking advantage of a crippled user interface. In order for consumers to understand the totality of the terms to which they are bound, a potentially promising transparency approach is mandating a one-page summary of all material terms—modeled on the spirit of a summary prospectus—as the first screen of all digital agreements. In the language of the SEC, the rationale behind the requirement of a summary prospectus is to offer concise standardized information to consumers, which allows them to compare terms across products.[73] The information market requires similar disclosure structures to build consumer resilience.

3. Modeling: Imposing Digital Reasonableness Standards

Particularly with respect to privacy settings on social network websites such as Facebook, a common consumer lament is: "There are way too many privacy settings, and they change the presentation constantly. I can't keep up, and I have no clue whether what I'm doing will actually set the preferences the way I want them to be. No average person can figure this out in a reasonable amount of time."

If the ability to set privacy settings is offered, these settings—as selected by the consumer—should constitute a material term of the agreement. Correspondingly, a material unilateral alteration of the terms

---

71. SEC Plain English Disclosure, 17 C.F.R. pts. 228–30, 239, 274 (2008).

72. *Id.*

73. SEC Enhanced Disclosure and New Prospectus Delivery Options for Registered Open-End Management Investment Companies, 17 C.F.R. pts 230, 232, 239, 274 (2008).

may constitute a breach of contract. Any alteration in interface that changes the spirit of consumer preferences will be perceived by a consumer as "unfair:" technology-mediated contracting lacks the back-and-forth consumers take for granted in real space. Although most consumers never negotiate the agreements they sign, the potential for negotiability appears to exist, at least superficially, in most cases. A human hands over a document for signature; presumably this human can engage in some degree of negotiation or at least answer questions about the contract. Although this person's incentives are not aligned with those of the consumer, psychologically, for a consumer, this person serves as a type of model with respect to the relationship. This, in turn, likely fosters feelings of self-efficacy and control. In digital spaces, no other human appears present, and this modeling aspect of the exchange is lost.

As I have argued elsewhere, technological skills vary dramatically across users, and this distribution is multi-modal, not necessarily "map[ping] onto chronological age."[74] As such, the imposition of a reasonableness standard for contracts in technology spaces accommodates this variation. Creating contracts that a reasonable consumer—as determined by empirical testing—can understand has a type of modeling function. The imposition of this "reasonable digital consumer" standard would perform a modeling function for consumers less skilled than average, urging them to improve and offering a target for their development.

4. Feedback Loops in Digital Contract: Offering a Live Human to Negotiate and Explain Terms

A final lament of many a user goes something like this: "None of my friends understand any of this stuff, either. I don't have anyone to ask for help with understanding a EULA or privacy policy, or anyone to ask questions of regarding what the company is doing with my information." In other words, consumers lack a feedback loop: they are asking social guidance in interpreting the situation. Companies rarely have a real-time virtual point of contact for inquiries about EULAs and privacy policies. However, they frequently have real-time shopping assistance. In other words, the possibility exists for the drafter to provide real-time feedback on contracts in technology-mediated spaces. However, even without real-time assistance, consumer questions regarding data privacy and information security, particularly subsequent to a known data breach, should be promptly answered through other means. Based on this author's experience, consumer inquiries regarding privacy and security inquiries are sometimes ignored even by large, reputable companies.

---

74. Matwyshyn, *supra* note 59, at 540.

III. CONCLUSION

Applying the concept of resilience, this Article has explored the possibility of crafting improved guidelines for fair trade practices in information contacts. Without meaningful guidance to improve data privacy and technology contracts, code creators have inadequate incentives to write more user-friendly and privacy-sensitive code. They believe their contracts to protect them from almost all liability, and that users are powerless to negotiate. Creators can impose their products on consumers on their own terms—terms which, as I have argued elsewhere, may be unconscionable from the perspective of a reasonable consumer.[75] Stating the argument another way, using language reflecting the spirit of the SEC's Plain English Rule, contracting practices that may have started out embodying the traditional resilience of contract law have crept into the realm of potentially embodying unfair trade practices. Using the language of developmental psychology, the current state of affairs in digital contracting actively erodes resilience rather than building it, an undesirable result that hampers the future of the information technology marketplace.

75. *See generally* Andrea M. Matwyshyn, *Technoconsen(t)sus*, 85 WASH. U. L. REV. 529 (2007).

The Internet Ecosystem: The Potential for Discrimination

Dirk Grunwald*

---

* Dirk Grunwald is the Wilfred and Caroline Slade Endowed Professor in the Department of Computer Science, the Department of Electrical and Computer Engineering, and the Interdisciplinary Telecommunications Program at the University of Colorado. He received his Ph.D. from the Department of Computer Science at the University of Illinois in 1989. Alongside his many gifted students, he has also studied specialties including computer networking, wireless networking, privacy in wireless networks, mechanisms to enforce and counter anonymity in the Internet, advanced techniques to compile languages for emerging hardware devices, and computer architecture and design. His current research involves the design and evaluation of Software Defined Radios and Cognitive Radios, or wireless systems that adapt to their environment and coordinate with one another to achieve high performance.

I. THE PREMISE BEHIND NETWORK NEUTRALITY

The premise behind the current debate in network neutrality was articulated in an FCC policy statement adopted in August 2005[1] that stated four goals for the Internet:

1. "[C]onsumers are entitled to access the lawful . . . content of their choice."[2]

2. "[C]onsumers are entitled to run applications and use services of their choice, subject to the needs of law enforcement."[3]

3. "[C]onsumers are entitled to connect their choice of legal devices that do not harm the network."[4]

4. "[C]onsumers are entitled to competition between network providers, application and service providers, and content providers."[5]

Rules that have been proposed since would extend these four core principles by adding two additional rules:[6]

1. A provider of broadband Internet access service must "treat lawful content, applications, and services in a nondiscriminatory manner."[7]

2. A provider of broadband Internet access service must "disclose such information concerning network management and other practices as is reasonably required for users and content, application, and service providers to enjoy the protections specified in this rulemaking."[8]

Broadly speaking, participants in the network neutrality debate use the same term to conflate two issues—accessing content of their choice

1. Appropriate Framework for Broadband Access to the Internet over Wireline Facilities, *Policy Statement*, 20 F.C.C.R. 14986 (2005) (including the publication of the original "four rules").

2. *Id.* at para. 4.

3. *Id.*

4. *Id.*

5. *Id.*

6. Preserving the Open Internet and Broadband Industry Practices, *Notice of Proposed Rulemaking*, 24 F.C.C.R. 13064 (2009) (containing two additional rules) [hereinafter Preserving the Open Internet *NPRM*].

7. *Id.* at para. 16.

8. *Id.*

and, more narrowly, enabling the development of a competitive environment for services, applications, and content providers by maintaining "neutral" access to the last link for consumers or the "public" Internet (the "access network").

The two primary concerns have been that access network providers would provide preferential treatment to specific uses of the network and may go so far as to block certain kinds of applications.[9] To support this concern, proponents of regulation point to a small number of documented cases where ISPs have blocked specific services (VOIP[10] and file sharing[11]). There is concern about a lack of transparency in network management and how that might diminish the opportunity for innovation in the Internet or unfairly limit competition. But the ability to limit access to Internet applications is not restricted to access networks. Such restrictions can be imposed by many components used to access Internet content, such as the browser and services or applications within the Internet.

Likewise, there are many ways to enable preferential access. In a 2007 article, this Author, along with Douglas Sicker, discussed aspects of current Internet access network designs that can lead to higher barriers for innovation and new services or can allow subtle forms of preferential network access.[12] We specifically focused on *asymmetric access links* and *content distribution networks* (*CDNs*). Asymmetric access networks make it more difficult for consumers to "self-publish," and commercial content distribution networks[13] can effectively provide "preferential access" to content provisioned on a CDN located within an ISP's network without actually violating "neutral" access network policies.

---

9. Formal Complaint of Free Press and Public Knowledge Against Comcast Corporation for Secretly Degrading Peer-to-Peer Applications, *Memorandum Opinion and Order*, 23 F.C.C.R. 13028 (2008) [hereinafter Free Press], *vacated by* Comcast Corp. v. FCC, 600 F.3d 642 (D.C. Cir. 2010).

10. *See* Madison River Comm., *Consent Decree*, 20 F.C.C.R. 4295 (2005).

11. *See* Free Press, *supra* note 9.

12. *See* Dirk Grunwald & Douglas Sicker, *Measuring the Network–Service Level Agreements, Service Level Monitoring, Network Architecture and Network Neutrality*, 1 INT'L J. COMM. 548, 551–52 (2007), *available at* http://www.ijoc.org/ojs/index.php/ijoc/article/viewFile/163/98. The article raised the issue of how "non-discriminatory" attributes such as asymmetric link access could impair expression and competition as much as access network management practices. *Id.* Most broadband access networks have higher download speeds than upload speeds. These communication asymmetries make it difficult for consumers to host services in their home or to generate content.

13. Examples of "Content Distribution Networks" (or CDNs) include Akami, Limelight, and Amazon Cloudcast. These services make multiple copies of content available at multiple physical locations in the Internet, improving the experience of accessing that content under periods of high demand. *See* Christopher S. Yoo, *The Evolution of Internet Architecture: Innovations in the Internet's Architecture that Challenge the Status Quo*, 8 J. TELECOMM. & HIGH TECH. L. 79 (2010).

We argued that these barriers impose as much risk as preferential treatment of access networks, but that network neutrality regulation focused solely on access networks would be unlikely to address these barriers.[14] Instead, the proposed regulations may hamper network innovation at the access network, as well as the core of the network, while still leaving open the door for anticompetitive actions that the regulations are intended to forestall.

This Article explores other parts of the Internet ecosystem and how they affect open and competitive networks. There is broad consensus that layers of the Internet ecosystem other than the access network may impact competition and innovation—the question remains as to whether new rules are needed. In the conclusion of a paper describing the economic history of price discrimination in telecommunications networks,[15] Andrew Odlyzko wrote:

> For telecommunications, given current trends in demand and in rate and sources of innovation, it appears to be better for society not to tilt towards the operators, and instead to stimulate innovation on the network by others by enforcing net neutrality. But this would likely open the way for other players, such as Google, that emerge from that open and competitive arena as big winners, to become choke points. So it would be wise to prepare to monitor what happens, and be ready to intervene by imposing neutrality rules on them when necessary.[16]

Odlyzko's point was that what he termed "cloud computing"[17] would become a more important marketplace for innovation than services integrated into access networks; his implication mirrors that of this Article —focusing on those access networks may distract from anticompetitive behavior in those other markets.

This Article is in agreement with Odlyzko's observation that other parts of the Internet ecosystem are equally powerful in determining the rich, competitive environment of the Internet and show this for past, current, and emerging parts of the Internet. At the same time, this Article argues that regulation and action—either that proposed for the access network or extending beyond those networks (through ambiguity or design)—should be applied only when clear harms are shown. The development of specific technologies coupled with the pace of technology development, the continued innovation of the Internet community, and the use of existing laws has served the Internet well.

---

14. Grunwald & Sicker, *supra* note 12, at 555–58.

15. Andrew Odlyzko, *Network Neutrality, Search Neutrality, and the Never-Ending Conflict Between Efficiency and Fairness in Markets*, 8 REV. NETWORK ECON. 40 (2009).

16. *Id.* at 57.

17. By this term, Odlyzko meant software services hosted on computers not located at a person's home or business. *See id.* at 41, 51, 57. Later, this Article will discuss that current common usage has two meanings for this term and will disambiguate those meanings.

The FCC's *Notice of Proposed Rulemaking* (*NPRM*), released in October 2009,[18] attempts to ensure a competitive marketplace, but it does so through regulating one subset of providers and certain specific network characteristics such as traffic priorities[19] and managed services (having multiple services use a single physical transport).[20] This focus ignores the fact that the Internet evolves over time and is far from a finished work. In fact, the National Science Foundation (NSF), the national agency that has long funded Internet research, has launched multiple research programs to define the future Internet.[21] Extending the existing Internet is difficult because it has become essential to society, but there are clear reasons to improve on the current design. Would regulation add yet more friction to the process of improving the Internet? Are we doomed to the Internet of today?

Rather than use words like "discrimination," network engineers prefer terms like "network management" and "prioritization."[22] One form of prioritization endemic to the Internet is "congestion control"; congestion occurs in a network when too many packets try to use the same resource (link or router). The Internet Protocol[23] handles congestion by simply discarding packets when resources are limited, but congestion requires that the transmitter slow down, or the network can enter a "congestion collapse" whereby no useful communication takes place.[24] The original Internet

---

18. Preserving the Open Internet *NPRM*, *supra* note 6.

19. *See id.* at para. 16. The use of the word "nondiscriminatory" in the proposed rules is regrettable. *Id.* From a technical perspective, discrimination can mean any form of differentiation, including simple traffic prioritization designed to improve performance; however, the word is laden with other meanings by events and history external to network engineering.

20. *Id.* at paras. 148–53.

21. The "Future Internet Directions" program (FIND) has funded research to address how parts of the Internet design need to change in response to new demands and technologies. *See NSF NeTS Find Initiative*, NAT'L SCI. FOUND., http://www.nets-find.net/ (last visited Feb. 21, 2011). The NSF Global Environment for Networking Innovations (GENI) program is funding the development of test platforms and new technologies for future Internets. *See* GENI: EXPLORING NETWORKS FUTURE, http://www.geni.net/ (last visited Feb. 21, 2011). Similar efforts are underway in Europe, Japan, and other countries as well through the Future Internet Research and Experimentation program. *See* FIRE, http://www.ict-fire.eu/ (last visited Feb. 21, 2011).

22. An overview of the history and design of congestion control and related network management techniques can be found in Steven Bauer, David Clark & William Lehr, The Evolution of Internet Congestion (2009) (unpublished paper) (on file with the Massachusetts Institute of Technology), http://www.tprcweb.com/images/stories/papers/Bauer_Clark_Lehr_2009.pdf.

23. The Internet Protocol specification is published by the Internet Engineering Task Force as an online document. INFO. SCI. INST., DARPA INTERNET PROGRAM PROTOCOL SPECIFICATION RFC 791 (rel. Sept. 1991), http://www.ietf.org/rfc/rfc791.txt.

24. For example, assume two transmitters are trying to use a single common link that has a capacity of 100 packets per second. Both transmitters want all of their data to be

design principles emphasized "end-to-end" control[25] and assumed that the computers at each end of a transmission would cooperate to prevent congestion collapse. In 1986, the network experienced a series of congestion collapses that reduced useful throughput by factors of 10 to 1000.[26] New congestion control methods were introduced then and have continued to be developed. Different congestion control methods, implemented on devices or working in concert with network routers, affect how competing network flows use the networks to improve the overall efficiency of a complex, distributed, and decentralized system. Would this research and innovation be possible with the proposed FCC rules in place?

Although the Internet is forty years old, the commercial Internet is only fifteen to twenty years old. New applications and an increased number of users change assumptions that network engineers have made and expose the network to new challenges with the concomitant need for new solutions. In an effort to maintain a rich Internet environment, the proposed regulations focus on access networks without considering how anticompetitive pressures can be applied in the remainder of the Internet. They also regulate a mechanism (traffic prioritization) that is used in congestion control, but at the same time is part of the basic Internet design. Likewise, although the FCC's *NPRM* addresses the distinction between the "managed" and "public" Internet, it does so in a limited way that may hamper innovation in "managed" networks or in the interface between private and public networks.

This Article argues that there are better ways to maintain a vibrant Internet. These include: having clear standards and methods for measuring what is actually happening in the Internet, as well as methods for reporting or disseminating policy to consumers; using existing agencies and policies;

---

received and will retransmit packets if they are discarded. If one transmitter injects 100 packets per second on to that link while the other injects 10, some packets will have to be discarded. Assuming a random discard policy, ninety-one percent of the discarded packets will be from the higher rate transmitter. If the transmitter determines that those packets were dropped, it would retransmit those packets in addition to the existing 100 packets per second, resulting in increased congestion. As more and more packets from the faster transmitter are dropped, it will increase the sending rate until its access link capacity is reached. This "congestion collapse" ensures that an increasing number of packets are discarded and also negatively affects the slower transmitter, because its packets will make up an increasingly dwindling portion of the packets that traverse the congested link.

25. *See* J. H. Saltzer, D. P. Reed & D. D. Clark, *End-to-End Arguments in System Design*, 2 ACM TRANSACTIONS COMPUTER SYS. 277 (1984). The core point of the paper concerned the engineering flexibility of having "the end points" (computers and servers) control what was communicated and how traffic was managed. *Id.* This was in stark contrast to the existing telecommunications systems that had "dumb end points" (telephones) and a smart network.

26. *See* Van Jacobson, *Congestion Avoidance and Control*, COMPUTER COMM. REV., Aug. 1988, at 314 ("[We] were fascinated by this sudden factor-of-thousand drop in bandwidth and embarked on an investigation of why things had gotten so bad.").

encouraging innovation and competition for access networks; and developing "best practices" that can be clearly understood by network operators, regulators, and consumers.

II. RISKS TO THE INTERNET ECOSYSTEM

The Internet is composed of many parts that make up the "experience" that end users now confront. Just as the phone network is made more useful by 411, white pages, yellow pages, 911, and other services or applications, the Internet is made more useful by domain names, browsers, search engines, and services that are integral to the web. Ensuring competition and a rich Internet environment by *solely* focusing on the local loop, as is being done with the Internet, clearly misses the mark—the entire "ecosystem" that influences either network experience is important.

To understand how applications and services can foster an anticompetitive environment, this Article examines a series of past concerns about Internet exclusion and market dominance, starting with the platforms that enabled web access, and stretching to services that now generate the most debate. These examples illustrate the rapid pace of innovation and demonstrate that the Internet often innovates its way out of anticompetitive markets; they also show that even when that does not happen, existing laws and regulations enforced by the Federal Trade Commission and the Justice Department can level the playing field.

A. Access to the Web—the Browser

The web browser is an application that has had almost total market dominance by multiple companies at different times. One of the earliest graphical Internet web browsers was Mosaic, developed by students and staff at the University of Illinois.[27] The Mosaic developers founded Netscape to commercialize the browser.[28] Although other companies, particularly Microsoft, developed other browsers in the mid-1990s, Netscape maintained approximately an eighty- to ninety-percent share of the browser market until Microsoft bundled its own product, Internet Explorer, with Windows 98.[29] Netscape's fortunes quickly soured as

27. The National Center for Supercomputing Applications at Illinois maintains a history of the development of Mosaic. *About NCSA Mosaic*, NCSA, http://www.ncsa.illinois.edu/Projects/mosaic.html (last visited Feb. 21, 2011).

28. See Jim Clark's book documenting the rise of the Netscape company. JIM CLARK, NETSCAPE TIME: THE MAKING OF THE BILLION-DOLLAR START-UP THAT TOOK ON MICROSOFT (1999).

29. *Id.* Other reports of browser market share are collected and referenced at the Wikipedia article, *Usage Share of Web Browsers*, WIKIPEDIA, http://en.wikipedia.org/wiki/Usage_share_of_web_browsers (last visited Feb. 21, 2011).

Internet Explorer reached a ninety-percent share of the browser market; Internet Explorer now has sixty-three-percent market share, having lost share to browsers developed in the last five years.[30]

It is rare for a market to switch from total domination by one product to another so quickly. However, as Netscape discovered, the problem with marketing a browser was how to monetize the product. Most businesses were hoping to use the browser to steer users to specific web properties.[31] Open standards allow rapid substitution of one product for another and can equally favor the adoption of software that "extends" those standards. Internet Explorer enabled Microsoft to launch protocols that favored other Microsoft products (either Windows desktops or Windows server). Chief among these were "ActiveX controls," a mechanism to embed software unique to Windows in a web page. Many of these "controls" provided mechanisms missing in the web (such as audio or video); because ActiveX only worked with Microsoft clients, the use of such controls drove many to rely on Microsoft software. The combined control of the most common operating system and the pre-installed browser brought on antitrust actions and an initial finding of monopoly power.[32]

Although Internet Explorer still dominates the browser market, alternate services and new technologies and standards eliminated much of the threat of Internet Explorer. AOL eventually purchased Netscape and much of the code-base was spun off into the popular open-source "Mozilla" and later "Firefox" browser platform.[33] Additional vendors, primarily Apple, Opera, and now Google, produced other competitive browsers. Increased broadband speeds and better software installation and update processes made it easier to install competing browsers. At the same time, browsers became ubiquitous, emerging as a universal way to access and control devices ranging from printers to alarm clocks—manufacturers wanted those controls to be universal. A widespread "open standards" effort ensued to identify browser techniques that limited users to Windows-based computers; lobbying and branding by the World Wide Web Consortium (W3C) led governments and many companies to eschew IE-specific mechanisms to focus on a "works with any browser" standard.[34] At

30. Browser adoption rates are highly regional at an international level. *See* Gregg Keizer, *See Google's Chrome Grabs No. 3 Browser Spot from Safari*, COMPUTERWORLD (Jan. 2, 2010), http://www.computerworld.com/s/article/9142958/Google_s_Chrome_grabs_No._3_browser_spot_from_Safari. There are few longitudinal academic studies of browser shares, but the Wikipedia article provides referenced studies from a variety of international website and Internet service provider measurements. *See Usage Share of Web Browsers*, *supra* note 29.

31. *See* CLARK, *supra* note 28.

32. *See, e.g.*, United States v. Microsoft Corp., No. 98-1232 (D.D.C. Nov. 12, 2002).

33. *See* CLARK, *supra* note 28.

34. *See* Tim Bresnahan, *A Remedy That Falls Short of Restoring Competition*,

the same time, the development of "Web 2.0" technologies such as Ajax around 2004,[35] coupled with increased broadband speeds, meant that many of the Microsoft-specific "ActiveX controls" could be replaced by software that worked across all browsers. The impetus for a standards-based browser has become particularly important as web browsers have become an integral part of mobile phones that are unable to use Windows-specific features, such as the iPhone.

Although Internet Explorer still dominates the browser market, that dominance connotes little economic advantage to Microsoft at this point; the majority of Microsoft profits are still generated from sales of Windows and Office rather than online products.[36] However, without the development of alternative software and open standards by organizations such as W3C, the present situation might not have come about and could rapidly change. It is arguable that the antitrust investigation of Microsoft was what led to the current situation. It is equally plausible that the development of mobile phones and the demands of that emerging non-Windows ecosystem, or the deployment of broadband and more interactive web pages using Ajax, forestalled the dependency on Microsoft-specific features. One thing is certain: competition, innovation, and existing legal recourse opened access to the Internet without the need for additional regulation.

B. Rich Internet Applications, Video, and the New Content Companies

Less well known than the "browser wars" is the (ongoing) battle for

---

ANTITRUST, Fall 2001, at 67. Similarly, in 1996, Tim Berners-Lee stated in the July issue of the MIT Technology Review "[a]nyone who slaps a 'this page is best viewed with Browser X' label on a Web page appears to be yearning for the bad old days, before the Web, when you had very little chance of reading a document written on another computer, another word processor, or another network." Herb Brody, *The Web Maestro: An Interview with Tim Berners-Lee*, TECHNOLOGY REV., July 1, 1996, at 33.

35. Ajax is a term used to describe one way in which "rich" web applications are developed using nothing more than standard web browser protocols. Gmail, released by Google in 2004, was one of the first widely known Ajax applications. Jesse James Garrett coined the term while working at Adaptive Path. *See* Jesse James Garrett, *Ajax: A New Approach to Web Applications*, ADAPTIVE PATH (Feb. 18, 2005), http://www.adaptivepath.com/ideas/essays/archives/000385.php for a readable description of the technology.

36. See MICHAEL CUSUMANO, MICROSOFT SECRETS: HOW THE WORLD'S MOST POWERFUL SOFTWARE COMPANY CREATES TECHNOLOGY, SHAPES MARKETS, AND MANAGES PEOPLE (1998) for details on Microsoft business strategy. The 2002 Annual Report for Microsoft indicates that Desktop and Enterprise Software (mainly Office and Word) contributed $23.8 billion to revenue in 2002 resulting in a $14.7 billion income while all Consumer Software Services Devices (web properties, ISP and game systems) revenue was $3.5 billion, resulting in a loss of $1.8 billion. MICROSOFT CORP., FORM 10-K, ANN. REP. (June 30, 2002).

"rich Internet applications" (RIAs).[37] The RIA is now a fundamental part of the Internet ecosystem. These environments provide extended usability to systems like Google Mail (Gmail), Netflix, Hulu, Microsoft Live, Yahoo! News, and many other websites—RIAs allow conventional "desktop" applications to be replaced by web-based applications. The features that made Internet Explorer indispensible in many areas were for "rich web applications"; RIA environments make that approach work across different operating systems. Microsoft sought to use the Windows infrastructure to allow developers to use existing Windows code in web applications. The primary alternative approach was Java, developed by Sun Microsystems, by which programmers could develop "applets," or programs that ran within a web browser. Although the Java language found extensive use in business software, applets experienced limited success, largely because the process of installing software was relatively complex. Macromedia Flash was introduced in 1996 and rapidly became the primary RIA tool; it is currently installed in more than ninety-percent of browsers and is used to power many video and online game sites.[38] Later entrants were Microsoft Silverlight (similar to Flash and Java) and Adobe AIR (developed as an extension to Flash when Adobe acquired Macromedia).[39]

Surprisingly, there has been little concern to date that any of these alternatives would preclude effective competition. In large part, this is because there are "open source" implementations of the dominant platform (Flash) and any one system is largely substitutable for the other (although not always on the same device). More importantly, existing and new standards-based technologies are replacing many of the functions for which developers turn to RIA frameworks. Microsoft argued this point in a 2007 response[40] to a motion by the State of California and several other states,[41] which argued that Microsoft's development of Silverlight should extend the

---

37. *See* Jim Rapoza, *RIA War Is Brewing*, EWEEK EMERGING TECH. (Apr. 11, 2008, 3:07 PM), http://etech.eweek.com/content/application_development/ria_war_is_brewing.html.

38. Adobe maintains statistics on the adoption or "penetration" of Adobe Flash at *Flash Player Version Penetration*, ADOBE, http://www.adobe.com/products/player_census/flashplayer/version_penetration.html (last visited Feb. 21, 2011).

39. The adoption rate of competing tools is collected by several online measurement forums; the reports at StatOwl.com show historical trends for the three main technologies, Flash, Java, and Silverlight. *Rich Internet Application Market Share*, STATOWL.COM, http://www.statowl.com/custom_ria_market_penetration.php (last visited Feb. 21, 2011).

40. Microsoft's Report Concerning the Final Judgments, United States v. Microsoft Corp., No. 98-1233 (D.D.C. Aug. 31, 2007), http://www.microsoft.com/presspass/download/legal/SettlementProceedings/08-30MSFTReportConcerningFinalJudgments.pdf.

41. Plaintiff States' Motion to Extend the Modified Final Judgment Until Nov. 12, 2012, New York v. Microsoft Corp., No. 98-1233 (D.D.C. Oct. 16, 2007), http://blog.seattlepi.com/microsoft/library/califfiveyears.pdf.

earlier antitrust actions.[42] Some Microsoft web services (such as Bing 3-D maps) still require Silverlight and ActiveX controls. Others argue that the required use of Silverlight for specific high-profile events (such as Olympic events and presidential inaugurations) and bundling of Silverlight with Windows 7 will raise the same anticompetitive issues that Netscape faced in the 1990s.[43]

The argument that "open" alternatives suffice is compelling. Applications developed by Google, such as Gmail, Maps and "Instant Search," only rely on JavaScript, a programming language that has long been a standard tool embedded in web browsers.[44] Rather than develop a new programming environment, Google, Apple, and Firefox have worked to greatly increase the usefulness of JavaScript, making that standard tool more suitable for many "rich" applications. The web standards community also developed HTML5,[45] the latest variant of the *lingua franca* of web browsers. That standard supplants many of the reasons RIA frameworks were needed, such as high performance video playback, access to geographic location, and support for storing and accessing data via the browser. These individual components allow large changes to applications—for example, using HTML5, Gmail can function more like a standard e-mail client allowing access to e-mail even when not connected to the Internet.

This analysis of RIA environments serves to show how regulation decisions are interconnected by past technology. Had Microsoft "won" the browser wars, most of this innovation would not have occurred—developers would have used Microsoft components rather than adopt a new RIA framework. This would have also altered the landscape of devices,

---

42. *See generally id.*; *see also* Todd Bishop, *Antitrust Filing Cites Microsoft Silverlight Concern*, *The Microsoft Blog*, SEATTLE POST INTELLIGENCER BLOG (Oct. 17, 2007, 11:57 AM), http://blog.seattlepi.com/microsoft/archives/123837.asp (offering an analysis of the filing).

43. *See* John Markoff, *Microsoft Leveraging Silverlight and Riling Critics*, N.Y. TIMES, Aug. 11, 2008, *available at* http://www.nytimes.com/2008/08/11/technology/11iht-stream11.1.15135139.html?_r=1 (arguing that there was significant evidence that Microsoft was pursuing such a strategy). Although Silverlight has notable successes including streaming Netflix videos, the concern that Silverlight would dominate the other technologies appears to be waning in 2011.

44. It should be noted that the development of JavaScript was not without contention. Netscape initially developed JavaScript; Microsoft developed a competing version and submitted that version for standardization. Rather than splintering web standards, JavaScript came to unify them through standardization efforts.

45. HTML5 is the fifth major revision to the core "language" used to describe web pages. The primary changes in HTML5 compared to earlier versions are standards for video, storing information at the browser, and a better way of drawing or displaying text and drawings. For a full specification, see *HTML5: A Vocabulary and Associated APIs for HTML and XHTML*, W3C EDITOR'S DRAFT, http://dev.w3.org/html5/spec/Overview.html (last visited Feb. 21, 2011).

such as the iPhone, that are used to access the web. The competitive alternatives are so diverse and rich that government intervention is not needed; rather, the past experience of the "browser war" shows that existing methods for intervention are possible and effective when needed.

C. Naming and Information Discovery

Names play a central role in the Internet—people need to be able to access websites and services. The Domain Name System (DNS), which translates names to IP addresses, is central to naming in the Internet. Naming is one of the clearest cases of regulation applied to Internet services, and a number of national and international laws, rules, and bodies have been created to address names, particularly as applied to commercial interests. With the rise of the commercial Internet, the Internet Corporation for Names and Numbers (ICANN) devised a Uniform Dispute Resolution Policy for the ownership of domain names clearly related to existing trademarks and properties.[46]

Today, search has taken on the importance originally attributed to DNS names. No part of the Internet Ecosystem would appear to be as important as search, as search is now a universal way for finding new information, even supplanting the common use of domain names. Many of the most common search terms on Google are the names of (often competing) web services, indicating that users rely on search for even trivial or well-known information.[47]

Should search be regulated? Recently, there have been calls for such regulation often based on the dominance of a single search engine.[48] While this rationale is similar to that of DNS, there is a distinct difference—DNS was a single system essential to the core operation of the Internet, while Google (for example) is one of many search services. Moreover, search services were not originally intended to identify commercial interests—they were intended to "discover information."

Although Google dominates current search services, there have been numerous popular search services over time—AltaVista, GoTo.com, Ask.com, Yahoo!, and different Microsoft systems. The current dominance of Google (currently estimated at approximately sixty-five- to eighty-five-

---

46. The ICANN policies are described at *Domain Name Dispute Resolution Policies*, ICANN, http://www.icann.org/en/udrp/#udrp (last visited Feb. 21, 2011).

47. The Google Trends service provides statistics on current and historical popular search terms. *Trends*, GOOGLE, http://www.google.com/trends (last visited Feb. 21, 2011). That information is collected into the Google Zeitgeist to give a yearly summary of search trends. *Zeitgeist 2010: How the World Searched*, GOOGLE ZEITGEIST, http://www.google.com/intl/en/press/zeitgeist2010/ (last visited Feb. 21, 2011).

48. *See generally* Oren Bracha & Frank Pasquale, *Federal Search Commission? Access, Fairness, and Accountability in the Law of Search*, 93 CORNELL L. REV. 1149 (2008).

percent of U.S. market share)[49] coupled with the consolidation of online advertising, has led some to call for regulation of search engines and search-based advertising to make it "neutral."[50] The key objection is that search (and Google specifically) is so influential on the way people find information that it constitutes a "gatekeeper" on the Internet.[51] In one *New York Times* Op-Ed article,[52] Adam Raff, founder of a company promoting an alternative search engine, describes how Google has promoted its own products (e.g., maps, shopping services) over that of other companies in search results.[53] It is difficult to know why a specific Internet tool falls from favor—for example, Google Maps is now preferred over MapQuest. Clearly, advertising a service is one reason, but so are features and usability. It is difficult to simultaneously argue that customers are unlikely to flock to a new search engine, but would rapidly switch to new mapping software simply because it is well advertised. Advertising drives the substantial growth of Google; existing antitrust measures would seem to govern and appear to have been successfully applied in specific instances, such as to counter the proposed joint Yahoo!-Google advertising pact[54] and exclusive licensing of digital books.

Many of the arguments for regulating search are based on the difficulty of effective competition.[55] Search is composed of three main components—crawling, indexing, and presentation. Crawling is the traversal of web pages—bringing the content of those pages to be indexed. Indexing records the information in the pages so that specific web pages can be quickly identified. Retrieval and presentation transform search requests into queries that search the indices and present the results to the users. Oren Bracha and Frank Pasquale argue that creating search engines is costly, but as with much of the infrastructure of the Internet, the software

---

49. *See, e.g.*, *Search Engine Market Share*, NETMARKETSHARE, http://marketshare.hitslink.com/search-engine-market-share.aspx?qprid=4 (last visited Feb. 21, 2011) (85%); Nathania Johnson, *comScore Shows Bing Growing in December 2009 Search Rankings*, SEARCHENGINEWATCH.COM (Jan. 19, 2009, 7:54 AM), http://blog.searchenginewatch.com/100119-075446 (65%). As with web browser choices, different search engines are popular in different markets. *See, e.g.*, *Search Engine Market Share*, NETMARKETSHARE, http://marketshare.hitslink.com/search-engine-market-share.aspx?qprid=4 (last visited Feb. 21, 2011).

50. *See* Bracha & Pasquale, *supra* note 48.

51. *See id.*

52. Adam Raff, Op-Ed, *Search, But You May Not Find*, N.Y. TIMES, Dec. 28, 2009, at A27.

53. *Id.*

54. Press Release, U.S. Dep't of Justice, Yahoo! Inc. and Google Inc. Abandon Their Advertising Agreement (Nov. 5, 2008), *available at* http://www.justice.gov/opa/pr/2008/November/08-at-981.html.

55. Pasquale makes this point specifically. Bracha & Pasquale, *supra* note 48, at 1179–81.

to develop effective and scalable search engines is now free. The Apache Foundation, an organization that manages the development of the free Apache web server, also distributes Nutch, an open source search engine,[56] and Lucene, a free indexing mechanism.[57] Yahoo! has also donated Hadoop, software designed to rapidly index large numbers of web pages.[58]

Although the software is free, adoption of new search engines depends on the utility they provide to users. This is usually based on the effectiveness of presenting the results of a search query. Ranking determines the order in which the most important search results are displayed. The GoTo.com search engine pioneered the "money talks" policy of paid search rankings and Google "AdWords" expanded that base with an auction-based scheme.[59] In many ways, the barriers presented by search engines and ad rankings are similar to the yellow pages. Businesses were at a disadvantage if they did not place paid advertisements in yellow page directories. One of the complexities that search companies face is that the variables governing advertisement (for example, placement, frequency, relation to search) are more complex than those used in static print media. Defining and communicating those characteristics and having customers understand them are complicated tasks. There is always a need for transparency so that advertisers understand what they are purchasing, particularly when competing "house brands" are also advertised, as Adam Raff argued.[60] This situation is similar to grocery stores that present their own house brand and a diverse array of competing brands whose placement is governed by a combination of consumer demand and "slotting fees."[61] Slotting fees have received much discussion as well as government scrutiny and enforcement actions at state and national levels.[62] It seems likely that

---

56. *About Nutch*, NUTCH, http://nutch.apache.org/about.html#Overview (last visited Feb. 21, 2011).

57. *Apache Lucene–Overview*, LUCENE, http://lucene.apache.org/java/docs/index.html (last visited Feb. 21, 2011).

58. *Hadoop at Yahoo!*, YAHOO! DEVELOPER NETWORK, http://developer.yahoo.com/hadoop/ (last visited Feb. 21, 2011).

59. A brief history of paid search is included in Andrew Sinclair, Note, *Regulation of Paid Listings in Internet Search Engines: A Proposal for FTC Action*, 10 B.U. J. SCI. & TECH. L. 353 (2004).

60. Raff, *supra* note 52.

61. "Slotting fees are fees manufacturers pay to retailers in order to obtain shelf-space." Robert J. Aalberts & Marianne M. Jennings, *The Ethics of Slotting: Is This Bribery, Facilitation Marketing or Just Plain Competition?*, 20 J. BUS. ETHICS 207, 207 (1999).

62. For Aalberts & Jennings's study of the issue, see *id.* In November of 2003, the FTC Staff released a study that examined when slotting occurred. FED. TRADE COMM'N, SLOTTING ALLOWANCES IN THE RETAIL GROCERY INDUSTRY: SELECTED CASE STUDIES IN FIVE PRODUCE CATEGORIES (2003). Lastly, Gregory T. Gundlach testified before the California State Senate Standing Committee on Business, Professions and Economic Development on February 9, 2005, detailing the problem slotting fees cause for businesses. *Slotting Fees–Fees Charged by Grocery Retailers for Shelf Space: Are They Stifling*

anticompetitive behavior in search would encounter similar scrutiny, and the FTC has already asked companies to disclose paid search results.[63]

Despite the dominance of Google in the search-based advertising market, the search market itself has seen considerable innovation, in part because there are many *corpora* over which to search and many methods to rank or present results. Real-time search, personalized search, social search, and peer-to-peer search tools are in active development. OneRiot is a start-up that recently partnered with Yahoo! to develop "real-time" search (or search about breaking events rather than historical documents)[64] and Lijit is a search engine focused on blogs and social networking.[65] Ask.com and Aardvark focus on casting questions that are understandable to people into search queries.[66] It may be that no search engine could compete with Google in the sense of becoming a multi-billion-dollar company; many will be acquired by existing search companies—indeed, Google acquired Aardvark in February 2010. It is also important to recognize that Google, as a company, is little more than ten years old.[67] Given the low barriers to entry (other than customers), there should be continued innovation in search.

It is clear that search has become as important as naming in the Internet; it also influences the experience that users have because they have come to rely on the speed and accuracy of search to locate services. What is not clear is whether additional mechanisms beyond current laws are needed to ensure a competitive and innovative Internet.

D. Content Distribution and Cloud Computing—the Invisible Ecosystem

The Internet has visible components, such as the browsers, rich application frameworks, and search engines, as discussed. Equally important is the invisible infrastructure that defines how the web and web services are implemented. This Section will describe services that

---

Competition?: Statement Before the Cal. State S. Standing Comm. on Bus., Professions and Econ. Dev., 2005 Sess. (Cal. 2005) (statement of Gregory T. Gundlach, Senior Fellow, American Antitrust Institute), http://www.antitrustinstitute.org/files/386.pdf.

63. The FTC responded to a complaint filed by Commercial Alert, a consumer advocacy group. The details of the response to the Commercial Alert case were disclosed by a publically available response. *Commercial Alert Letter*, FED. TRADE COMM'N, http://www.ftc.gov/os/closings/staff/commercialalertletter.shtm (last visited Feb. 21, 2011).

64. *About*, ONERIOT, http://www.oneriot.com/about (last visited Feb. 21, 2011).

65. *Company: Who Is Lijit?*, LIJIT, http://www.lijit.com/company (last visited Feb. 21, 2011).

66. *About Ask.com*, ASK.COM, http://www.ask.com/about (last visited Feb. 21, 2011); *About Aardvark*, AARDVARK, http://vark.com/about (last visited Feb. 21, 2011).

67. *Google History*, GOOGLE, http://www.google.com/corporate/milestones.html (last visited Feb. 21, 2011).

dramatically lower the barriers for creating new web services. Just as open-source tools such as Nutch, Lucene, and Hadoop reduce the technical barriers for developing a new search engine or service, new business models and technology reduce the operational barriers to deploying and scaling those services.

Content distribution networks (CDNs), co-location, and peering arrangements are some of the most critical elements of the Internet ecosystem that affect the web as it is used today. A CDN is an organized network of computers that are often placed "close" to Internet users. Commonly accessed content is then stored on those computers and requests by web users are directed to "nearby" or lightly loaded computers. Content distribution networks can be used to save bandwidth since the content for a popular item does not need to be fetched from a distant location; this was the basis for the concern that focusing solely on the access network would not prevent performance discrimination.[68] However, with the drop in price for Internet bandwidth, CDNs have become useful primarily because they provide a way to provide *scalable* service. The canonical example for this is the success that Victoria's Secret (a retailer) had in hosting online content before and after using a commercial CDN.[69] In the initial offering, demand for the retailer's content exceeded the capabilities of its own web services, but successive offerings using a CDN were much more successful.[70]

The web would present a very different experience without CDNs, but the use of a CDN provides as much opportunity to discriminate performance as subtle packet differentiation or "traffic shaping" on an access network. Indeed, comments in FCC filings indicate that ISPs in China market their own content networks and hosting services as providing better access to their own clients.[71] In a competitive marketplace, the difference in performance is less a conspiracy than the result of innovative network architectures. Different combinations of CDNs and network management lead to differing degrees of efficiency, but efficient network architectures can still enable competition.[72] At the same time, CDNs

---

68. *See* Grunwald & Sicker, *supra* note 12.

69. A case study is available from Akamai, a cloud-based service provider. *Victoria's Secret Web Site Raises the Bar on Customer Experience with Content Delivery from Akamai and IBM*, AKAMAI, http://www.akamai.com/html/customers/case_study_victoria.html (last visited Feb. 21, 2011).

70. *Id.*

71. Comments of Daniel Scherlis, Notice of Ex Parte Communication, FCC GN Docket No. 09-191 (rel. Jan. 15, 2010). It should be clear that his comments concern the Chinese Internet market where two large companies dominate, but his experience serves as a cautionary note on the importance of competition.

72. Researchers are only recently beginning to study the economic benefit of different CDN organizations. *See* Wenjie Jiang et al., *Cooperative Content Distribution and Traffic*

enhance the ability of a web company or organization to successfully connect with readers without having to invest huge sums in capital infrastructure.

In a *Wall Street Journal* article in 2008, Vishesh Kumar and Christopher Rhoads noted arguments that such "fast-track" access violates net neutrality.[73] The fact is that *most commercial content* on websites is distributed using CDNs and that there is significant competition in CDNs in the United States.[74] The proposed FCC rules do not seem to address the importance of content distribution systems within the Internet ecosystem. This omission is arguably good, because no concrete harms have been shown—indeed, the existing "fast-track" access has enabled more companies to scale to meet web demand. But the omission highlights the rather arbitrary nature of the proposed FCC rules. The proposed rules would arguably also prohibit new services or offerings by "network operators" that could achieve the benefits of CDNs using different technical means, thus increasing competition in this segment of the Internet ecosystem.

Peering relationships between different ISPs, application providers, and Tier-1 network providers also enable "fast tracks" for information.[75] Most of those peering relationships have been historically "settlement free" because they benefit both parties and because traffic demands were symmetrical.[76] Increasingly, the line between "backbone," application, and edge network provider have blurred. Google and large CDN companies

---

Engineering in an ISP Network, SIGMETRICS/PERFORMANCE 2009: PROCEEDINGS OF THE ELEVENTH INTERNATIONAL JOINT CONFERENCE ON MEASUREMENT AND MODELING OF COMPUTER SYSTEMS (2009).

73. *See* Vishesh Kumar & Christopher Rhoads, *Google Wants Its Own Fast Track on the Web*, WALL ST. J., Dec. 15, 2008, at A1.

74. Dan Rayburn maintains a list of current CDN vendors; as of August 11, 2010, it listed around fifty companies. *See* Dan Rayburn, *Updated List of Vendors in the Content Delivery Network Business*, CDNLIST.COM (Aug. 11, 2010, 12:01 AM), http://www.cdnlist.com. While this market is currently very competitive, a history of significant price decreases indicates that consolidation may occur. It is often difficult to find authoritative pricing for Internet services, but a history of CDN pricing is published at a website called *The Business of Online Video*. *See Q4 CDN Pricing Detailed, Down 20% in 2010, Expected to Remain Stable Next Year*, BUS. ONLINE VIDEO, http://www.cdnpricing.com (last visited Feb. 21, 2011). The prices for delivered data declined twenty percent in 2010, although pricing is expected to remain more stable in 2011.

75. Christopher Yoo has published a very readable, accurate, and timely article concerning the rapidly evolving world of peering. *See* Yoo, *supra* note 13.

76. Peering is a complex subject that straddles engineering, law, and business. William Norton, a peering consultant, runs an education site and pricing data repository. His articles on the history of peering are available at *The Evolution of the U.S. Internet Peering Ecosystem*, DRPEERING INT'L, http://drpeering.net/white-papers/Ecosystems/ Evolution-of-the-U.S.-Peering-Ecosystem.html (last visited Feb. 21, 2011).

such as Limelight now run some of the largest Internet backbones.[77] At the same time, "edge" network companies such as Comcast, AT&T, and Verizon also carry considerable corporate or "non-public" network traffic.[78] Amid the consolidation in networking companies, "paid peering" has emerged as a way to enable content providers or other co-location companies to reduce the *cost* of access while improving performance for their hosted partners.[79] Content distribution networks (and peering) improve performance; being excluded from such interaction would raise costs or limit competition. Reaching a sizable population would be possible, but would require significant investment to be "scalable."

The proposed FCC rules do not clearly indicate whether peering and content distribution relationships constitute "neutral" access or in what situations they constitute "discriminatory" access. Again, this is arguably good, because there are few instances in which concrete harms have been demonstrated. In the past, the Internet has been "partitioned" because Internet providers could not agree on pricing for transit or peering relationships,[80] and more consumers have experienced network problems from these business disputes than those affected by the rules in the proposed FCC regulations. Is regulation needed to cover peering? History indicates that existing dispute resolution mechanisms (for example, lawsuits, agreements, and contracts) can resolve these problems. This lends credence to the argument that those same mechanisms will ensure competition in other Internet services such as CDNs.

Just as CDNs developed out of a need to replicate and distribute "static" content, a new market, "cloud computing," has emerged as a technology that subsumes CDNs and facilitates even faster changes in technology. Cloud computing providers such as Amazon EC2, Rackspace, AT&T, IBM, Microsoft, and several others run warehouse-sized data centers on which customers can lease and run customized software. Combined with "virtualization technology," which lets users capture the

---

77. Detailed reports of data collected at a major Internet exchange are reported in the *ATLAS Internet Observatory 2009 Annual Report.* C. LABOVITZ ET AL., ATLAS INTERNET OBSERVATORY 2009 ANNUAL REPORT (2009), http://www.nanog.org/meetings/nanog47/presentations/Monday/Labovitz_ObserveReport_N47_Mon.pdf.

78. *Id.*

79. Despite the rather arcane history of peering arrangements, some access network providers (such as Comcast) have clearly articulated rules for how peering relationships are established. *See, e.g.*, *Comcast Settlement-Free Interconnection (SFI) Policy*, COMCAST, http://www.comcast.com/peering/ (last visited Feb. 21, 2011).

80. For example, in 2008, Sprint and Cogent networks "de-peered" their networks, causing service disruptions between Sprint and Cogent customers. *See* Om Malik, *Cogent, Sprint Disconnect Networks, May Cause Web Slowdown*, GIGAOM (Oct. 30, 2008, 10:50 PM), http://gigaom.com/2008/10/30/cogent-sprint-un-peer-may-cause-web-slowdown.

entire configuration of a computer in a form that can be shipped off to a remote data center, cloud computing has changed the economics of establishing Internet services. Cloud computing systems can typically be leased by the hour and new online services can be launched quickly. For example, in early 2010, Amazon's EC2 (a service that popularized the cloud computing model) rented individual "machines" for $0.02 per hour to $2.48 per hour depending on the machine resources.[81] More importantly, since the leasing is "per hour" and because machines can be "turned on" quickly, software can be designed to use resources as needed.

Cloud computing has accelerated the deconstruction of monolithic software systems into components of a "service-oriented architecture" that can be used in multiple services. Examples include Twilio, which integrates the legacy telephone network and provides voice-guided phone services.[82] Such services, coupled with the ability to rapidly deploy systems using cloud computing, allow developers to innovate in a select part of the software systems. But all these components—CDNs, cloud computing, software as a service (SaaS) systems—are rapidly becoming integral to the way that applications and services are deployed on the Internet. How will they be affected by regulation?

III. THE RISKS OF REGULATION IN THE INTERNET ECOSYSTEM

There are several risks to the proposed network neutrality rules. These concerns include the lack of clarity as to whether "neutral" networks even exist or are beneficial, the uncertainty concerning how services and applications should be treated, and the risks of mandating monitoring for legal content and innovation in network management. This Section addresses a general concern about the ability or wisdom of applying regulation in an era of fast-paced technology development by examining a particular Internet application regulated by the FCC.

A. Insensible Neutrality

Proponents of network neutrality legislation assume that people could agree on what a "neutral" network is and that any management other than existing prioritization methods will break applications. Is it possible for consumers to spot a "non-neutral" network? If neutrality cannot be measured or sensed, it is difficult to know when it is being violated or if it

81. The listed prices are for machine instances, but any practical use of the service requires network bandwidth and storage, which are priced separately. *See Amazon Elastic Compute Cloud (Amazon EC2)*, AMAZON.COM, http://aws.amazon.com/ec2/#pricing (last visited Feb. 21, 2011).

82. *How It Works*, TWILIO CLOUD COMMS., http://www.twilio.com/how-twilio-works (last visited Feb. 21, 2011).

is even important. In an earlier work, this Author, along with Sicker, detailed how the lack of clearly stated service level agreements for residential service and the multi-party nature of the Internet make it difficult to know what is affecting performance and who is responsible.[83] Studies by networking researchers in 2003[84] (and also more recently in 2007[85]) have shown through careful measurement that the major performance limitations faced by most broadband users (such as latency, bandwidth, and jitter) occur because of the technologies used in the "last mile" access network—the connection to an individual house. At the same time, a study conducted of Internet users in the United States and Europe in 2009 showed that users' *home networks*, and in particular the use of "WiFi" wireless networks, impose more latency and variability than the access network itself.[86] These measurement studies were conducted so broadly (across multiple ISPs in multiple countries) that they indicate that latency limitations and variability exist in most access networks. These limitations are caused by pressing existing infrastructure (cable and phone lines) into service for purposes they were never intended to serve, rather than by anticompetitive actions.

Because the Internet is composed of many pieces made by different parties, it is difficult to understand what causes specific problems. This is true even for experts—in a network measurement study, members of the University of Colorado at Boulder research group (of which this Author is a contributor) initially reported many types of network sessions were being blocked; upon further analysis (and after much embarrassment) we had to retract that report because the problems were caused by a home networking router.[87] This action occurred only when the home router was overloaded, but if the cause was not immediately clear to networking researchers, it is unlikely that an average consumer could identify similar problems.

As is clear by the success of existing applications, Internet protocol

---

83. Grunwald & Sicker, *supra* note 12, at 550.

84. Aditya Akella et al., *An Empirical Evaluation of Wide-Area Internet Bottlenecks*, IMC'03: PROCEEDINGS OF THE 3RD ACM SIGCOMM INTERNET MEASUREMENT CONFERENCE 101–13 (2003).

85. Marcel Dischinger et al., *Characterizing Residential Broadband Networks*, IMC'07: PROCEEDINGS OF THE 2007 ACM SIGCOMM INTERNET MEASUREMENT CONFERENCE 43 (2007).

86. Gregor Maier et al., *On Dominant Characteristics of Residential Broadband Internet Traffic*, IMC'09: PROCEEDINGS OF THE 2009 ACM SIGCOMM INTERNET MEASUREMENT CONFERENCE 90 (2009).

87. *See* Karl Bode, *University of Colorado Researchers Retract Claims*, BROADBAND DSLREPORTS.COM (Apr. 7, 2008), http://www.dslreports.com/shownews/Comcast-Now-Forging-Packets-For-All-TCP-Traffic-93388. The issue was traced to a home router—the processors in many inexpensive home routers are too slow and not designed to handle high-traffic loads. When subjected to high loads, those routers also close connections using mechanisms that mimicked the mechanism used by Comcast.

and application designers understand that minor fluctuations in latency and bandwidth go with the territory of the current Internet. Applications and various parts of the broader "Internet architecture" are designed to accommodate those variations; there is good reason to believe that the design principles used in existing applications could overcome "subtle preferential treatment" just as they overcome the highly variable best-effort characteristics of the Internet. For example, video distribution systems came to rely on "faster-than-real-time" downloads to successfully deliver video on the existing Internet.[88] Despite the broad success of VoIP companies such as Vonage, Skype, and the like, highly interactive applications (voice or video communication and interactive gaming) are usually thought to be sensitive to latency. However, comments submitted to the FCC by interactive game developers indicate that the current Internet is suitable for those applications.[89]

All of this indicates that improving the speed of Internet access, rather than fixing current network designs into law, better serves consumers.

B. Fostering a Competitive Ecosystem

The proposed FCC rules affect only one part of the network, but performance and the user experience are affected by many parts of the network. Both content distribution and cloud computing resources are distributed globally and interconnected by private IP networks; since these are not "public networks," these facilities are free to prioritize traffic for payment without violating the proposed network neutrality rules. Singling out one part of the Internet for regulation does not seem to ensure the goal of competitive networks that respond to consumer needs.

There is continued vertical integration of the Internet market wherein "access network" providers also become CDNs, or application companies (like Google) or retailers (like Amazon.com) become cloud computing providers. It is unclear how proposed regulations that distinguish between "public" and "private" networks will apply as those network companies recombine and change form. This requires either greater clarity as to when the proposed network neutrality rules apply, or, better yet, a "wait and see attitude" with action taken when anticompetitive harms actually occur.

C. Regulating Legal Content

The proposed neutrality rules focus on *lawful* content, and there have been both calls and proposals for applying "deep packet inspection" to

88. Andrew Odlyzko, The Delusions of Net Neutrality 4–5 (Aug. 31, 2008) (unpublished paper) (on file with School of Mathematics, University of Minnesota), http://www.dtc.umn.edu/~odlyzko/doc/net.neutrality.delusions.pdf.

89. Comments of Scherlis, *supra* note 71, at 1.

assist in enforcing intellectual property ownership.[90] These efforts pose considerable costs and significant risks, both of misidentifying legal content as illegal and of failing to identify illegal content. Researchers have shown that anyone (including inanimate objects) can be implicated in file sharing.[91] Existing file-sharing systems are far from "stealthy" and are easy to monitor. Illegal file sharing is already hidden using "anonymity overlays"[92] and simple protocol extensions make it much more difficult to decidedly identify illegal file-sharing activity.[93]

At the same time, the rapid commoditization of co-location services, cloud computing, and content distribution networks are also affecting illegal content. Not only can new companies be launched quickly, but less legal Internet services are also possible. One of the many reasons that "peer-to-peer" (P2P) applications are popular is because they allow people to use their own infrastructure for file sharing. With the emergence of inexpensive cloud computing and other leased computing services, there has been a surge in the amount of Internet traffic for "hosted file services" at the expense of P2P services,[94] making it easier for file sharing to use those high-performance systems rather than rely on the low-bandwidth uplinks common to the asymmetric network architectures used for access.

The rapid change in infrastructure that drives much of the Internet ecosystem illustrates the challenge to monitoring unlawful content. In two short years, "bandwidth intensive" applications such as video and file sharing have moved to systems using the same protocols and service

---

90. AT&T has stated that it will filter Internet content for such purposes. *See* Tim Wu, *Has AT&T Lost Its Mind? A Baffling Proposal to Filter the Internet*, SLATE (Jan. 16, 2008, 10:15 AM), http://www.slate.com/id/2182152. Similar statements have recently been made by Comcast CEO Brian Roberts. *See* Kenneth Corbin, *Comcast Set to Enter Copyright Wars*, DATAMATION (Jan. 27, 2010), http://itmanagement.earthweb.com/cnews/article.php/3861096/Comcast-Set-to-Enter-Copyright-Wars.htm.

91. *See* Michael Piatek et al., *Challenges and Directions for Monitoring P2P File Sharing Networks – or – Why My Printer Received a DMCA Takedown Notice*, HOTSEC'08, PROCEEDINGS OF THE 3RD CONFERENCE ON HOT TOPICS IN SECURITY (2008) (showing that the method used in identifying file sharing is susceptible to false accusations).

92. Several such systems exist, the most common of which is the "Tor Network." *See* Damon McCoy et al., *Shining Light in Dark Places: Understanding the Tor Network*, PROCEEDINGS OF THE 8TH INTERNATIONAL SYMPOSIUM ON PRIVACY ENHANCING TECHNOLOGIES 63 (2008).

93. Again, several such systems exist. The system designed by our research group is very efficient and uses existing BitTorrent protocols. Kevin Bauer et al., *BitBlender: Light-Weight Anonymity for BitTorrent*, PROCEEDINGS OF THE WORKSHOP ON APPLICATIONS OF PRIVATE AND ANONYMOUS COMMUNICATIONS (2008).

94. This trend has been reported in numerous venues. One of the more detailed studies was *ATLAS Internet Observatory 2009 Annual Report*, which was presented to the 2009 NANOG network operators meeting. LABOVITZ ET AL., *supra* note 77. The report showed a dramatic increase in "hosted HTTP" services rather than the expected increase in P2P services. *Id.*

providers as "legitimate" services. Because those systems use encryption, any mandated monitoring of such traffic will be both expensive and error prone. Stopping illegal content by monitoring traffic requires that *all* traffic be monitored and the costs to implement this will be borne by all users of the Internet. Pushing this requirement on all network providers imposes a significant cost to benefit a different industry.

D. Curtailing Innovation in Network Management

The proposed neutrality rules distinguish between "managed" and "public" services, but the discussion about what constitutes managed services is relatively ad hoc and clearly captures the status quo rather than what is possible. An existing example would be having a distinct network service for latency-sensitive traffic, such as voice. Some existing "competition-friendly" networks use a managed network exclusively for one of many possible voice services and relegate "best effort" and streaming video services to other networks all carried on the same fiber.[95] Similar capabilities are present to varying degrees in almost all other access networks. Commercial Ethernet uses 802.1Q (Virtual LAN) and 802.1P (Class of Service) to provide such managed networks.[96] New home network technologies such as Multimedia over Coax Alliance (MoCA) and HomePlug are rapidly being developed and will allow different managed streams to be carried over the same physical cable.[97]

What have been missing are standards to link the differing streams in access network media to similar capabilities in home networks. A generalized capability to have multiple streams of data for multiple classes of service simplifies the distinction between "managed" and the "public" Internet and would allow additional managed services (for example, video-conferencing could extend current "triple play" networks) or service offerings that let consumers choose between multiple service qualities. Some of these mechanisms are being developed,[98] but such innovation will

---

95. Several Internet technologies promote "line sharing." The Ethernet-based architecture of the UTOPIA network is one of the most versatile designs. *See* Ken Moerman et al., *Utah's UTOPIA: An Ethernet-Based MPLS/VPLS Triple Play Deployment*, IEEE COMM. MAG., Nov. 2005, at 142.

96. Dirceu Cavendish, *Operation, Administration, and Maintenance of Ethernet Services in Wide Area Networks*, IEEE COMM. MAG., Mar. 2004, at 72–79.

97. HOMEPLUG: POWERLINE ALLIANCE, http://www.homeplug.org/home (last visited Feb. 21, 2011); MOCA, http://www.mocalliance.org/ (last visited Feb. 21, 2011).

98. The problem arises because attempts at standardizing "resource reservation" for the Internet have not been successful for a variety of business and technical reasons. However, most networks that makes up the Internet have mechanisms to reserve resources for specific tasks. For example, cable modems use the PacketCable standard, which uses a technique called "Reserved Services Domain" to handle managed services. The step that is lacking is connecting services in one network (e.g., PacketCable) to another (e.g., MoCA).

likely be halted if ambiguous regulation is in place.

Similarly, many existing access network technologies have impediments that limit performance; even seemingly high-performance networks such as DOCSIS cable modems benefit from "management" mechanisms to overcome such impediments.[99] Even long-studied systems in the Internet benefit from continued improvement. Congestion control algorithms are used to balance the performance of one "flow" of information versus another at all scales of the Internet. Recently, the design and "fairness" of these algorithms is being reexamined by the technical community. Steven Bauer, David Clark, and William Lehr published a very readable history of congestion control.[100] Internet connections "self-regulate" the bandwidth they use—without such self-regulation, TCP connections would only be limited by the ability of the sender to transmit data. Those algorithms seek to balance congestion in the network with the ability of the receiver to accept packets. The original algorithms sought to allocate each "flow" a fair share of bandwidth.[101] That design decision was reflective of the Internet at the time. Per-flow fairness is one reason why P2P applications exert more pressure on networks than, for example, simple host-based streaming—P2P applications use many connections to download content, and each is striving for a "fair share" of the access network. There are ongoing efforts to evolve network congestion control algorithms to include information from the network in order to build a more responsive and efficient network; network neutrality legislation seemingly precludes such efforts. These efforts include both the access network and congestion control at routers in the "core" of the Internet.[102]

---

99. DOCSIS cable modem networks tend to have "bursty" uplink connections, and this causes TCP/IP throughput to be lower than what the downlink can support. *See* Jim Martin, *The Impact of the DOCSIS 1.1/2.0 MAC Protocol on TCP*, CONSUMER COMMUNICATIONS AND NETWORKING CONFERENCE 302 (2005). This particular study examines the effectiveness of "TCP ACK compression" to see if it overcomes the problems in the physical access network. *See id.* This mechanism monitors TCP/IP connection characteristics and delays specific uplink traffic at the cable modem to eliminate redundant acknowledgement messages. As described in the study, this basic mechanism has been studied in other domains, but is rarely applied. *Id.*

100. Steven Bauer, David Clark & William Lehr, The Evolution of Internet Congestion (2009) (unpublished paper) (on file with the Massachusetts Institute of Technology) (prepared for the 37th Research Conference on Communication, Information and Internet), http://www.tprcweb.com/images/stories/papers/Bauer_Clark_Lehr_2009.pdf.

101. A "flow" represents a data connection between end points on a source and a destination in the Internet. Originally, each "application" (such as a web browser or file transfer program) would use a single flow at a time and "per flow" fairness results in "per user" fairness; over time, applications began to use more flows for performance and flexibility. The Evolution of Internet Congestion describes the history of these developments. *Id.*

102. The Evolution of Internet Congestion above describes some research studies on this topic. *Id.* One of the more readable descriptions is Matthew Mathis, *Reflections on the TCP*

E. Technology on Internet Time

The FCC orders affecting the AOL Instant Messaging system during the Time Warner and AOL merger provide a historical lesson about the risks and challenges of predicting the path of technology and the impact that regulation has on that path.[103] Instant messaging (IM) emerged in the mid-1990s as a popular communication system based on a long history of "computer chat" systems in place since the early 1970s. Messaging or "talk" applications were initially used on local area networks where the communication latency was sufficiently low. Because "chat" programs allowed users to communicate over long distances in near-real-time, they became increasingly popular on systems run by companies such as CompuServe, Prodigy, AOL, and others. As with much of the online content of those systems, chat systems were initially "walled gardens" that served only the members of those services.[104] As the commercial Internet evolved and became popular in the mid- to late- 1990s, there was greater interest in having IM systems operate across multiple services.

Instant messaging is notable because it is one of the few Internet technologies to have been affected by FCC and FTC orders. This occurred during the merger between AOL and Time Warner; Gerald Faulhaber wrote an excellent analysis and history of the reasoning behind orders affecting AOL Instant Messaging (AIM).[105] Lehman Brothers valued AIM as $5.8 billion during the merger in 2000.[106] AIM had 130 million members or users and appeared to have considerable market dominance over nascent IM alternatives such as Microsoft MSN Messenger.[107]

Prior to the merger, AOL and Microsoft had engaged in the "IM wars" wherein AOL exploited a security flaw in the AIM software to block interoperation with competing services, such as Microsoft Messenger.[108] Microsoft and other IM companies lobbied for open access to the AIM service as a condition of merger. Faulhaber argues that this was one of the first times that network effects were used as an argument in regulatory oversight in the absence of specific harm.[109] It was thought that if Time

---

Macroscopic Model, ACM SIGCOMM COMPUTER COMM. REV., Dec. 2008, at 47–49.

103. Comments in FCC Cable Services Bureau CS Docket 00-30, Time Warner Inc./AOL Time Warner Inc. Transfer of Control Applications; FTC Docket No. C-3989, America Online, Inc., and Time Warner Inc.

104. *See* Gerald Faulhaber, *Network Effects and Merger Analysis: Instant Messaging and the AOL-Time Warner Case*, 25 TELECOMM. POL'Y 311 (2002).

105. *See id.*

106. Comments of Covington & Burling at 1, Applications of America Online, Inc., and Time Warner Inc. for Transfers of Control, CS Docket No. 00-30 (rel. Sept. 27, 2000).

107. Louise Rosen notes this in *Why IM Matters So Much*, UPSIDE TODAY, Sept. 19, 2000, which appears in Comments of Covington & Burling, *supra* note 106, at 3.

108. *See* Faulhaber, *supra* note 104, at 314–17.

109. *See generally id.*

Warner were able to block other IM systems from access to their cable modem networks, AIM would have significant advantage. This was thought important because it was clear that as network speeds increased, IM systems would evolve into a series of services (such as video chat or file transfer) that would expand on the value of the existing systems.[110] The "names and presence directory" (NPD) was seen as being a critical infrastructure for IM services that precluded interoperability with other services.[111] AOL resisted efforts to publish clear protocol standards or allow interoperation between its NPD and other software, asserting concerns of "security" and "privacy" for its users.[112]

The FCC conditions for the AOL and Time Warner merger prohibited the use of new "advanced" videoconference extensions unless standardized server-to-server interoperability mechanisms were implemented.[113] Today, AIM is one of many protocols. Although AOL still has the largest number of users, IM has diminished in importance and multiple competing protocols and systems have emerged. Today, it would be fanciful to imagine that AIM adds $5.8 billion of value for AOL. What happened?

In large part, the efforts of AOL to block use of its services spurred development of competing services—this was apparent even at the time the merger conditions were being debated.[114] In addition to the MSN Messenger system, several "open source" efforts were developed to produce scalable messaging platforms with the most successful being Jabber, which produced the XMPP protocol.[115] These multiple implementations allowed companies to launch their own private and customized IM services because the cost of deploying the technology had been greatly reduced. People learned that adopting a new IM system was not hard. In part, the plurality of systems and the willingness to adopt new IM systems accelerated the use of IM and messaging systems for business applications. One of the complications of using AIM for business purposes was that AIM was often blamed for security lapses, and businesses had

110. *Id.* at 317–19.

111. *Id.*

112. AOL's concern about security and privacy was disingenuous given AOL's reliance on a "buffer overflow" attack to block competing services; that same attack could be used to compromise the customer's computer.

113. *See* Applications for Consent to the Transfer of Control of Licenses and Section 214 Authorizations by Time Warner Inc. and America Online, Inc., Transferors, to AOL Time Warner Inc., Transferee, *Memorandum Opinion and Order*, 16 F.C.C.R. 6547, para. 167 (2001) [hereinafter Time Warner & AOL Transfers].

114. *See* Jim Hu, *AOL's Lead in Instant Messaging Arena Dwindles*, CNETNEWS.COM (Nov. 16, 2000), http://news.cnet.com/AOLs-lead-in-instant-messaging-arena-dwindles/2100-1023_3-248700.html?tag=mncol;1n.

115. XMPP stands for Extensible Messaging and Presence Protocol. *About*, XMPP STANDARDS FOUND., http://xmpp.org/about-xmpp/ (last visited Feb. 21, 2010).

poor controls over the identity, security, privacy, and logging needed when applying AIM to business applications.[116] In particular, the Sarbanes-Oxley Act of 2002 and other reporting and disclosure rules, precipitated by various financial scandals, make it more important to keep accurate records and logs of communication between investors and financial advisors, as well as between people in the investment community. This led several companies to stop using public IM networks in favor of in-house networks.[117] Eventually, those IM systems used web browsers rather than requiring extra clients to be downloaded. The development of "Web 2.0" technologies such as Ajax changed the IM experience afforded by a browser interface so that it was equal to that of dedicated software. This allowed businesses to maintain control over "customer chat" and integrate those chat records with other "customer relationship management" (CRM) software that records customer names, account numbers, service and sales calls and all customer interactions.

The pace of technology adoption and the peculiar needs of companies seeking to employ IM systems means that although AOL's system is still the largest IM system, there has been no stranglehold on innovation or capabilities. The pace of this innovation was addressed in the FCC merger memorandum:

> Finally, it might be thought that in the rapidly changing technology of the Internet, even network effects and AOL's present position in the market would not prevent successful entry by IM providers other than AOL, that a new breakthrough technology might become available and would be superior enough to AOL's service to overcome the network effects flowing from its NPD, and cause users to shift *en masse* away from AOL. . . . We see no evidence at this time, however, of such a new breakthrough technology strong enough to overtake AOL's NPD.[118]

With the benefit of hindsight, we see that within two to four years after the merger orders were written, rich IM competition developed. Customers did not shift *en masse* away from AOL because they did not need to—they

116. The AIM instant message system had numerous security flaws that were used to block interoperability but could also be used by attackers against computer security. AIM also functioned by sending all data to AOL and, in later versions, that communication was encrypted, making it impossible to record the "plain text" version of the conversation. The Instant Messaging market split into "Enterprise" and "Consumer" instant messaging in 1998 with companies such as Lotus, Microsoft, and others providing solutions with features specifically for business uses. Wikipedia has a history and supplementary references on those developments. *Instant Messaging*, WIKIPEDIA, http://en.wikipedia.org/wiki/Instant_messaging (last visited Feb. 21, 2011).

117. *See, e.g.*, Thomas Hoffman, *Sarbanes-Oxley Trumps IM at Some Firms: Concerns About Security, Archiving Prompt Companies to Unplug Instant Messaging Systems*, COMPUTERWORLD (Aug. 8, 2005), http://www.computerworld.com/s/article/103752/Sarbanes_Oxley_trumps_IM_at_some_firms.

118. Time Warner & AOL Transfers, *supra* note 113, at para. 167.

simply used other technologies in concert with AIM.

Hindsight certainly helps in seeing trends, but some trends are apparent only when other technologies arise. One of the FCC's concerns with the AOL and Time Warner merger was that it might lead to a new dominant signaling and communication system by the introduction of new services over AIM.[119] This did not come to pass because alternate services became available (and were easy to adopt), mechanisms existed to work around restrictions, and open standards reduced the barrier for entry. The rapid evolution of technology was in contrast to most of the history of telecommunications, and this rapid evolution made it difficult to estimate the impact of regulation.

IV. MAINTAINING A VIBRANT INTERNET ECOSYSTEM

Technology on the Internet moves both more slowly and more quickly than most technology overseen by traditional regulation. VoIP technologies were in place almost a decade before they became widely adopted. Promising technologies such as AIM arose, peaked, and then diminished in value dramatically within that same period of time. The technology for one application was largely a substitute for the other,[120] but that was not clear at the time.

Regulation may not always be the best way to maintain a vibrant Internet. Standard methods for measuring what is actually happening in the Internet can help identify the root cause of complex service problems. Standard methods for reporting or disseminating policy to consumers in understandable terms can reduce confusion about services and performance guarantees. Existing agencies and policies can be used to maintain competition. Increased innovation and competition for access networks can provide consumers access to the competitive services in the Internet ecosystem. Lastly, developing "best practices" that can be clearly understood by network operators, regulators, and consumers will set "networking norms" that highlight the violation of those norms.

A. Measure and Report

Clearly identifying problems in the Internet and apportioning blame is very difficult. Consumers on access networks typically want answers to three questions: Can I access a specific service? Is the latency or quality of

119. *Id.* at para. 2.

120. The XMPP protocol used by Jabber and Google Talk has been extended as "Jingle" by Google to enable voice calls. *What Is Google Talk?*, GOOGLE CODE LABS, http://code.google.com/apis/talk/ (last visited Feb. 21, 2011). Similarly, voice systems such as Skype added support for basic and "enhanced" IM services. *Instant Messaging*, SKYPE, http://www.skype.com/intl/en/features/allfeatures/instant-messaging/ (last visited Feb. 21, 2011).

that service acceptable? Is there a bandwidth problem for a specific service?

Consumers often jump to conclusions when a service or site is blocked or unavailable. Services may be blocked by an ISP—or, the service may actually be down. Alternatively, parts of the Internet protocols not under control of the ISP (such as DNS) may misdirect traffic. In extreme cases, events halfway around the world may block services.[121]

The debate concerning network neutrality has prompted the development of several measurement tools to determine if application blocking or data modification is occurring. Examples include the "Switzerland" tool, developed by the Electronic Frontier Foundation;[122] the "Glasnost" tool, developed by The Max Planck Institute;[123] and the "Measurement Lab" consortium that supports both education and analysis tools.[124] These tools either detect specific problems (e.g. BitTorrent blocking) or identify factors that may delay communication. They are first steps in helping consumers identify what may be wrong and assisting in network monitoring. However, they are still primitive and require considerable sophistication to deploy and interpret.

It would be better for ISPs to be transparent about their network management policies and network conditions. Many ISPs block services that appear to arise from "malware"; sometimes those services are actual but uncommon services. For example, Scherlis notes that game developers often need to contact ISPs to remove blocked services that are mis-identified as malware.[125] At the same time, consumers are typically unaware when one of their home computers or devices is launching network attacks on others.

---

121. One example occurred in February 2008, when the government of Pakistan ordered access to YouTube to be blocked within Pakistan. *See* Danny McPherson, *Internet Routing Insecurity: Pakistan Nukes YouTube?*, ARBOR NETWORKS (Feb. 25, 2008), http://asert.arbornetworks.com/2008/02/internet-routing-insecuritypakistan-nukes-youtube/. The network operators for Pakistan Telecom implemented that order by issuing a "black hole route." *See id.* This is a method whereby a network router advertizes that it has an efficient route to the designated host but then actually discards that traffic. That "black hole route" was then published to other ISPs, causing a large part of the world to think that Pakistan had a very good connection to YouTube; this caused broad outages for YouTube. *See id.*

122. *See, e.g.*, *Switzerland Network Testing Tool*, ELECTRONIC FRONTIER FOUND., http://www.eff.org/testyourisp/switzerland (last visited Feb. 21, 2011).

123. Marcel Dischinger et al., *Glasnost: Enabling End Users to Detect Traffic Differentiation*, PROCEEDINGS OF NSDI '10: 7TH USENIX SYMPOSIUM ON NETWORKED SYSTEMS DESIGN AND IMPLEMENTATION 405 (2010).

124. Measurement labs arose from an effort by a number of companies, university faculty, and Internet researchers to determine technical approaches to measuring the network access characteristics. *About Measurement Lab*, M-LAB, http://www.measurementlab.net/content/about-measurement-lab (last visited Feb. 21, 2011).

125. Comments of Scherlis, *supra* note 71.

What is missing is a mechanism or protocol for communicating current management and policy information to consumers. Developing standards or protocols for informing customers about "suspicious" traffic would remove much of the confusion when an application stops working. There are existing protocols, such as Simple Network Management Protocol (SNMP) and Remote Network Monitoring (RMON), designed to communicate network performance, but these protocols are designed for network management rather than consumer enlightenment—they provide too much detail for consumers and provide no insight into what steps can be taken to correct problems. Through efforts such as the P4P consortium,[126] ISPs have found that it is possible to work with applications to reduce bandwidth demands and costs. Similar tools for communicating with consumers would likely improve customer service and help reduce network security problems. Efforts to inform consumers about broadband capabilities would allow broadband providers to compete based on those different services without consumers complaining about hidden differences. The British regulator, Ofcom, has established a voluntary "Code of Practice" for ISPs that communicates much of this information to consumers prior to sale and during service.[127]

B. Maintain Competitive Applications, Content, and Services

Content distribution and cloud computing services dramatically reduce the infrastructure cost for computing and web applications, allowing noncommercial groups to rapidly scale their efforts. Software innovations and business models that can exploit these new platforms are enabling even more rapid innovation. Vertical integration in these markets may or may not lead to anticompetitive behavior; however, these technologies are so new that it is not clear whether they will remain in their current form or if concerns about fair competition will last longer than the technology itself. The possibility of antitrust enforcement from the FTC and the Justice Department will foster more innovation than enacting preemptive and broad rules to regulate these hybrid "private/public" networks.

Predicting the future of technology is difficult, as evidenced by the analyses of the predicted outcome of the competition surrounding AOL Instant Messaging. That regulation was eclipsed by the reality of rapid technology development, external technology, and changes in business practice and usage patterns. Although there is certain to be consolidation in

126. P4P is a reporting method that allows P2P software to learn the "topology" of ISPs, allowing the P2P software to avoid expensive or congested links. *The P4P Working Group*, PANDO NETWORKS, http://www.pandonetworks.com/p4p (last visited Feb. 21, 2011).

127. *Voluntary Code of Practice: Broadband Speeds*, OFCOM, http://www.ofcom.org.uk/telecoms/ioi/copbb/copbb/ (last visited Feb. 21, 2011).

the "cloud computing" ecosystem, it remains to be seen whether the consolidation will foster anticompetitive behaviors.

C. *Maintain Competitive Networks with Transparency and Clarity*

Business networks (primarily Ethernet) have many mechanisms to improve flexibility, control performance, and diagnose problems. Consumer access network technology is only beginning to see similar development, and there is a real risk that regulation will curtail investment in or development of those technologies. At the same time, certain services benefit from separation from general best-effort traffic—this is why many businesses use different "virtual private networks" to separate different kinds of traffic. As home users expand the range of services they use, consumers may be better served by technologies that enable multiple network services, each with different qualities.

Likewise, innovations in congestion control will continue and can be implemented in many parts of the networks. Researchers are exploring the tension between enforcing congestion control at the end-points (such as a laptop or cell phone, where it may take years to upgrade or replace all the software) versus upgrading specific routers or other parts of the network. Precluding implementation at the access network will simply increase the costs of network management. Rather than exclude specific mechanisms such as congestion control, regulation should be used to foster goals such as competition.

D. *Keep Ahead of the Technology*

The Internet is complex, encompassing both traditional communication services as well as computer systems, novel services, and rapidly evolving technology. Developing an ongoing process for discussing and analyzing the interplay between the different technologies is critical. There are specific actions that can foster more thoughtful review, such as creating an organization to provide independent and informed counsel to policy makers about the Internet ecosystem as a whole. This is a difficult charge because some emerging trends are not apparent until they are established practices. The other action is to counter specific concerns that have been indicated by prior regulators and develop standards or tools to mitigate those concerns.

There are many bodies that examine and discuss how Internet technology should be developed; other groups discuss business practices, and yet others research new techniques or services. It is equally important to have a continued and informed discussion about how technology, business, and new services affect future policy so that policy makers can stay ahead of the technology. It is useful to guide technology before it is

widely deployed because that lessens the cost of regulation.

Two such examples are the "network effects" of systems such as instant messaging, and the "stickiness" of specific e-mail addresses. As an example, although there have been calls for "e-mail portability," there has been little serious study of the concept. However, "identity" on the Internet is one of the key features that makes network effects important. Although AIM was not the only messaging tool available, moving to another system entailed rediscovering the online identity of your friends. Now, when instant messaging has been replaced with social networking, the same issues that were raised about AIM "stickiness" may be raised about Facebook or MySpace. Here, the technical community is moving faster than the regulatory world—there have long been Internet standards, such as DNS, for "machine portability," and now there are developing standards, such as OpenID,[128] for "people portability." Such identity systems could have significant impact when widely adopted, but it is also important to understand and clarify how such systems will interact with regulation.

V. REGULATION SHOULD BE A PROCESS, NOT A PRODUCT

This Article has argued that regulation or legislation that simply affects control of the access work policies while ignoring the impact of the rest of the Internet ecosystem is a disservice to consumers. At the same time, regulation or legislation that affects the *entire* Internet is overreaching and also not needed.

To date, most of the network neutrality discussion has been heavily influenced by existing telecommunications regulation—this is natural since most regulation seeks to model new systems after old. This has led regulators to focus on "bits in flight"—for example, the regulation of access networks—while largely ignoring the "bits at rest"—content distribution networks—that make up much of the Internet. That distinction between basic and information services is rapidly being challenged by the development of an integrated Internet ecosystem. Focusing on "bits in flight" also impacts the ability of regulators (or even technology pundits) to predict the evolution of services. This Article highlighted the example of AOL Instant Messenger, arguing that the comparison between AIM and the existing communications systems missed the rapidity with which new and competing systems could be *deployed* using the existing infrastructure. Standardization and open software and protocols also meant that the cost of *developing* a new system was radically reduced compared to existing telecommunications systems. The rapid evolution of the Internet makes it difficult to ensure that regulation is still meaningful by the time it is

128. *What Is OpenID?*, OPENID, http://openid.net/get-an-openid/what-is-openid/ (last visited Feb. 21, 2011).

developed.

True network neutrality is about competition and innovation, and any such discussion must involve the full Internet ecosystem. It is clear that narrowly defined rules affecting one part of that ecosystem are not the best solution to maintaining a competitive and responsive Internet. Existing legislation—primarily antitrust laws in the case of browsers and the threat of similar laws in advertising-based search—are being applied and should be able to address future anticompetitive actions. At the same time, consumers would benefit from competition, innovation, and better information about the services available to them.

Wireless Efficiency Versus Net Neutrality

Charles L. Jackson*

---

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

I. INTRODUCTION

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

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

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

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

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

II. CONGESTION IN THE INTERNET

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

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

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

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

would otherwise be.

A. Controlling Internet Congestion

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

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

1. Internet Congestion Control on the Honor System

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

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

---

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

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

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

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

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

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

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

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

---

superhighway.html (emphasis added).

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

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

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

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

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

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

---

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

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

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

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

12. *Id.* at 9.

13. *Id*.

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

15. RFC 2914 states:

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

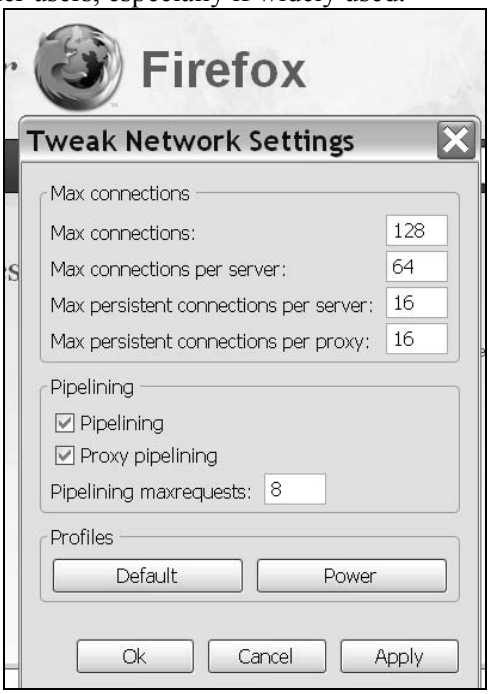


Figure 1. Firefox network control panel showing a maximum of 16 persistent connections rather than the RFC 2616 maximum of 2.[16]

The Internet community is well aware of the congestion risk created

---

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

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

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

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

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

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

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

---

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

18. *About FireFox's Connection*, PINGUY'S WEBSITE, http://pinguy.infogami.com/blog/39l5 (last visited Feb. 21, 2011). Other blogs also suggest tuning Firefox to increase performance, but do not explain the negative consequences for others. *See* Sandip Dedhia, *21 About:Config Hacks(Tweaks) for Firefox 3*, BLOGSDNA (June 22, 2008), http://www.blogsdna.com/372/21-aboutconfig-hackstweaks-for-firefox-3.htm; Serdar Yegulalp, *Hacking Firefox: The Secrets of About:Config*, COMPUTERWORLD (May 29, 2007, 12:00 PM), http://www.computerworld.com/action/article.do?command=viewArticleBasic&taxonomyName=Networking+and+Internet&articleId=9020880&taxonomyId=16&pageNumber=5; Damien Oh, *28 Coolest Firefox About:Config Tricks*, MAKETECHEASIER (Aug. 21, 2008), http://maketecheasier.com/28-coolest-firefox-aboutconfig-tricks/2008/08/21. The help page for the Opera browser states, "It is recommended to keep the default setting of 16 [maximum connections to a server], but you can try changing the maximum number of connections to a single server if you are experiencing problems with browsing speed." *Advanced Preferences: Network*, OPERA HELP, http://help.opera.com/Windows/10.63/en/network.html (last visited Feb. 21, 2011).

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

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

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

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

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

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

---

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

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

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

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

2. More Recent System Collapses

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

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

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

---

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

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

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

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

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

---

25. Plonka, *supra* note 24.

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

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

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

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

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

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

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

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

---

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

30. *Id.*

31. *Id.*

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

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

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

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

36. Baldor, *supra* note 35.

37. *Id.*

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

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

3. Use of Established Congestion-Avoidance Technologies

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

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

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

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

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

4. Security

Adoption of the proposals mandating undifferentiated treatment of

---

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

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

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

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

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

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

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

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

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

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

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

---

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

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

46. *Id.*

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

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

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

B. Impacts of Eliminating ISPs' Congestion Control Tools

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

---

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

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

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

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

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

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

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

III. WIRELESS NETWORKS AND NETWORK NEUTRALITY

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

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

A. Priority Routing Expands Capacity

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

---

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

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

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

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

decreases cancel and total power from the base stays even.

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

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

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

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

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

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

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

---

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

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

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

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

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

B. Priority in the Backhaul Network

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

1. Separation of Control Signaling and User Information

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

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

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

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

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

---

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

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

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

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

2. Converged Networks

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

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

3. Network Neutrality and Backhaul Networks

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

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

---

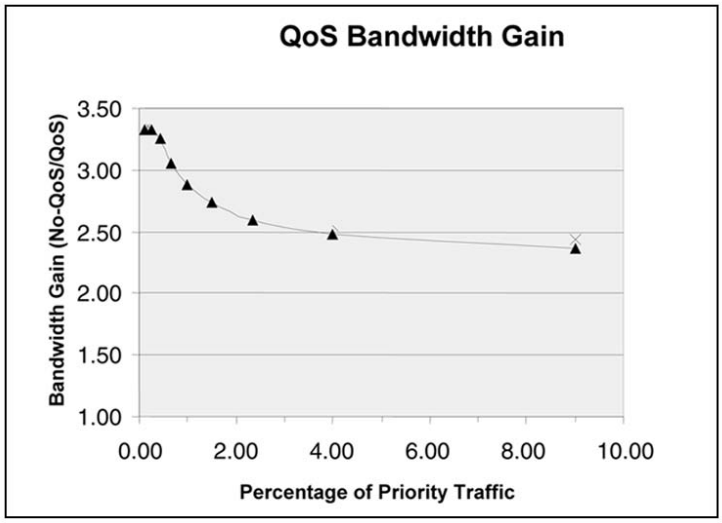
62. *Id.*

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

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

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

Figure 2. Capacity benefits of priority routing



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

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

higher-priority and lower-priority applications.

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

C. Cross-Layer Design

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

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

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

---

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

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

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

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

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

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

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

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

---

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

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

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

D. Efficiency

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

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

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

---

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

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

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

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

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

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

E. Handset Attributes and System Capacity

1. Receiver Sensitivity

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

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

74. *See* 47 C.F.R. § 20.18(b).

75. *See* 47 C.F.R. § 20.18(e)–(j).

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

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

---

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

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

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

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

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

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

2. Vocoder Performance

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

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

---

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

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

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

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

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

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

3. Other Handset Attributes That Affect System Capacity

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

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

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

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

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

4. Handset Attributes and Service Quality

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

5. Poor Handsets or Poor Networks?

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

---

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

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

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

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

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

6. Network Standards Evolution

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

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

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

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

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

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

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

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

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

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

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

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

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

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

V. CONCLUSION

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

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

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

---

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

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

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

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

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

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

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

---

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

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

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

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

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

---

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

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

Spectrum Miscreants, Vigilantes, and Kangaroo Courts: The Return of the Wireless Wars

Christian Sandvig*

---

* Christian Sandvig is associate professor of Communication at the University of Illinois at Urbana-Champaign. This material is based on work supported by the National Science Foundation under Grants No. 0308269 and 0546409. Portions of this material were supported by funding from the Higher Education Funding Council for England, and part of this research occurred during a visiting fellowship at the Oxford Internet Institute at Oxford University. An earlier portion of this Article concerning jamming as competition was presented to the symposium, "The Economics, Technology and Policy of Unlicensed Spectrum" by the kind invitation of Michigan State University. An earlier portion of this Paper concerning the visualization of wireless was presented to the 33rd Telecommunications Policy Research Conference (TPRC) on Communication, Information, and Internet Policy in Arlington, Virginia. Remarks from this Paper were also given at the kind invitation of the Center for Technology, Innovation, and Competition at the University of Pennsylvania Law School as part of a conference titled "Rough Consensus and Running Code: Integrating Engineering Principles into the Internet Policy Debates." The Author would like to thank Rivka Daar, Lindsay Hinkle, and Katie Kuppler for research assistance essential to this project.

The requirement that radio users obtain the government's permission in advance before transmitting has been a foundational feature of communications regulation for about eighty years. However, the recent regulatory expansion of "open" regimes for managing the electromagnetic spectrum, such as the increase in license-exempt and "light" licensed frequencies in several countries, may change all of that, and this prospect has created excitement among observers of wireless telecommunications and communication law. Garage door openers, cordless phones, and baby monitors, it is hoped, were just the first kinds of "radio stations" one could have without a license. Under open regimes, more people will have more wireless devices in their hands than ever before, and they will be able to use them in new ways. Proponents hope that more use, more efficient use, and more application innovation will result. However, the fate of services in these bands—and of the open spectrum model itself—now rests with user behavior. As of this writing, no one is sure of the answers to basic questions such as when (or if) these open bands of the electromagnetic spectrum will become congested with too many users, if they will fail due to congestion, or, more generally, what it is exactly that people will do with these new wireless freedoms. While allusions to "tragedies of the commons"[1] and their inevitability or avoidability have been widespread in writing about license-exempt spectrum, little is empirically known.

In effect, license-exempt bands are a partial return to communication policy's "state of nature"—what will people do without government? (Or, more properly, what will people do when the role of government is changed and the requirement for prior permission to transmit is removed?) Using two case studies drawn from a larger project across six countries, this Article considers the case of Wireless Internet Service Providers trying to use "open" spectrum, and chronicles their successes and failures. It shows, perhaps unsurprisingly, that when legal constraints are removed, users make their own order and are bound by their own local and differing standards of fairness and propriety. The topic of this Article could be identified by the keyword "*shared* spectrum," used in the literature—but in

1. Garrett Hardin, *The Tragedy of the Commons*, 162 SCI. 1243 (1968).

what follows it is clear that sharing sits alongside selfishness, coexistence with extortion, and formal law with kinship and neighborhood customs.

First, this Article will outline the theoretical approach embodied by these observations, an approach grounded in the anthropology of law and derived from Moore's process theory of law:[2] here, glibly labeled "telecommunications policy from below." Second, it will introduce Wireless Internet Service Providers (WISPs) and this project's methods in studying them. Next, it will present two detailed case studies from 2003 to 2005. The first case study describes an entrepreneurial project in a small city that never quite got off the ground because the spectrum never looked empty enough, while the second focuses on a "war" between two competing WISPs that evokes the world before the enforcement of radio regulations—the "Wild West" of radio, as some have called it. Finally, this Article will end by drawing conclusions about the future of open spectrum regimes and the utility of studying the administrative law processes of telecommunications with a process theory of law, or "telecommunications policy from below."

I. TELECOMMUNICATIONS POLICY FROM BELOW: THE THEORY OF LAW AS PROCESS

If one is interested in the study of telecommunications policy, one almost always assumes that the action can be found in bodies one would identify as "policymaking" (legislatures and regulators), and that the appropriate object of study is a law or ruling—or, more expansively, an elite debate about one. Certainly the world outside these fora is crucial to research on telecommunications policy, but the outside world makes its entry via logical arguments in legal analysis, in descriptions of technological changes, in ideal hypothetical cases, in secondary reporting of market research, and in economic simulations of reason. Research on telecommunications policy is in this way dominated by a philosophically conservative approach to law—an approach encouraged structurally by the political economy of the policymaking process.[3]

If inherent in all of these diverse approaches is an overarching philosophy of law, the closest may be legal formalism. A policy researcher never need mention that "the law" of interest is the law as it exists written on a page or that the appropriate focus of a research project should be a patriciate debate over a present or future policy. The analyst's goal is usually to determine how a current or proposed law (on a page) is right or wrong. Research in this tradition has produced useful and even brilliant

2. SALLY FALK MOORE, LAW AS PROCESS: AN ANTHROPOLOGICAL APPROACH (1978).

3. *See, e.g.*, THOMAS STREETER, SELLING THE AIR: A CRITIQUE OF THE POLICY OF COMMERCIAL BROADCASTING IN THE UNITED STATES 113–162 (1996).

work. However, although the experts toiling with such research are more savvy than any layman about the intrigues and interpretations that surround every line of formal law, these intrigues rarely appear in mainstream analyses of telecommunications policy, and they are almost never the focus of it.

In contrast, this Article takes a methodological approach derived from the anthropology of law, specifically Moore's theory of law as process.[4] This approach can be contrasted with other research on telecommunications policy by two central differences: First, it considers the law as it appears in the lives of people who are not policymakers. These lives occasionally appear in formalist telecommunication "user studies,"[5] but it is significant that what is *analyzed* in other studies is the communication technology (as in "telephone users" or "Internet users") and not the law. Telephone users are studied so that the right law can be written. The right law will then define the system that these telephone users use. "Yet although everyone acknowledges that the enforceable rules stated and restated in legal institutions, in legislatures, courts and administrative agencies, also have a place in ordinary social life, that normal locus is where they are least studied."[6] That is, telephone and other kinds of users are also users of the law. To care about this is not just to go looking for the same law in a different place, but instead it is looking for a different law. That is, not the telecommunications policy that is written, but the policy that you can get away with. It is well known that only a rare few will ever encounter formal legal proceedings, and even these laws oftentimes become operative only when a certain kind of person claims to know about it and presses for its enforcement.[7] And so, telecommunications policy then is not just responsible for systems and markets, it is also something that users (or rather, people) directly think about, interpret, manipulate, and even create in the course of their experiences with communication systems. This framework's second departure from other approaches to telecommunications policy then is that it takes as its object the intrigues and interpretations that surround law. Indeed, without formalist law as an object, this approach asserts that the surrounding impermanent perceptions are in fact the substance of the law. The rules are whatever we believe the

---

4. MOORE, *supra* note 2.

5. *See, e.g.*, Christian Sandvig, *Public Internet Access for Young Children in the Inner City: Evidence to Inform Access Subsidy and Content Regulation*, 19 INFO. SOC'Y 171 (2003) (discussing previous user studies by this Author); Milton L. Mueller & Jorge Reina Schement, *Universal Service from the Bottom Up: A Study of Telephone Penetration in Camden, New Jersey*, 12 INFO. SOC'Y 273 (1996).

6. MOORE, s*upra* note 2, at 55 (citing Paul Bohannan, *The Differing Realms of the Law*, 67 AM. ANTHROPOLOGIST 33 (1965)) (internal citation omitted).

7. *See* MOORE, s*upra* note 2, at 79.

rules are, no matter what the law books say. Statutes are one way of talking about rules and order, among many other ways.

Many precedents for this approach exist. Although Moore's theory of law as process has not been employed in telecommunications, excellent previous scholarship in communications policy has focused critically on the political culture surrounding law. Perhaps most memorably, some scholars have considered administrative agencies like the FCC as an interpretive community, and have analyzed communications policy symbolically rather than institutionally.[8] There have also been approaches to law that are methodologically similar to this one via oral history.[9] Previous scholars have rejected legal formalism by turning to the critical legal studies movement,[10] but here we will instead turn to socio-legal studies[11]—a pluralistic scheme for studying the law that is inclusive of legal anthropology.[12]

Indeed, it may be clearest to say at the outset that telecommunications policy has always been ruled "from below" as much as from above. Midwestern farmers in the first decades of the twentieth century were running illegal telephone systems over barbed-wire fences and using their farm kitchen as the exchange.[13] In the 1960s, ordinary people with no technical experience were using then-illegal network attachments in their homes.[14] Commercial broadcasting was brought to the United Kingdom in part by Radio Caroline and other commercial broadcasters intentionally testing and even flaunting broadcast rules.[15] Yet aside from some discussion of radio pirates,[16] empirical analyses of these minor telecommunications criminals do not typically appear in law journals.

---

8. STREETER, *supra* note 3, at 114–16.

9. *See, e.g.*, Robert B. Horwitz, *Broadcast Reform Revisited: Reverend Everett C. Parker and the "Standing" Case (*Office of Communication of the United Church of Christ v. Federal Communications Commission*)*, 2 COMM. REV. 311, 313 (1997).

10. *See, e.g.*, Thomas Streeter, *Beyond Freedom of Speech and the Public Interest: The Relevance of Critical Legal Studies to Communications Policy*, 40 J. COMM. 43 (1990).

11. *See, e.g.*, D.J. Galligan, *Introduction*, 22 J. L. SOC'Y 1 (1995).

12. *See, e.g.*, Peter Just, Review Essay, *History, Power, Ideology, and Culture: Current Directions in the Anthropology of Law*, 26 L. & SOC'Y REV. 373 (1992); *see also* Sally Falk Moore, *Certainties Undone: Fifty Turbulent Years of Legal Anthropology, 1949–1999*, 7 J. ROYAL ANTHROPOLOGICAL INST. 95 (2001).

13. CLAUDE S. FISCHER, AMERICA CALLING: A SOCIAL HISTORY OF THE TELEPHONE TO 1940, at 43 (1994).

14. *See, e.g.*, Use of the Carterfone Device in Message Toll-Telephone Service, *Decision*, 13 F.C.C.2d 420 (1968).

15. Douglas A. Boyd, *Pirate Radio in Britain: A Programming Alternative*, 36 J. COMM. 83, 86 (1986).

16. *See, e.g.*, Buck Endemann, Comment, *Keelhauling Pirates: How Ex Parte Seizure of Non-Interfering LPFM Does Not Further the FCC's "Public Interest*,*"* 43 SAN DIEGO L. REV. 661, 692–97 (2006).

It is obvious that changing the penal code's sanction for (or definition of) assault will not eliminate assault. Similarly, radio laws will always have radio pirates. Indeed, manipulating the formal criminal law may define illegality and change rules, penalties, and their enforcement, but all of this may have little relation to what happens in your neighborhood. The same is true for telecommunications policy, as this Article will show.

II. RETURNING WIRELESS TO ITS "STATE OF NATURE"

This Article considers the promise and viability of open spectrum regimes by investigating how a few interlocutors make order in the electromagnetic spectrum. The process theory of law would hold that invoking law in a social situation is a way to symbolically communicate, establish, maintain, or undermine order against a background assumption of absolute indeterminacy.[17] This way of thinking about law may have seemed unnecessary to discussion of the electromagnetic spectrum until quite recently. After all, it seems that order among users of the spectrum has been solidly achieved by government regulation of the time, power output, location, and frequency to be used by radio transmissions. Users of the spectrum may have seemed like only second- or third-hand users of the law, as their awareness of national spectrum allocation rules might have been limited to the concept of a "channel" when turning the knob on an old television set. Laws about spectrum allocation were fixed both in law books and in tuned crystals, and there might have seemed to be little that communicators could do to interpret or even interact with them.

But as alluded to in the Introduction, spectrum users are now being expected to interact with each other and the spectrum in novel ways. Unlicensed spectrum "parks"—like the U.S. allocation of 2.435–2.465 GHz[18] (most commonly used for "Wi-Fi" wireless data networking, cordless phones, and other unlicensed uses)—confine all users to a narrow slice of spectrum and dictate that users must accept whatever interference results. Some scholars have advanced the prospect that the liberalization of spectrum should continue such that most or all spectrum is open.[19]

The closest historical analog to this situation may be radio broadcasting circa 1920 in the United States. At this time, although there was something called a "license," there was no governmental enforcement authority. All users had a limited choice of frequencies (all commercial stations were at 360 meters or, after 1922, at either 360 or 400 meters),

---

17. MOORE, *supra* note 2, at 49.

18. 47 C.F.R. §§ 15.245–248.

19. *See, e.g.*, Yochai Benkler, *Some Economics of Wireless Communications*, 16 HARV. J. L. & TECH. 25, 29 (2002) (providing a review and proposal of the liberalization of the entire spectrum).

faced power limits, and had to accept any interference that resulted.[20] At first, within the overall framework of shared licenses, stations that encountered interference made "simple agreements" and "handshake pacts" with each other to reduce interference.[21] Specifically, within the conditions for transmission specified by the government, individual stations haggled over the more limited options left open to them by federal law, creating new norms, formal agreements, and informal agreements at the local level that were within or even superseded the federal rules.[22] Stations arranged to manipulate the times they transmitted (e.g., making voluntary frequency-sharing schedules), the locations of their transmitters (e.g., dividing up the transmission area amongst themselves), their power within the limits specified by the government, and their frequencies within what the government allowed.[23] The U.S. Department of Commerce sponsored conferences to encourage this kind of self-regulation from 1922 to 1925.[24] However, before long, "owners of stations who believed themselves to be interfered with took matters in their own hands," leading "eventually to a warlike atmosphere"[25] and ultimately the breakdown of order into chaos.

Although it is difficult to clearly see such a distant past, a common interpretation among radio historians is that at some point, after the local arrangements began to fail, all systems of order in the spectrum failed. Stations "jumped without restraint to new wave lengths. . . [and] also jumped their power"[26] even beyond federal limits. "By the end of 1926 it was impossible in most geographical areas to receive a consistent broadcast signal" due to interference between stations.[27] This crisis is the genesis story of modern telecommunications regulatory agencies, which are often described as born to bring order from this chaos.[28] (Although this is a received view of the creation of such agencies, that view has attracted scholarly criticism.[29]) The solution by 1934 was a much more rigid

---

20. *See* SUSAN J. DOUGLAS, INVENTING AMERICAN BROADCASTING 1899–1922, at 316 (1987).

21. GEORGE H. DOUGLAS, THE EARLY DAYS OF RADIO BROADCASTING 92–93 (1987).

22. *See id.*

23. *Id*.

24. *Id.* at 93.

25. *Id.*

26. 1927 FED. RADIO COMM'N. ANN. REP. 10–11, http://www.fcc.gov/Bureaus/Mass_Media/Databases/documents_collection/270701.pdf.

27. DOUGLAS, *supra* note 21, at 95.

28. *See, e.g.*, CHRISTOPHER H. STERLING & JOHN M. KITTROSS, STAY TUNED: A HISTORY OF AMERICAN BROADCASTING 141–47 (2002).

29. *See* Robert Horvitz, *Marconi's Legacy: National Sovereignty Claims in Radio*, 1ST COMMUNIA WORKSHOP ON "TECHNOLOGY AND THE PUBLIC DOMAIN" (2008), http://www.communia-project.eu/communiafiles/ws01p_Marconis%20Legacy%20National%20Sovereignty%20Claims%20in%20Radio.pdf.

allocation of the electromagnetic spectrum that largely eliminated shared use of bands except among radio amateurs—centralizing the administration of the spectrum and creating the FCC and its foreign counterparts.[30]

Today's unlicensed electronic consumer devices might appear to be nothing like the radios of the 1920s. Yet the recent experience of users of shared frequencies (now often called "open spectrum," "unlicensed," or "license-exempt" regimes) show the striking parallels between 1927 and today. Chiefly, the experience so far shows the widespread reappearance of multiple interlocking and overlapping systems of rules derived from a wide variety of sources of authority—federal administrative law, municipal authorities, private mediators, bilateral contracts, friendships, rivalries, family ties, and neighborhood norms. Multiple fields of rulemaking have always existed, but they are now more salient.[31] As federal policy has thrown the spectrum open to "individual" firms and actors, it is clearer than usual that this is the stuff from which individual action is knit. Proposals for open spectrum now pose a question. Were the "broadcast wars" of the 1920s the interregnum between two regimes of order: the garden and the license? Or was instead all of licensing (1934 to 2005) the interregnum between two periods of open wireless: the broadcast wars of then and today?

A. To the Trenches of License-Exempt Spectrum

To consider the uses of more open spectrum, this Article turns to the specific instance of wireless Internet. One of the most significant developments in the use of license-exempt spectrum from 2000 to 2005 has been the rapid emergence of a Wireless Internet Service Provider (WISP) industry based on license-exempt spectrum—a frequent estimate is that there are 3,000 commercial WISPs in the United States alone.[32]

---

30. *See* Communications Act of 1934, ch. 652, 48 Stat. 1064 (codified as amended at scattered sections of 47 U.S.C.) (2006) (establishing the FCC).

31. In some contexts, the word "fields" automatically evokes Foucault. The connection intended here is legal anthropology's semi-autonomous social field. *See, e.g.*, Sally Falk Moore, *Law and Social Change: The Semi-Autonomous Social Field as an Appropriate Subject of Study*, 7 L. & SOC'Y REV. 719 (1973).

32. *See, e.g.*, Marlon K. Schafer, *Mandatory FCC Form 477 Info*, WIRELESS INTERNET SERVICE PROVIDERS ASS'N MAILING LIST (Feb. 2, 2006), http://www.mail-archive.com/wireless@wispa.org/msg03551.html; *FCC Form 477 Forum*, BROADBAND DSLREPORTS.COM, http://www.broadbandreports.com/forum/r13791564-FCC-Form-477 (last visited Feb. 23, 2011). During this time period the Wireless Internet Service Provider Association had 400 members in the United States. About 500 WISPs filed the FCC's Form 477, leading to concerns about underreporting. For a discussion of reporting, see Kristopher Twomey, *FCC Form 477*, ISP-PLANET (Mar. 6, 2006), http://www.isp-planet.com/fixed_wireless/politics/2006/form_477.html. WISPA then conducted a survey of sales by equipment manufacturers and subsequently estimated the total number of WISPs to be about 3,000, a figure that was then accepted in the government and has been often

Sophisticated users of the electromagnetic spectrum from this industry are in an interesting position. They are frequently well-educated, technically trained engineers within an established white-collar profession. If it is possible to speak so broadly about the engineer's disposition toward the law, it could be said that a person from this background is likely to be exceedingly conscientious about following it, even though when the rules about radiocommunication are strict, the engineer's own skills provide ample means to bend or break them.

As an introduction to this industry's experience of the law, consider Tim Pozar, an engineer with "traditional" radio experience and the founding member of the Bay Area Wireless Users Group of San Francisco, California.[33] In 2002, Pozar wrote the definitive legal guide for the emerging community of Wi-Fi engineers, and he began it with the sentence, "I am not a lawyer."[34] The guide, titled "Regulations Affecting 802.11 Deployment," started as a presentation transcript posted to Pozar's personal website and then became so well known that it was eventually included in a popular wireless reference book as an appendix.[35] A point worthy of note about Pozar's presentation is that he approached radio regulation like a quest: the law in this area was a distant thing that an engineer would be unlikely to know but quite likely to break.[36] In response, Pozar conscientiously interviewed telecommunications lawyers and read law books until he could produce a very comprehensive list of the relevant sources of authority at different levels and agencies of government and also their specific rules. His presentation included explanations of limits on the height of towers, rules against swapping antennas between different equipment manufacturers, a discussion of the risk of wireless networks interfering with aircraft RADAR, the limit of a wireless worker's maximum exposure to electromagnetic fields, and much more.[37] He concludes by recommending "[c]oordination with other users"[38] in unlicensed bands, and it is this coordination that provides local order without licenses.

---

repeated. *See, e.g.*, John M. R. Kneuer, Deputy Assistant Sec'y for Comm. and Info., U.S. Dep't of Commerce, Speech on Promoting U.S. Broadband Deployment and Economic Growth at the Mississppi Technology Alliance Sixth Annual Conference on High Technology (Nov. 9, 2005).

33. *About*, BAY AREA WIRELESS USERS GROUP, http://www.bawug.org/about/ (last visited Feb. 23, 2011).

34. Tim Pozar, *Regulations Affecting 802.11 Deployment* (Mar. 10, 2004), http://www.lns.com/papers/part15/Regulations_Affecting_802_11.pdf.

35. ROB FLICKENGER, BUILDING WIRELESS COMMUNITY NETWORKS, app. A, at 137–56 (2003).

36. *See generally id.*

37. *See generally id.*

38. *Id.* at 156.

The case studies that appear below also reflect encounters with this distant law. These cases arose from a comparative cross-national study of WISPs in six countries: the United States, Canada, Ireland, the United Kingdom, New Zealand, and Australia. For the larger project, sixty-three groups were chosen that identified themselves as WISPs in 2003 and were affiliated with the "open wireless," "community wireless," "municipal wireless," or similar movements.[39] This includes groups that call themselves commercial, noncommercial, and governmental. The groups range from the quite formal to the quite informal. To be included, the group had to have a web presence in 2003. Extensive, ongoing participant observation was conducted with two of these groups (one in the United States, one in the United Kingdom), while members of an additional sixteen groups have so far been visited at least once by researchers who interviewed participants and organizers in an open-ended format. Researchers also attended meetings (if possible) and received a tour or demonstration of the network (if relevant). All groups (including the remaining forty-five) were analyzed by quantitative and qualitative content analysis of online material about them (often including extensive mailing list archives). This larger project is ongoing and the case studies presented below represent early results from the visits to those sixteen groups that are worthy of considerable attention on their own. Names of the people, places, and organizations involved in the following case studies have been changed.

Methodologically, this approach could be characterized as interviews with users of electromagnetic spectrum regulation, or more formally, what Yin would term a holistic, multiple case study research design organized around literal replication.[40] This Article will present material from two cases, but other cases from this overall study have appeared elsewhere,[41] and the research methods have been described in more detail elsewhere.[42]

III. CASE 1: MONROEMESH'S FAILURE TO SHARE

Monroe is a small city of about 210,000 people in the Midwestern United States, and is a county seat.[43] The main industries there are white-

39. Or in some cases, the groups were referenced on web pages about those terms.

40. ROBERT K. YIN, CASE STUDY RESEARCH: DESIGN AND METHODS (3d ed. 2003). In Yin's terms, both case studies presented here represent failures of order (literal replication) that were arrived at in different ways. *See id.*

41. *See* Christian Sandvig, *An Initial Assessment of Cooperative Action in Wi-Fi Networking*, 28 TELECOMM. POL'Y 579 (2004); Christian Sandvig, *Wireless Play and Unexpected Innovation*, *in* DIGITAL YOUTH, INNOVATION, AND THE UNEXPECTED 77 (Tara McPherson ed., 2008).

42. Christian Sandvig, *How Technical Is Technology Research?*, *in* RESEARCH METHODS FROM THE TRENCHES 141 (E. Hargittai ed., 2009).

43. In the 2000 U.S. Census, Monroe's per capita income in 1999 dollars was about

collar services (notably health, finance, and insurance), retail trade, government, and education. (Monroe is known for its local university and very high levels of education.) These industries are slowly displacing a historical focus on light manufacturing in a period of economic growth that has been continuing for over ten years. Monroe is set among very slight hills. At the time the proto-WISP group that is the subject of this case study was founded, high-speed wired broadband Internet service was already widely available via cable modem and DSL, and prices were falling. This case study is based on interviews with Terry and Dave in 2003 and 2004. Terry and Dave are both white, well-educated men in their twenties who decided to found a WISP in Monroe—optimistically named MonroeMesh.[44]

Terry is a local engineer with a passion for tinkering with wireless equipment. Before founding MonroeMesh, Terry gained previous experience with a rare, proprietary unlicensed wireless data technology called RLAN.[45] In the late 1990s, using twelve radios scavenged from a friend's failed electronic coupon printing business, Terry built an RLAN network in a small town near Monroe as a hobby. The network wirelessly connected his friends to the Internet (via an ISDN line) and to a shared file server. The connection served most anywhere within an area of about fifteen square miles. Terry and friends then had mobile Internet connectivity at a time when this was so rare as to be almost unknown.[46] "We had an old Sparc 2 sitting at [a friend's] house with a 22-gig SCSI hanging off it—that was our central depository. We had mp3s, video, whatever we want," Terry said. "It was pretty kick ass."

This earlier technology, RLAN, used shared spectrum, but the other users were encountered so rarely that the fact of spectrum sharing was not especially obvious, and the high power of the RLAN radios gave an extra feeling of security. "RLAN was so powerful that, if you were close enough,

---

$23,000. There are 3,000 people per square mile. 7.6 percent of the population has no secondary school diploma.

44. The network did not use what is called "mesh" technology—an advanced technique—but Terry and Dave had hoped to. In a typical wireless network there are dedicated routers to relay traffic and end nodes that originate it in a configuration like the hub and spoke of a bicycle wheel. In a home Wi-Fi network, for example, a dedicated device called an "access point" acts as the hub (and as a router), while the connected devices (laptops, game consoles, etc.) communicate only with the access point and not with each other. In the more complicated mesh network configuration that was just emerging at this time, every device could acts as a relay for other nearby devices and there may be no central routers.

45. RLAN was a proprietary first-generation wireless product operated at 900 MHz. It was developed by DCA, a company that was later acquired by Attachmate. RLAN was then discontinued.

46. Metricom Ricochet, the only widely known mobile Internet company at that time, did not serve Monroe.

it would overload the front end on whatever tuner you had . . . it was a very robust, very good solution," Terry explained. When it ever became noticeable that other users did exist and there was a technological skirmish between systems, the skirmish was fine with Terry because his radios always won. Once, in the manner of a confidential aside, he mentioned, "We walked over the analog cell phones. You heard a clicking sound in your earpiece if you were using a cell phone anywhere in town." He paused, then continued, "I thought that was pretty cool."

The introduction of the first cheap "Wi-Fi" wireless Internet consumer products in 1999 spurred new wireless networking projects across the world.[47] At the same time that Tim Pozar was getting excited about Wi-Fi in San Francisco, Terry and Dave met in the city of Monroe and formed a working group that included four other technically inclined people interested in capitalizing on the new possibilities of unlicensed spectrum. The initial goals of the hobbyists were carried over from the earlier RLAN network: "I wanted to be able to drink beer and have my PDA get me alerts from work," explained Terry. Dave added, "I wanted Internet access out at the lake."

They soon conceived of a city-wide transport network that a variety of local service providers could connect to and jointly pay for. For instance, the local radio station could join and then use MonroeMesh's transport network to link remote broadcasts (such as DJs or live music events) with the studio. This was already being done with 900 MHz ISM band equipment, but the radio station reported that because there were so many other users on that band in town "that first leg was horrible—they could rarely get more than voice grade communication." Dave and Terry saw the possibility for higher-quality digital sound broadcasts if they moved over to Wi-Fi. Additionally, a local Internet service provider (ISP) agreed to connect to the network and provide a connection to the Internet with an authentication scheme—the MonroeMesh city-wide network would then be a way to connect to the local ISP from a laptop while outside (and without dialing in or using any wires).

As a beginning, using their own money, the group built three nodes on tall buildings in Monroe (including one at the radio station), and the limited network functioned just as it was intended to. But as they planned for the network's expansion, Terry and Dave had more and more misgivings. Terry explained, "We did some surveying. In one instance, we

47. *See* François Bar & Hernan Galperin, *Building the Wireless Internet Infrastructure: From Cordless Ethernet Archipelagos to Wireless Grids*, 54 COMM. & STRATEGIES 45, 52 (2004). "Wi-Fi" is a brand and not an acronym—as coined, it did not stand for anything. It was meant to evoke the "Hi-Fi" or "high fidelity" of audio equipment. It is an industry consortium's name for interoperating radios that comply with the IEEE 802.11 standards. *See id.*

were on top of [a particular] building, just looking around, and we found ten, fifteen networks. All just hammering away. All just blasting away, making noise. And in the same area! That's an issue."

Later along in the widespread adoption of Wi-Fi these numbers (ten or fifteen) seem small and would not be a concern, but this was 2002. The Wi-Fi equipment in question (an access point) usually sends out a signal called a beacon even when it has no traffic. This "hammering away" almost surely consisted of beacons detected by the site survey software Terry and Dave were using. Packet data traffic like Wi-Fi (and all Internet traffic) is notoriously "bursty"—seeing a few other users on the same band would not mean the band was full of traffic, simply that others had set their equipment to use that frequency if and when they had something to transmit. Both Dave and Terry were well aware of the technical details of the communications protocol; nonetheless, the electromagnetic spectrum felt full to them, and the presence of fifteen other possible users was "an issue."

As another way to facilitate the sharing of spectrum, the Wi-Fi communication standard calls for the band to be further divided into overlapping channels in a way similar to older cordless phones—if a cordless phone user heard static on one channel, the user could move a switch on the phone to transfer to a different frequency. The same held true for Wi-Fi.[48] But Terry explained that regardless of the number of channels there were, someone else might still be using them:

> If I were to use any of the channels that are available to me, one to eleven let's say. No matter what I've picked, I'm asking for loss. There's no technical way I can avoid loss with the gear that I'm given. Or I can get. Anything that we can find is going to fall down at some point. And even though these [other] networks are not necessarily very active, they're still producing traffic.

The notion of overlapping channels itself has been problematic for the engineering community, as traffic on an adjacent or even nearby channel implies some (though not complete) degradation in service quality. One comment on a trade press article that appeared around this time stated forcefully: "The 802.11b standard gives us 14 channels to work with, right? Wrong! Sorry, it's really only three useful ones."[49]

The MonroeMesh group's hesitation in this case does not seem to be

---

48. For instance, the IEEE 802.11b standard defines fourteen channels; one through eleven are used in the U.S., one through thirteen in most of Europe, and only channel fourteen in Japan. Jim Geier, *Assigning 802.11b Access Point Channels*, WI-FI PLANET (Feb. 11, 2002), http://www.wi-fiplanet.com/tutorials/article.php/972261.

49. This quotation comes from a comment that has since been removed. It was originally posted on WI-FI PLANET. *Id.* For example, these channels would be one, six, or eleven in the United States. The channels in 802.11b center on frequencies in 5 MHz steps, but a transmission is 22 MHz wide, meaning that transmissions on adjacent channels overlap significantly. *See id.*

about responding to interference, but sharing—the only sharing arrangement acceptable to them would have been one without any possibility of degradation whatsoever Dave explained:

> You're sitting at a table with two network engineers that would rather build nothing than to build something that doesn't work the way it should. Because the user expects the network to be a utility. When they turn on the light switch, the lights come on. When they call 911, the ambulance is at their door. It's the same thing.

More critically, it seems that the notion of increasing the use of license-exempt spectrum creates an irresolvable tension for some wireless engineers—while municipal, community, and entrepreneurial groups like this one sometimes state that they aim to "democratize" access to the spectrum, their engineers hate the thought of sharing spectrum (or worse, losing a spectrum war) to an unskilled, uncredentialed consumer who purchased equipment at a local superstore and knows nothing about radio. While this group and others state that they hope to rely on cheap consumer equipment to reduce prices, their professional identity rebels at the idea of using consumer-grade equipment—Terry and Dave noted archly that "it makes no sense" to "use commodity equipment in an infrastructure role." This tension between credentialed engineers and amateurs has existed since the beginning of the idea of the spectrum.[50]

After their initial burst of activity, MonroeMesh experienced several frustrations trying to negotiate for the placement of their radios on tower sites. Terry and Dave were also astonished by the effect of the weather in the Midwestern United States on their outdoor equipment. Next they became disappointed at the limited reach and few features of the Wi-Fi equipment when compared to Terry's more powerful RLAN radios of the past. MonroeMesh ran out of steam and the group dissolved in 2004, with three stations and twenty users, and without formally incorporating or taking in any money. "Maybe twenty users wasn't enough to legitimate [sic] me donating gear and doing all this work," Terry said. "I'm sorry for being so cynical." Still, both Dave and Terry listed Wi-Fi experience on their resumes, and both quickly moved on to higher-paid jobs working on wireless systems—Terry with an out-of-state telephone company[51] looking to move into wireless; Dave in the IT department of a large organization.

A. The Engineer's Perception of Congestion and Beauty

The MonroeMesh case reinforces an important lesson about technology—the need to pay due attention to the way things look as well as

---

50. *See* CAROLYN MARVIN, WHEN OLD TECHNOLOGIES WERE NEW: THINKING ABOUT ELECTRIC COMMUNICATION IN THE LATE NINETEENTH CENTURY 40 (1990).

51. Actually, it was a competitive local exchange carrier (CLEC).

the way they work. Dave and Terry's experience of the electromagnetic spectrum came to them from the user interface of their mapping software.[52] The popular software Netstumbler, for example, presents the user with a list of detected networks where each new station identified is added to the list as it is identified.[53] No measure of the amount of traffic is shown. No measure of the amount of traffic is shown. This means that Netstumbler's screen could show fifteen networks nearby, yet they might all be silent. Dave and Terry could see the spectrum as though it were "full" because fifteen networks are listed on the screen. While national spectrum regulators and the Wi-Fi protocol designers would see that same spectrum as empty, they would be looking at it through a different portal. Functionally, Netstumbler was open spectrum's user interface.

More significantly, when we look at the MonroeMesh case in order to understand the many overlapping obligations governing Terry and Dave's behavior, it is clear that their professionalization as engineers is the controlling one. While this may not be so for Terry, for other engineers who are now coming to wireless systems with a background in computer software rather than in radio, the uncertainties of the radio environment are traumatic. Terry and Dave did not want MonroeMesh to work as much as they wanted it to be beautiful to engineers, and this couldn't be accomplished within their other constraints. This suggests that a significant obstacle for proponents of open spectrum may be the cultural connotations of sharing and the socialization of engineers.

IV. CASE 2: THE PLANETREE FOREST SPECTRUM WAR

The next case study includes threats of litigation and (at the time of interview) an ongoing government investigation. To permit the parties involved to speak at all about these events (especially because the electromagnetic spectrum enforcement community is so small, even across six countries), this Article will not reveal which country of the six (United States, Canada, United Kingdom, Ireland, Australia, New Zealand) is the home of "Planetree Forest." While the cultures and laws of the six nations vary, the actual national law in this case makes surprisingly little difference to what happens in Planetree Forest.

This case study concerns the relationship between two WISPs, here called TownNet and SATNet. The materials for this case study come from interviews with the two cofounders of TownNet (Alan and Philip), a

52. For a further discussion of mapping, see Christian Sandvig, *The RED Project: Rendering Electromagnetic Distributions*, VECTORS (Fall 2007), http://vectors.usc.edu/projects/index.php?project=87.

53. NETSTUMBLER.COM, http://www.netstumbler.com/2007/04/17/about/ (last visited Feb. 23, 2011).

private mediator working at a not-for-profit organization who was called in to adjudicate the following dispute, and two government officials from the national communications regulator who were in a position to be familiar with the regulations relevant to the dispute.[54] The lack of SATNet interviews make the picture of events unfortunately one-sided, but the two available parties (a SATNet employee and the SATNet founder) both declined requests for interviews—probably for reasons that will become clearer below. Public information about the dispute was also consulted. All descriptions here refer to the time of the interviews: the middle of 2004. As most of the information comes from TownNet interviews, the dispute will be told from the perspective of Alan and Philip, the cofounders.

Planetree Forest has a lower unemployment rate than surrounding areas and higher levels of education.[55] The main industries are farming and the light manufacture of furniture, precision machinery, and clothing, and the landscape is marked by farms, river valleys, and fifty-four small towns (the population of the largest is 3,000). At the time the two groups described here began operations, only dial-up Internet access was available.

TownNet was the creation of Alan and Philip, two white, well-educated professionals. Philip was a former telecommunications engineer, and Alan was a manager. Both lived in Planetree Forest. They gathered about six other local professionals—including two accountants, a marketing manager, a property developer, someone from the municipal government's IT department, and a telecommunications market researcher—who were dissatisfied that no broadband Internet service was available. They wrote a business plan that projected that they could build a sustainable (break-even) service using the latest wireless technology with about USD $50,000 for ten towns in Planetree Forest. At the time, large telecommunications carriers had publicly claimed that it was not profitable enough to deploy broadband service in Planetree Forest. In response to a national policy to accelerate the deployment of broadband in rural areas, government subsidies were available from several agencies at different levels of government. Alan and Philip of TownNet received a zero-interest loan of about USD $5,000 and quit their jobs. They subsequently received an additional $50,000 in the form of a state-level development grant and expanded the network to include sixteen towns. However, TownNet was not the only group interested in using new wireless technology to bring high-speed Internet to Planetree Forest.

---

54. To protect the confidentiality of the TownNet founders, the government officials were not told that the interviews were in reference to the specific dispute.

55. The area ranks in the eighty-ninth percentile for per capita income when ranked against other areas in the country. There are 348 people per square mile. 21.9 percent of the population has no secondary school diploma.

A. SATNet and the Informal Spectrum Negotiation

Another company, SATNet, operated on a for-profit basis and was run from a town outside the area. SATNet's founder had previous experience in information technology, and SATNet had been providing Internet service to hotels when they saw an opportunity to provide wireless Internet service in Planetree Forest. Both SATNet and TownNet offered very roughly comparable wireless Internet service,[56] and both designed their network to use license-exempt bands for many necessary links (although both also used other bands). Note the neighborhood relationships involved in their first contact, which was about potential interference, described here by Alan:

> What [SATNet] basically said to us was, "could you change the channel please?" But they didn't ask us, they told their customer who happened to be our customer's landlord. Who then told his brother-in-law, who was our customer. It was the brother-in-law who finally introduced both parties.

Surprisingly, it turned out that SATNet was also receiving a substantial government subsidy to provide service to Planetree Forest in competition with TownNet. SATNet may have received as much as about USD $30,000 from a different agency at a different level of government (from a fund for the promotion of local businesses).

To reconcile this competing use of public funds, the municipal government asked both parties to come in for a voluntary meeting. Alan reflected that:

> At first [SATNet] tried to sell us equipment, a client device they had built. [We didn't buy.] We initially signed a [non-disclosure agreement] with them. We were going to work together. We came up with this idea of sharing the spectrum. We said we'd use only a given channel. We came up with a reasonable plan and they seemed happy at the meeting.

A government official from the agency that gave TownNet the bulk of their funding asked for a second meeting, intended to be a "technical meeting" between the engineering staff of both groups. The second meeting went badly, as Alan explained:

> [SATNet] said, "We were here first, tough." Their stated objective was to close the space down so that no one else could move in. Part of the agreement was that they'd provide us with a list of where we could use what channel and we would provide a list of what areas we had covered. That agreement has sat in abeyance. They [didn't] do anything and we haven't changed anything.

---

56. SATNet provided a speed of 512 Kbps for about USD $24.99 per month, and served eight towns in Planetree Forest. TownNet offered comparable service to sixteen towns (including eight towns also served by SATNet), with a speed of 512 Kbps for about USD $29.99 per month and 1 Mbps for $82.99 per month.

Worried, Alan of TownNet then posted to a mailing list for coordination between WISPs. His post read in part:

> I think we're going to have a problem with spectrum issues. Does anybody have any advice on arbitration over use of channels? These people won't negotiate—it's likely to end up in court.

SATNet forwarded Alan's post to TownNet's funding agency with the addendum: "This guy's a troublemaker."

B. From Negotiation to Jamming

By this point, TownNet and SATNet had an antagonistic relationship, to be sure, but antagonism should not necessarily be worrying. They are competitors requiring the same resource (part of the 2.4 GHz band) that they both saw as scarce. Next, according to Philip, SATNet's strategy for winning changed to enforcement of a first-come, first-served model of the band, and the instrument used changed from negotiation to technology. Philip elaborated:

> When we moved one [TownNet] link to 300 yards and crossed two of their long links we found that we couldn't do anything. We stick up an antenna and do Netstumbler and get a long list of [SATNet] out there. [Before], we were picking channels that were well separated, the noise floor appeared nice and quiet, and [now] at a matter of a few hundred yards with line of sight we couldn't see a thing. There was no signal, nothing. . . . Then customers started to complain that their own [indoor] home networks stopped working. . . . So [the regulator] in the end started some sort of investigation.

This mysterious failure of all of the open spectrum to be open occurred just after a number of relevant developments in the larger Wi-Fi engineering community.

First, widespread publicity appeared about wireless Internet's newly discovered susceptibility to "logical" jamming. Briefly, digital wireless systems use a "listen before talking" procedure to reduce the chance that a transmission will collide with one from another station. For example, the equipment used by TownNet and SATNet employ Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) that includes a process called "clear channel assessment." When a wireless card performs a clear channel assessment as a prelude to "talking," if another station is transmitting, the card will wait a "backoff" interval and then perform the assessment again.[57] Researchers noticed that directing a wireless card to transmit a continuous pattern of bits will cause all other devices within range to always conclude that the channel is busy, and wait indefinitely.[58]

---

57. Chris Wullems et al., *A Trivial Denial of Service Attack on IEEE 802.11 Direct Sequence Spread Spectrum Wireless LANs*, *in* WIRELESS TELECOMMUNICATIONS SYMPOSIUM 129, 131 (2004).

58. *Vulnerability Note VU#106678: IEEE 802.11 Wireless Network Protocol DSSS*

Similarly, simple scripts appeared on the World Wide Web at this time that allowed a wireless device to "use" all of its available capacity by transmitting nonsense. Rumors were also flying in the wireless community about new equipment proposed by manufacturers that would increase the speed of data transmission by using all of the available channels simultaneously across multiple bands, using multiple radio chipsets.[59] Finally, online discussion boards reported that commercial telecommunications carriers had begun to raise the transmission power on their equipment to nearer the legal maximum,[60] presumably to drown out competing signals. It is not clear what exactly happened to the open spectrum in Planetree Forest in 2004, but Alan is convinced that:

> They were over power. [SATNet] was using amplifiers. There are a number of technologies that cause denial of service that are actually very difficult to pinpoint and I'm . . . convinced that they were using something. They really didn't like competing with anybody. [They thought,] "The more channels I grab means the less competition."

C. From Jamming to Extortion

At this point, the municipal government asked TownNet and SATNet to return for a third of what Alan called (with a chuckle) "these arbitration-confrontation meetings." This time they also invited an outside mediator, and Alan said the SATNet tactic moved from jamming to extortion:

> We agreed that [SATNet] would let us use channel one and they'd use the rest. He went away with this agreement but he had this list of [other] demands still in place. He wanted us to pay for reconfiguring their network. [He said] we were going to pay them [thousands of dollars] to implement the changes that were necessary.

Even after the agreement, nothing changed immediately, until a few days later. Alan said, "[O]n the day the regulator knocked on their door, that all stopped and suddenly our customer[s' equipment] burst into life." Alan and Philip explained that *both* SATNet and TownNet were found to be using Taiwanese equipment that was not certified for operation in the country, and had secondary harmonics in a licensed band (thereby violating

---

CCA Algorithm Vulnerable to Denial of Service, U.S. COMPUTER EMERGENCY READINESS TEAM, DEP'T. OF HOMELAND SEC., http://www.kb.cert.org/vuls/id/106678 (last visited Feb. 23, 2011); *see* Wullems, *supra* note 57, at 131.

59. For instance, rumors were that Sony's proposed "Hi-Bit Wireless" strategy for consumer electronics would involve multiple simultaneous Wi-Fi channels (and therefore multiple radios) in the same device. *See, e.g.*, *Sony Air Board in March*, AKIBALIVE (Jan. 19, 2004), http://web.archive.org/web/20080105114258/http://www.akibalive.com/archives/000514.html.

60. *See, e.g.*, John Foust, *2Wire and SBC Interference?*, BAY AREA WIRELESS USERS GROUP MAILING LIST (Jan. 20, 2004), http://www.mail-archive.com/wireless@lists.bawug.org/msg05848.html; John Foust, *Re: SBC Routinely Installing 2Wire 400 mW AP/FW*, BAY AREA WIRELESS USERS GROUP MAILING LIST (Apr. 21, 2004), http://www.mail-archive.com/wireless@lists.bawug.org/msg06334.html.

certification laws and causing harmful interference in an easily detectible way). While charges of jamming—especially by denial of service using random or nonsense traffic—would be almost impossible to prove, luckily the fact that they both bought very cheap uncertified equipment over the Internet gave a national regulatory official the reason to inspect the premises, and all interference then disappeared. Both TownNet and SATNet have continued to compete, but with no interference. Both providers were warned about the uncertified equipment and stopped using it. There were no formal legal sanctions.

D. Primacy of Local Versus National Sources of Adjudication

Local ties interconnect all aspects of TownNet's story. A local government employee was one of the founding members, and they were able to secure free access to antenna locations and some development money (at least the initial zero-interest loan) in part through existing relationships in the community where they lived. But one relationship not yet discussed is essential to understanding this case, and that is the one between the TownNet founders and the official who worked for the national communications regulator. The official also lived in Planetree Forest and wanted broadband in his community; after meeting the founders at an early public organizing meeting about broadband, they became friendly.

"He's very professional. But he does keep us informed," Alan noted. Philip seconded, "He has access to spectrum analyzers and all those kinds of things; we used to regularly bring him in because getting a hold of that stuff is expensive."

In another context, Alan explained, "He compartmentalizes his advice as well. Sometimes it was a formal warning." Philip added, "He still gives us quite a lot of technical help."

With this relationship in the foreground, the fact that the regulatory official, "in the end started some sort of investigation," shifts in meaning. At first glance, Alan and Philip successfully and justly defended themselves against a variety of assaults—requests to change frequencies, demands for outrageous payments, jamming of the airwaves. Their successful defense "in the end" required the mobilization of central state authority, which looms as a final arbiter after more local systems of order fail in adjudication. At second glance, Alan and Philip are well-connected local experts who are already enmeshed in the apparatus of the state at a variety of levels: through grants, their own board of directors, and their acquaintance with a federal enforcement officer. Their competitor is "run from out of town rather than here," "has just come from the outside," and is "notorious" because of earlier suggestions of dishonesty. He "has a

reputation," and is "not in it for the community." His character can be judged by poor engineering decisions (his network is bridged), poor results (his network is high latency), and a lack of manual skill (he cannot do his own crimping).[61] Rephrased, his network is simple, slow, and he cannot connect two wires together. Led by a nontechnical profiteer and outsider, SATNet had already been "selected for failure,"[62] and all that remained was to decide the venue for the defeat and choose its justification. The outcome was decided by these measures of SATNet's character, and the federal enforcement action was not a last resort when local measures failed, but was itself an implementation of the local decision. That is, SATNet was first and foremost *not* a violator of national spectrum allocation and certification rules through the use of uncertified equipment. Instead, SATNet was first and foremost (in Alan's words), "a bad neighbor." Even though both networks were warned, only one network was triumphant and TownNet was delighted with the result.

The Planetree Forest case does not force a choice between two competing conceptions of regulation—it is not true that this incident must either indicate that the federal government is the *resolution* for intractable local disputes, or that it is the *expression* of local decisions about character. Both of these can be true, even within the same case. To cement which of these is more in play in this particular incident, let us close this case study with Philip's observation about the role of government in the regulation of the spectrum. Reflecting on the successful (for them) conclusion of the TownNet versus SATNet confrontation, Philip stated, "Part of the problem with a lot of regulators is that they're too heavy with the regulation. If you're being a bad neighbor, someone needs to officially remind you that you're being bad. No more than that."

Philip is not asking for more central control of the spectrum because the force of SATNet's comeuppance was not derived from central authority—enforcement was here a reminder of who could mobilize federal enforcement on behalf of *local* norms and attach distant officialdom to their victory. Of course, the quotation also restates the common notion that more

---

61. Bridged networks do not employ routing. Crimping is the process of connecting wires and electrical connectors by deforming them with a tool called a crimper. Latency refers to the time that elapses between a request for data and the beginning of data transfer.

62. "Selection for failure" is a concept introduced to legal anthropology by Moore to explain the action of multiple cultural dimensions that underpins some legal reasoning. Specifically, selection for failure has come to mean the process by which a party is culturally prejudged to lose or win in adjudication. The role of the law in these situations is then to externally rationalize and justify a decision arrived at much earlier, rather than to make any new decisions. *See* Sally Falk Moore, *Selection for Failure in a Small Social Field: Ritual Concord and Fraternal Strife Among the Chagga, Kilimanjaro, 1968–69*, *in* SYMBOL AND POLITICS IN COMMUNAL IDEOLOGY 109 (Sally Falk Moore & Barbara G. Myerhoff eds., 1975).

regulation is always bad. That is, regulators can never win, even when they can win your broadcast war.

V. CONCLUSION

It is worth remembering that "new laws are thrust upon going social arrangements in which there are complexes of binding obligations already in existence. . . . The social arrangements are often effectively stronger than the new laws."[63] There is no doubt that large national telecommunications carriers have complexes of binding obligations and ongoing social arrangements, but there are not that many of them. In counterpoint, a more open electromagnetic spectrum policy and the appearance of the Jeffersonian ideal of free competition between decentralized small enterprises like these WISPs have combined to produce an eruption of thousands of local spectrum confrontations where users with a wide variety of backgrounds and skill levels wrestle with new wireless technology, spectrum laws, and each other.

In these cases, we have seen that the engineer's allegiance to principles of engineering as a profession can be far stronger than any allegiance to communications regulation, and that even as engineers they have an allegiance to *their* system that is much stronger than any to *the* system. Similarly, it could be said that TownNet was the injured party in the Planetree Forest spectrum war—at least if jamming occurred—but relationships with members of the Planetree Forest community and a national regulatory official were critical in bringing the war to a resolution, whereas the facts of the dispute were not so critical. (Recall that SATNet was not charged with jamming anything.) Engineers in the field and regulators in the government disagreed as to whether a portion of unlicensed spectrum was empty or full, with all of the engineers in these two cases seeking unopposed access to a non-overlapping channel. In addition, these users of telecommunications laws "knew" the spectrum by both reading the law and using a free software program (Netstumbler).[64] The software program and some of the particularities of its design were very influential in seeing the spectrum as "full," in effect becoming the electromagnetic spectrum's user interface.

A. The New Role of the License-Exempt Regulator

This Article opened by posing questions about the future of open spectrum. No evidence found so far in these two cases or in the larger

63. MOORE, *supra* note 2, at 58.

64. While MonroeMesh only used Netstumbler in this case, TownNet "knew" the spectrum both through Netstumbler and by carefully monitoring the performance characteristics of its wireless network.

research project suggests that open spectrum will fail due to a tragedy of the commons. In several instances described here, engineers have been extremely overcautious in predicting the failure of unlicensed spectrum due to congestion. While this potential congestion has been something of an obsession, it has yet to appear. These engineers, like the MonroeMesh group, were sometimes unwilling to develop unlicensed systems because of a misapprehension that the mere existence of other users (inferred via 802.11b beacons) implied actual or imminent congestion. In one instance described—the alleged Planetree Forest jamming that occurred over a period of weeks in 2004—the spectrum appeared to be "actually" full, in that TownNet's equipment would not work in one area of Planetree Forest. However, this was an instance of aggression, not overgrazing (or congestion)—a critical difference.

Practically, this empirical evidence suggests that widespread use of wireless without licenses intimates a new kind of relationship with the electromagnetic spectrum. Regulators have never conceptualized the spectrum as something that has a user interface, and telecommunications policy research has never particularly focused on messy local situations. Nonetheless, the larger study of WISPs in six countries finds that whenever you have a local WISP, you have a local situation. Examples abound of local spectrum coordination and negotiation. These negotiations have occasionally turned hostile, as negotiations do, but this does not indicate a failure of negotiation. It is this image of spectrum use that regulators will find increasingly useful. Within the enforcement bureaus of the world, it is common to think that determining a source of unlicensed interference is a probably impossible task, but this is true only if the entity trying to do so is an enforcement bureau. Local providers have an intimate knowledge of what goes on on the rooftops of their town—and even what goes on inside. Mechanisms like searchable public databases in bands where registration is required would also aid local coordination, and WISPs have tried to set up their own databases for this purpose where registration is not required. A more clearly defined and promulgated set of unlicensed norms would also be a benefit. If under "open spectrum" models the government is not going to go away, this indicates a much increased role for enforcement (of, for example, certification, certification violations, and jamming) and public education (usually only weakly attempted by national regulators). This obviously implies more work: the unlicensed regulator is entering into this new relationship with a much larger number of local users.

B. Embedding Spectrum Negotiation in Software Will Not Change This Situation

It is tempting to conclude that the kinds of choices made by the

groups discussed in this Article result only from temporary lags in wireless technology. An argument in this vein might say that if in the future the negotiations for available spectrum (e.g., the choice of channel) are fixed in software algorithms, these local wireless providers will have nothing to talk about, and "broadcast wars" will go away. (Or, they will go away again.) Benkler hopes for as much by framing the problem of interference as one for product manufacturers to sort out in a world without carriers.[65] However, for a variety of reasons beyond the scope of this Article, one may also argue that carriers are unlikely to vanish. In that case, removing a decision from human agency to software algorithm simply changes the tools available to carriers—it is no balm for their desire to both operate in license-exempt bands and at the same time have some control over their operating situation. The promise of new configurability and user-driven innovation[66] is one of the reasons that unlicensed is attractive in the first place. Restricting the parameters that users can manipulate to build new services via some prior specification of algorithms (or requiring secrecy that discourages users from writing their own algorithms) runs exactly counter to the overall project of unlicensed.[67] Indeed, it precludes the innovation these new bands are supposed to produce.

C. The Future of the Approach "From Below"

This may be a historical moment when the study of telecommunications policy critically needs to pay attention to the law as it is lived. Unlike some other domains of law, communications policymakers and researchers have often assumed that the law is never particularly "operative" (Moore's term) in the lives of everyday people. Arcane communications rules are written for a small audience of industry insiders. In a magazine parody, a political writer noted that this culture in U.S. telecommunications policy was so insular it should be called "FCC World," and that it has "perhaps five thousand denizens."[68] Just after that article appeared in Spring 2003, that sort of thinking was jolted by an FCC vote on the relaxation of media ownership caps.[69] While the topic of media ownership caps is usually considered arcane, a record-breaking two million e-mails, telephone calls, and faxes poured into the FCC about it.[70] The

65. *See generally* Benkler, *supra* note 19.

66. *See generally* ERIC VON HIPPEL, DEMOCRATIZING INNOVATION (2005).

67. *See generally* Christian Sandvig et al., Hidden Interfaces to "Ownerless Networks," Presented to the 32nd Conference on Communication, Information, and Internet Policy (2004) (unpublished paper), http://research.niftyc.org/Hidden_Interfaces.pdf.

68. Nicholas Lemann, *The Chairman*, NEW YORKER, Oct. 7, 2002.

69. For a summary, see Press Release, FCC, FCC Sets Limits on Media Concentration (June 2, 2003), http://fjallfoss.fcc.gov/edocs_public/attachmatch/DOC-235047A1.pdf.

70. *See, e.g.*, Comments in FCC Media Bureau Docket 02-277, 2002 Biennial

surprise for many analysts was that it was possible at all for two million people to become aware that administrative regulation on media ownership existed in the first place. This increasingly active public, recent consumer protection efforts, and liberalization of the electromagnetic spectrum are all reinforcing trends. For instance, consider the new availability of wireless devices that do not require licenses in many countries; the new regime of low power FM licenses from the FCC; the use of a non-exclusive "light-license" of £1 per year for some wireless broadband services (with a simple online form) by the U.K. Office of Communications;[71] and the creation of an online National Do Not Call Registry by the U.S. Federal Trade Commission.[72] These are examples of situations in which much larger numbers of people are now expected to be aware of and interact with an administrative regulator in fairly unprecedented ways. The public is more active in arcane venues (where it perhaps is not welcomed by insiders), while at the same time new federal decisions (presumably endorsed by insiders) about things like telephone privacy now presuppose the participation of every citizen of the country in an administrative regulation. In this environment, we need more attention to the law as it is lived, and users of telecommunications as users of telecommunications law. In this, theoretical frameworks like Moore's theory of law as process are valuable tools.[73]

While parts of this Article have described events that are somewhat sensational by the standards of telecommunications policy, the future of "open spectrum" remains bright. In that spirit, the Article will close with one final return to the city of Monroe and to Planetree Forest. While these case studies describe a variety of different actors and events, there are many things wireless engineers will always agree on. In fact, when asked to describe the biggest problems facing wireless communication, no one mentioned congestion or regulation. Dave of MonroeMesh sighed and replied wistfully, "If only we got another tall building." A year later in

---

Regulatory Review–Review of the Commission's Broadcast Ownership Rules and Other Rules Adopted Pursuant to Section of the Telecommunications Act of 1996; Comments in FCC Media Bureau Docket 01-235, Cross Ownership of Broadcast Stations and Ownership; Comments in FCC Media Bureau Docket 01-317, Rules and Policies Concerning Multiple Ownership of Radio Broadcast Stations in Local Markets; Comments in FCC Media Bureau Docket 00-244, Definition of Radio Markets; *see also* Lemann, *supra* note 68.

71. This applies to fixed stations in 5 GHz Band C (5725 to 5850 MHz). "Light licensing" generally refers to the ability to obtain a non-exclusive license, pay a nominal licensing fee, or receive automatic license approval (also called "registration"), or all of these. *See Section 1*, OFCOM, http://stakeholders.ofcom.org.uk/market-data-research/telecoms-research/bbresearch/wireless_update/wirelessbroadband/section1 (last visited Feb. 23, 2011).

72. Rules and Regulations Implementing the Telephone Consumer Protection Act of 1991, *Report and Order*, 18 F.C.C.R. 14014, para. 28 (2003).

73. *See generally* MOORE, *supra* note 2.

Planetree Forest, the reply was also quick: The biggest problem? "It's the trees."

Access to Media All A-Twitter: Revisiting *Gertz* and the Access to Media Test in the Age of Social Networking

Ann E. O'Connor*

"[T]he individual's right to the protection of his own good name reflects no more than our basic concept of the essential dignity and worth of every human being—a concept at the root of any decent system of ordered liberty."[1]

---

* J.D. Candidate, Indiana University Maurer School of Law, May 2011; B.A. in Journalism, *magna cum laude*, The George Washington University, 2006. The Author would like to thank Professor Fred Cate and the entire staff of the *Federal Communications Law Journal* for their input and invaluable assistance in the completion of this Note. In addition, the Author would like to thank Joe and Cinda O'Connor and Collin McCready for their tireless patience and support.

1. Gertz v. Robert Welch, Inc., 418 U.S. 323, 341 (1974) (citing Rosenblatt v. Baer, 383 U.S. 75, 92 (1966) (Stewart, J., concurring)) (internal quotation omitted).

I. INTRODUCTION

Since the introduction of the actual malice requirement for public figures in defamation cases,[2] the test employed by courts to distinguish those public figures from private individuals has frequently included an inquiry as to the level of access to media the plaintiff enjoys. This determination has been one part of a multifactor test used to establish whether the plaintiff is in fact a public figure who then must prove actual malice in order to be successful with a defamation claim. Once the plaintiff is found to be a public figure by way of this test, the burden on the plaintiff is significantly higher—making the likelihood of success much lower. Because of the resulting difficulty for the public figure plaintiff, it is important that the test in place appropriately measures the plaintiff's role within the controversy and in the public eye.

The definition of what comprises the media has changed in recent years—blogs are no longer at the periphery of the media world, but have found a place within mainstream media as a source and as a tool. The line has further blurred with more widely accessible and user-friendly services that allow users to share with an Internet audience at large; with the advent of such social networking tools as Facebook, YouTube, and Twitter, it has grown easier for anyone and everyone to access the media in one way or another. With the current media landscape such as it is—political candidates announcing their plans to run for office via Twitter and Facebook,[3] widely followed print columnists employing blogs in their daily research, corporate America using YouTube videos to reach a wider advertising audience[4]—it is time to reconsider what exactly "access to media" means. Without such a reconsideration, the access to media factor in the public figure test in defamation law is outdated; furthermore, without appropriate reconsideration in the context of technological advances, this

2. N.Y. Times Co. v. Sullivan, 376 U.S. 254, 279–80 (1964).

3. Russell Lissau, *Candidates Like This. Following Obama's Lead, Hopefuls Embrace the Internet*, CHI. DAILY HERALD, Jan. 2, 2010, at 1.

4. *See, e.g.*, *Gatorade Mission Control*, YOUTUBE, http://www.youtube.com/watch?v=InrOvEE2v38 (last visited Feb. 23, 2011) (highlighting the corporation's use of social networking to better access its marketing audience).

test may lead to inaccurate conclusions as to who is a public figure, based on judicial confusion as to what access means.

This Note will present the history of the public-private distinction, beginning in Part II with the Supreme Court's decision in *New York Times Co. v. Sullivan*, where the Court announced the test applicable for public officials in defamation law—requiring a heightened burden to prove a defamation case when a public official alleges defamation. This case began a series of decisions by the Court in which the test was further refined, and the class of people who were required to meet the "actual malice" standard of proof was both clarified and expanded—by the time the Court decided *Gertz v. Robert Welch, Inc.*, those who must prove actual malice included public figures. With *Gertz*, the Court attempted to set forth explicitly the appropriate test for determining whether or not a person alleging defamation is in fact a public figure and must therefore prove actual malice. Because of the added—and not insignificant—burden placed on plaintiffs who are found to be public figures, the Court established a test by which public figures may be proven as such. This required showing that first, she has either achieved pervasive fame or notoriety because of his position in society,[5] or that because of her role in the controversy at issue in the story, she is a public figure for purposes of coverage pertaining to that controversy. For the latter aspect of the test, the Court required either a showing that she had voluntarily thrust herself into the issue and taken on a position at its forefront, or that she had been involuntarily drawn into that issue.[6]

As one aspect of this determination, the Court instructed that an inquiry as to whether or not the plaintiff had access to the media to adequately redress the claims made against her should be employed.[7] For this prong of the test, the Court concluded that an individual of prominence would have ways to access the media and therefore to address the public. It left the test at that, without delving into the adequacy required of that response, nor the mode or medium of access that would satisfy the requirement.

This Note will then go on in Part III to give an overview of how that access test has been applied by lower courts, and the results lower courts have come up with when grappling with what exactly access to media means. There is not a clear consensus across all jurisdictions as to the importance of this prong of the test, nor as to what exactly is required to find that access to media is present in a particular context. Indeed, it does not even seem clear what constitutes "media" for the purpose of showing

5. *See infra* note 35 and accompanying text.
6. *See infra* note 37 and accompanying text.
7. *See infra* note 39 and accompanying text.

media access by the plaintiff.. This struggle has continued, and in recent years, has run up against the technological developments and trends in the area of online media and user-generated content. Part IV of this Note will provide an overview of the changing nature of the media landscape, noting the striking increase in the number and variety of ways that individuals can access larger audiences through the Internet—and the very fact that such networking has become ubiquitous (indeed, almost expected) in today's society. The effect of such universal access and networking should not go unnoticed by courts when they are considering an individual who is claiming defamation, but such access does not necessarily equate to the level of access imagined by *Gertz* when the Court established that the ability to redress defamation claims is a factor to be considered.

This Note will then argue in Part V that the access to media test is no longer applicable as it currently stands in this age of widespread access to media, and as such may no longer appropriately serve as a safeguard for private plaintiffs as it was initially envisioned by the *Gertz* Court. In order to do what the Court initially intended of it, the access to media test must take into account what the definition of "media" actually means today, and it then must be adequately tailored to reflect the trend of social networking and many-to-many online communication.[8] It is not enough to accept the ability to access some form of media—instead, the test must be appropriately limited in order to find only those who have the ability to access a similarly situated audience through a similar means of communication as having adequate means of redress through the media.

II. THE PUBLIC-PRIVATE DISTINCTION IN DEFAMATION LAW

Prior to 1964, defamation law was exclusively governed by state law,[9] but that changed with *New York Times Co. v. Sullivan*.[10] The case came before the Supreme Court in a time of political change, and with it came a sea of change for the legal world, as well; the Court's decision was "one of the most famous and important cases in all of constitutional jurisprudence."[11] With this decision, the Court gave a constitutional backbone to the law of defamation—recognizing the First Amendment importance of core political speech and the need to provide publishers with "breathing space" for such speech to occur.[12] In subsequent cases, the

8. *See infra* note 128 and accompanying text.

9. Erik Walker, *Defamation Law: Public Figures—Who Are They?*, 45 BAYLOR L. REV. 955, 956 (1993).

10. N.Y. Times Co. v. Sullivan, 376 U.S. 254, 254 (1964).

11. MARC A. FRANKLIN, DAVID A. ANDERSON & LYRISSA BARNETT LIDSKY, MASS MEDIA LAW 271 (7th ed. 2005).

12. *N.Y. Times*, 376 U.S. at 272 (citing NAACP v. Button, 371 U.S. 415, 433 (1963)) (internal quotation omitted).

Court broadened the scope of the rules set forth in *New York Times* to occupy the area of defamation law by issuing a series of constitutional decisions,[13] each decision building upon the last.

A. New York Times Co. v. Sullivan

The case with the most significant impact on defamation law began in the arena of the civil rights movement. It stemmed from a full-page editorial advertisement that ran in the *New York Times*[14] that included statements about police and official action against civil rights demonstrators that had taken place in Montgomery, Alabama.[15]

The ad contained some apparently false statements regarding the events that had occurred in Montgomery.[16] A claim was brought by the Commissioner of Public Affairs in Montgomery, L.B. Sullivan, who alleged that the advertisement concerned him because of his role in supervising the Montgomery Police Department.[17] Sullivan claimed that the charges asserted by the advertisement were leveled at him simply because of the nature of his duties[18] and that he had therefore been libeled by the advertisement.[19] The trial court agreed, finding the advertisement libel per se,[20] a ruling that was upheld by the Supreme Court of Alabama.[21]

In a unanimous decision to reverse the ruling, the U.S. Supreme Court held that the rule of law applied by the Alabama courts was "constitutionally deficient for failure to provide the safeguards for freedom of speech and of the press that are required by the First and Fourteenth Amendments in a libel action brought by a public official against critics of his official conduct."[22] In addition, continued the Court, a public official must prove that the publication acted with "actual malice," that is, "with knowledge that it was false or with reckless disregard of whether it was

---

13. Thomas Kane, *Malice, Lies, and Videotape: Revisiting* New York Times v. Sullivan *in the Modern Age of Political Campaigns*, 30 RUTGERS L.J. 755, 762 (1999).

14. *N.Y. Times*, 376 U.S. at 254.

15. *Id.*

16. *Id.* at 258–59 (including other falsehoods such as that the song the ad indicated was sung by the demonstrators was mistaken; that the reasons for the expulsion of some of the students were mischaracterized; that the campus dining hall was never padlocked; that students had protested by boycotting classes rather than refusing to register for classes; that the police never surrounded the campus, though they were deployed on three occasions; and that Dr. Martin Luther King, Jr. had only been arrested four times rather than seven).

17. *Id.* at 256.

18. *Id.* at 258 ("Respondent and six other Montgomery residents testified that they read some or all of the statements as referring to him in his capacity as Commissioner."). Sullivan was never actually referred to by name in the advertisement itself. *See id.*

19. *Id.* at 256.

20. *Id.* at 262.

21. *Id.* at 263.

22. *Id.* at 264.

false or not."[23]

This holding by the Court marked the first time that the First Amendment played a role in defamation law; the Court upheld these protections as necessary to give freedom of expression the "breathing space" it requires.[24] This was a recognition by the Court of the potential for a "chilling" effect if such core political speech was not protected.[25] For the first time, the bright line that protected plaintiffs from untrue speech was blurred—the actual malice test ultimately protected those speakers who acted without legitimate awareness of the falsity of their speech when speaking about public officials. The aim was to allow discourse concerning public officials as it advanced the introduction of important ideas into the marketplace.

Following this decision, a series of cases fell into line before the Court. Over the next decade, one case after another was decided that expanded upon or clarified the Court's decision in *New York Times*. Most significantly for purposes of this Note, the Court expanded the class of individuals who were subjected to the actual malice requirement to include not just public officials, but also public figures.[26]

B. Gertz *and the Origins of the Access to Media Test*

In 1973, a case came before the Court regarding a Chicago attorney, Elmer Gertz, who was representing the family of a youth who had been shot by a Chicago policeman. His defamation case resulted from an editorial in *American Opinion*[27] that accused Gertz of being a "Communist-fronter"[28] and of being a member of an organization that had planned a Communist attack on the Chicago police.[29] However, the issue in this case

---

23. *Id.* at 280.

24. *Id.* at 271–72.

25. Kane, *supra* note 13, at 771.

26. This extension was officially made in the companion cases *Curtis Publishing Co. v. Butts* and *Associated Press v. Walker*, 388 U.S. 130, 155 (1967), but the test for determining how a plaintiff should achieve the status of public figure was set forth in *Gertz v. Robert Welch, Inc.*, 418 U.S. 323, 351 (1974).

27. *American Opinion* is a publication of the John Birch Society. *Gertz*, 418 U.S. at 325. At the time, the publication was reporting on a supposed Communist conspiracy against law enforcement. *Id.* For more information about the John Birch Society, see *About the John Birch Society*, JOHN BIRCH SOC'Y, http://www.jbs.org/about (last visited Feb. 23, 2011).

28. *Gertz*, 418 U.S. at 326.

29. *Id.* (noting that the organization was the National Lawyers Guild, of which the plaintiff was in fact a member, but that there was no evidence that he or the organization had taken any part in planning the demonstrations during the 1968 Democratic Convention, as asserted by the article). Significantly, in light of the actual malice standard, the Court noted that the editor of *American Opinion* had made no effort to verify the charges against Gertz, despite an editorial introduction to the article that claimed extensive research had been

that was the focus of much of the Court's discussion[30] was that of Gertz's presence in the public realm—or lack thereof.

Two years earlier, the Court decided *Rosenbloom v. Metromedia, Inc.*, in which it concluded that the *New York Times* standard applied in such cases that concerned matters of public or general concern[31]—a holding that would certainly lend itself to application in this case because of the publicity surrounding the youth's death in Chicago. However, the plurality decision in *Rosenbloom* left no clear guidance for the application of the *New York Times* standard, so the Court in *Gertz* had to revisit the decision in order to place its holding in the "proper context."[32] In doing so, the Court determined that *Rosenbloom* extended the application of the *New York Times* standard to a degree that the Court found "unacceptable,"[33] leaving otherwise private plaintiffs without an adequate legal remedy for defamatory falsehoods injurious to their reputations.[34] Under the precedent set by *Rosenbloom*, any time a private plaintiff found himself involved in a story of interest to the public, he would be required to prove *New York Times* actual malice. With *Gertz*, the Court took a step back from this broad view of the standard for the sake of protecting the truly private plaintiff. The Court recognized that a story garnering media attention does not necessarily make every individual involved in that story a public figure without more. Had it left the test as it was, any person mentioned in any story in the media would automatically meet the *Rosenbloom* standard and be required to show actual malice. This was a burden the Court was not willing to force upon all individuals without requiring a more searching inquiry into their actual role in the issue, and whether they were capable of responding to any allegations leveled at them.

With such concerns in mind, the Court held that the standard for determining whether a plaintiff is in fact a public figure should require looking to the reach of the plaintiff's prominence. On the one hand, courts must consider whether he has achieved "pervasive fame or notoriety . . . for all purposes and in all contexts,"[35] making him a general-purpose public figure. On the other hand, a court must consider whether it is dealing with a plaintiff who has voluntarily injected himself into, or has been drawn involuntarily into, a public controversy such that he "becomes a public

---

conducted. *Id.* at 327.

30. The Court also discussed at length the appropriate level of proof necessary for plaintiffs depending upon whether they are classified as public or private figures. *Id.* at 342–48. However, this aspect of the Court's holding is not relevant to the discussion here.

31. 403 U.S. 29, 44 (1971).

32. *Gertz*, 418 U.S. at 333.

33. *Id.* at 346.

34. *Id.*

35. *Id.* at 351.

figure for a limited range of issues"[36]—the limited-purpose public figure.[37] In describing how a plaintiff might voluntarily inject himself into an issue, the Court stated that he must "thrust himself into the vortex of [the] public issue, [or] engage the public's attention in an attempt to influence its outcome."[38] A key aspect to the Court's reasoning was the fact that public figures, like the public officials discussed in *New York Times*, also tend to have more effective opportunities to redress such defamatory statements by maintaining regular access to the media.[39]

This final point—the self-help available to public figures—has remained a factor in subsequent defamation cases without adequate consideration of its context at the time of the *Gertz* decision and its changing context in light of today's media landscape. The Court addressed the issue quite simply in *Gertz*, stating merely: "Public officials and public figures usually enjoy significantly greater access to the channels of effective communication and hence have a more realistic opportunity to counteract false statements than private individuals normally enjoy."[40]

The Court treated the notion of access with little explanation, because at the time there existed only one definition of what media could mean, so invariably the media world in which the plaintiff was defamed would be similar to, if not the same as, the type in which that plaintiff could attempt to respond. The Court made no reference to whether there was a differentiation necessary when the defamation appeared in national media versus local media, but it seemed to accept that media, generally speaking, meant the print and broadcast media of the day. Thus it was in those media that defamation could be expected to originate, and it was in those same media that the plaintiff should seek to rebut such defamation.

III. THE ACCESS TO MEDIA TEST IN ACTION

Since *Gertz*, the access to media element of the public figure test has been used frequently by the Supreme Court, as well as by lower courts

---

36. *Id.*

37. It is clearly much more common for an individual to rise to the level of public figure in the context of one particular controversy. Consider, for instance, Bernard Madoff, who was little known outside Wall Street prior to his arrest and conviction for "the biggest financial swindle in history." Robert Frank & Amir Efrati, *'Evil' Madoff Gets 150 Years in Epic Fraud*, WALL ST. J., June 30, 2009, at A1. For a person to achieve pervasive fame or notoriety, it is generally understood that his name must be universally (or at least widely) recognizable. Examples might include the late Michael Jackson or Oprah Winfrey, figures who are not linked to one particular achievement or controversy but who are recognizable in all contexts.

38. *Gertz*, 418 U.S. at 352.

39. *Id.* at 344; *see also* RODNEY A. SMOLLA, LAW OF DEFAMATION § 2.05 (1st ed. 1986).

40. *Gertz*, 418 U.S. at 344.

(though not with complete consistency[41]), to separate the categories of defamation plaintiffs. The Court has continued to justify and explain the element,[42] and lower courts have continued to rely on it, frequently citing to *Gertz* for the basis of the test.[43]

A. The Role of the Test in Categorizing Plaintiffs

Members of the Court have seen private individuals' inability to access the media as a vulnerability, one that justifies protection of the private individual by not requiring her to prove actual malice under *New York Times*.[44] In *Hutchinson v. Proxmire*, the Court provided further elucidation of the rule, and found that it is not sufficient merely to show that the plaintiff is able to respond to the defamatory statements and have such responses published in order to establish that he has access to media.[45] Instead, the plaintiff must have what the Court describes as "regular and continuing access to the media," as such is "one of the accouterments of having become a public figure."[46] In addition, in order for such access to be sufficient for the purposes of the *Gertz* test, it must command enough media attention to effectively rebut the defamatory statements[47] (despite the Court's concession in *Gertz* that rebuttal "seldom suffices to undo harm of defamatory falsehood"[48]).

However, it is not clear to what extent the plaintiff must have the ability to rebut defamatory statements. In *Hutchinson*, the Supreme Court attempted to provide more guidance as to this factor,[49] and in doing so, it created ambiguity as to the threshold for sufficiency when it comes to rebuttal or media access. This decision by the Court obscured the notion of what type of access is necessary, giving weight to the ability to access the media on a regular basis, rather than simply for the purpose of rebuttal in

---

41. See discussion of *Waldbaum v. Fairchild Publications, Inc.*, 627 F.2d 1287 (D.C. Cir. 1980) *infra* pp. 13–15.

42. *See, e.g.*, Milkovich v. Lorain Journal Co., 497 U.S. 1, 15 (1990); Phila. Newspapers, Inc. v. Hepps, 475 U.S. 767, 774 (1986); Hutchinson v. Proxmire, 443 U.S. 111, 136 (1979).

43. *See, e.g.*, Wood v. Hustler Magazine, Inc., 736 F.2d 1084, 1090 (5th Cir. 1984); Street v. NBC, 645 F.2d 1227, 1234 (6th Cir. 1981); Mzamane v. Winfrey, 693 F. Supp. 2d 442, 501 (E.D. Penn. 2010); Chapman v. Journal Concepts, Inc., 528 F. Supp. 2d 1081, 1090 (D. Haw. 2007); Desai v. Hersh, 719 F. Supp. 670, 673 (N.D. Ill. 1989); Reader's Digest Ass'n v. Superior Court, 690 P.2d 610, 615 (Cal. 1984); Stolz v. KSFM 102 FM, 35 Cal. Rptr. 2d 740, 743 (Ct. App. 1994); Ellerbee v. Mills, 422 S.E.2d 539, 540 (Ga. 1992).

43. *Philadelphia Newspapers, Inc.*, 475 U.S. at 789 (Stevens, J., dissenting).

45. *Hutchinson*, 443 U.S. at 136.

46. *Id.*

47. Wolston v. Reader's Digest Ass'n, 443 U.S. 157, 171 (1979) (Blackmun, J., concurring).

48. Gertz v. Robert Welch, Inc., 418 U.S. 323, 344 n.9 (1974).

49. *Hutchinson*, 443 U.S. at 136.

response to media attention in the alleged defamation.[50] For lower courts, this has resulted in a trend of paying "lip service to the media access requirement,"[51] but without a clear consensus on what its weight should be, nor on what "access to media" means.[52] As one federal court put it, the resulting analysis for courts in determining who is a public figure has become "much like trying to nail a jellyfish to the wall."[53]

An example of a lower court's struggle with the access to media factor was demonstrated by the Fourth Circuit in *Hatfill v. New York Times Co.*[54] In this case, the access to media factor was used as one of several factors that were determinative of the plaintiff's status as a limited-purpose public figure.[55] Hatfill, a well-regarded scientist in his field of study, was accused by a columnist in the *New York Times* of sending letters containing anthrax to members of Congress and news organizations.[56] The court considered his renown in the field of bioterrorism and biological weapons, and therefore his ability to gain attention from media and the public in that arena, as sufficient for showing that he had continuing access to the media.[57] Instead of focusing on whether he could access the same types of media that had published the allegedly defamatory statements, the court focused on his ongoing relationship with scientific journals and experts in the field as proving sufficient access to channels of communication.[58]

Hatfill cited the Fourth Circuit's decisions in *Reuber v. Food Chemical News, Inc.*[59] and *Fitzgerald v. Penthouse International*.[60] In *Fitzgerald*, the court announced a five-factor test for determining whether the plaintiff is a public figure.[61] The first factor asked whether "the plaintiff had access to channels of effective communication."[62] When the court in 1990 again was faced with a defamation claim in *Reuber* by a plaintiff who purported to be a limited-purpose public figure, the court applied this same *Fitzgerald* test and focused on the plaintiff's activity within his field of expertise, including lectures he had given and reports he had published.[63] In

---

50. *Id.*

50. Walker, *supra* note 9, at 976.

51. *Id.*; *see, e.g.*, Lerman v. Flynt Distrib. Co., 745 F.2d 123, 136–37 (2d Cir. 1984).

53. Rosanova v. Playboy Enters., 411 F. Supp. 440, 443 (S.D. Ga. 1976), *aff'd*, 580 F.2d 859 (5th Cir. 1978).

54. 532 F.3d 312 (4th Cir. 2008).

55. *Id.* at 318–19.

56. *Id.* at 314, 320–21.

57. *Id.* at 322.

58. *Id.* at 320–21.

59. 925 F.2d 703 (4th Cir. 1991).

60. 691 F.2d 666 (4th Cir. 1982).

61. *Id.* at 668.

62. *Id.*

63. *Reuber*, 925 F.2d at 708.

this instance, it was within this scientific arena that Reuber's reputation had come under fire, and based on that fact, the court found that looking at these channels and his access therein was the appropriate inquiry in considering where that reputation could be redeemed.[64] "The inquiry into access to channels of communication proceeds on the assumption that public controversy can be aired without the need for litigation and that rebuttal of offending speech is preferable to recourse to the courts."[65]

The court in *Reuber* unnecessarily went on to note that it was significant that the plaintiff there had not attempted to rebut the statements through those channels to which he had access. However, *Gertz* did not ever clearly state that an attempt at rebuttal is necessary, but rather the appropriate inquiry is only whether the individual had the *opportunity* to do so based on his status.[66] And so in *Hatfill*, the Fourth Circuit correctly stepped back toward the *Gertz* conclusion and away from the analysis that the *Reuber* court had engaged in. The *Hatfill* court determined that it is not required that rebuttal be attempted, merely that the plaintiff's capability to do so be considered in weighing the individual's potential access to media.[67]

But in *Hatfill*, the court also seemed to disregard the importance of its position in *Reuber* that the channels of communication that are considered "effective" for the purposes of response are those same channels in which the reputation of the plaintiff was first at issue.[68] When the *Hatfill* court relied on this precedent, it mistakenly relied upon the attention Reuber had garnered within the same arena in which he was defamed—the court treated this as a signal that a visible reputation within a scientific community was sufficient to show access to channels of effective communication.[69] The error the *Hatfill* court committed when drawing its comparison to *Reuber* was its disregard for the fact that Reuber, unlike Hatfill, was alleging defamation in the same arena in which he had gained public recognition; in contrast to Reuber, Hatfill was alleging defamation in the *New York Times*—clearly *not* a scientific journal or science-specific publication. And while the court engaged in a discussion of the various times he had been interviewed by or mentioned in similar such media

64. *Id.*

65. *Id.* at 708–09 (citing Gertz v. Robert Welch, Inc., 418 U.S. 323, 344 (1974)).

66. Gertz v. Robert Welch, Inc., 418 U.S. 323, 344 (1974).

67. *See* Hatfill v. N.Y. Times Co., 532 F.3d 312, 317–18, 322 (4th Cir. 2008).

68. *Reuber*, 925 F.2d at 708–09.

69. *Hatfill*, 532 F.3d at 322 ("In *Reuber*, we found that the plaintiff had testified before Congress and the Environmental Protection Agency; had given lectures on subjects related to the allegedly defamatory articles in which he was mentioned; had provided interviews to a newspaper; and had published several relevant scientific papers. If Reuber's access to channels of communication was sufficient, so too is Dr. Hatfill's." (citation omitted)).

outlets, the court did not make it clear that it was on the basis of his ability to access those outlets that his access to media was considered sufficient.[70] Indeed, it is not clear from the court's analysis whether it would have been merely sufficient for the purposes of the access to media test to show that Hatfill enjoyed renown in the field of bioterrorism, or whether it was the fact that he had also had been interviewed for both newspapers and television reports that satisfied the requirement.[71]

It is that latter level of effectiveness that would seem to be the one considered and set forth by *Gertz*, since the *Gertz* Court was aiming at the notion of rebuttal—the ability to mitigate harm done by the purportedly defamatory statement by accessing the same or a substantially similar audience.[72] Merely showing that a plaintiff enjoys some access to some form of media is not sufficient; in *Hatfill*, it must have been his access to the same or substantially similar outlets to the one in which the defamatory material appeared that proved he had the appropriate level of access to media to satisfy that prong of the limited-purpose public figure test.

B. Departure from the Access to Media Test

Other lower courts have not given this media access factor the same weight as the courts in the decisions discussed above; and some have found that it is not necessarily an integral part of the test in determining whether the individual is a public figure—despite references to *Gertz* and use of its language in stating the rule to be applied.

For instance, in *Waldbaum v. Fairchild Publications, Inc.*,[73] the district court set forth a three-part rule for determining whether the plaintiff is a limited-purpose public figure. First, there must be a public controversy

---

70. *See id.* at 321–22.

71. If it is the latter that the court intended to point to, then this would seem to satisfy the *Gertz* test as the court originally imagined it. That is, if it was because he was quoted in an article in the *Washington Post* and featured in a news broadcast on ABC News, as well as the variety of different media outlets that ran stories featuring comments by Hatfill in the days following the initial allegation, then this would seem to show that he had access to effective channels of communication that are in the same vein as the media outlet that initially published the allegedly defamatory statement (the *New York Times*). However, it is the court's reliance on *Reuber* that blurs its conclusion because of the different categories of media involved in the two cases. In *Reuber* it was only necessary to show that the plaintiff had access to scientific journals and similar such outlets; in *Hatfill*, the plaintiff's access must go significantly beyond the scientific community. Therefore, stating "If Reuber's access to channels of communication was sufficient, so too is Dr. Hatfill's," *Hatfill*, 532 F.3d at 322, seems to underestimate the level of access necessary for Hatfill to satisfy the *Gertz* test.

72. Gertz v. Robert Welch, Inc., 418 U.S. 323, 344 (1974) (emphasizing the ability of plaintiffs to "counteract false statements" when considering what "effective communication" means).

73. 627 F.2d 1287 (D.C. Cir. 1980).

or a dispute that has received media attention because of its potential impact.[74] Second, the plaintiff's role in the controversy must be analyzed by considering whether he or she has in fact, as set forth in *Gertz*, "thrust" himself or herself into the public controversy.[75] Finally, the *Waldbaum* court considered the defamatory statement and its relationship to the plaintiff's role in the controversy.[76]

The court in *Waldbaum* makes no mention of access to effective channels of communication in order to respond to the defamatory statements, and, similar to *Waldbaum*, many courts have relied on such tests that do not use the access to media factor.[77] In fact, one such court makes a note of the lesser importance of the access to media factor of the test, even when it is used by courts, before proceeding to decline to use the test itself: "Almost anyone who finds himself in the middle of a controversy will likely have enough access to the press to rebut any allegedly libelous statements, thus satisfying the Supreme Court's first concern. It is perhaps because of this that the Court has regarded the second justification as more important."[78]

And so lower courts continue to regard the limited-purpose public figure test with some confusion, and without a consistent voice. These courts have attempted to use the guidance offered by the Supreme Court by way of *Gertz* and *Hutchinson*, but have not managed to reach a consensus on the importance of the access to media prong of the test.[79] While it is clear that the Court regarded the role of the plaintiff in the controversy itself as an important determination for a court to make when assessing the classification of the plaintiff, his access to media was certainly an aspect the Court considered essential in *Gertz* and *Hutchinson*. It is just the exact nature of this access that was not clearly defined.

---

74. *Id.* at 1296.

75. *Id.* at 1297; *see also Gertz*, 418 U.S. at 345.

76. *Waldbaum*, 627 F.2d at 1298.

77. *See, e.g.*, Parsi v. Daioleslam, 595 F. Supp. 2d 99, 105 (D.D.C. 2009); Framsted v. Mun. Ambulance Serv., 347 F. Supp. 2d 638, 662 (W.D. Wis. 2004); Howard v. Antilla, No. 97-543-M, 1999 U.S. Dist. LEXIS 19772, at *3–4 (D.N.H. Nov. 17, 1999).

78. Clyburn v. News World Comms., Inc., 903 F.2d 29, 32 n.2 (D.C. Cir. 1990).

79. From lower courts' downplaying of what "access to media" actually means in the defamation context, it often appears that the heart of the overall test to determine whether a plaintiff is public or private is in fact the role the individual played in the controversy—whether he had voluntarily injected himself in it or thrust himself to the forefront. This is in keeping with the discussion in *Gertz* that emphasized that the heart of the issue was not the relative ease with which the public individual can access the media, but the very fact that he brought publicity upon himself in the first place. *Gertz*, 418 U.S. at 344, 345. Eliminating the access to media test, however, ignores the Court's added concern about plaintiff's ability to respond to the allegations, and therefore to effectively redress the claims made against him. *See id.* at 344.

C. The Test as Imagined by the Gertz *and* Hutchinson *Courts*

In considering *Gertz* and *Hutchinson* together, the Court's aim with the access to media element of the test seems to be weighing the plaintiff's ability to command media attention in order to redress claims leveled against him.[80] This would seem to resemble something more like the Fourth Circuit's description in *Reuber* of the capability to access "the fora where [the plaintiff's] reputation was presumably tarnished and where it could be redeemed."[81] It was the Supreme Court's goal to consider when a plaintiff would be able to effectively limit the damage done to him by defamatory statements in the media, and lower courts that recognize the importance of that aspect of defining a public figure have continued to use that element of the test.[82]

The difference between a public figure and a private individual changes the nature of what the plaintiff must prove in a defamation case, and a public figure—more capable of accessing the media and therefore of clearing his name—has a more difficult burden of proof. So it is the private figure that the *Gertz* and *Hutchinson* courts were considering; it is the private figure—who is unable to effectively stave off the negative comments made against him by responding with his own comments—that the Court was interested in protecting. Thus the Court's concern was allowing those private individuals to prove their case and receive their remedy through the courts.

However, the actual use of the test as applied by the lower courts[83] often looks primarily at the first factor in the limited-purpose public figure test—that of the plaintiff's role in the controversy—and less so at the plaintiff's ability to respond effectively to defamatory statements that appear in the media. Doing so, in fact, may seem logical in today's world of twenty-four-hour news cycles and fully integrated media outlets; the media are not only more accessible for the private individual, but in fact are at his fingertips.[84] But this is not necessarily the most protective approach for private individuals. This media world, with so many eyes on so many

---

80. *Id*. at 345; Hutchinson v. Proxmire, 443 U.S. 111, 136 (1979).

81. Reuber v. Food Chem. News, Inc., 925 F.2d 703, 708 (4th Cir. 1991).

82. *See, e.g.*, Wells v. Liddy, 186 F.3d 505, 531, 534 (4th Cir. 1999); Douglass v. Hustler Magazine, 769 F.2d 1128, 1141 (7th Cir. 1985); Chapman v. Journal Concepts, Inc., 528 F. Supp. 2d 1081, 1093 (D. Haw. 2007).

83. *See* Mark D. Walton, *The Public Figure Doctrine: A Reexamination of* Gertz v. Robert Welch, Inc. *in Light of Lower Federal Court Public Figure Formulations*, 16 N. ILL. U. L. REV. 141, 166 nn.134–35 (1995).

84. *See, e.g.*, Aaron Perzanowski, Comment, *Relative Access to Corrective Speech: A New Test for Requiring Actual Malice*, 94 CALIF. L. REV. 833, 836 (2006) ("The public figure doctrine fails to account for access to means of corrective speech so prevalent on the [I]nternet. But ironically, the ability to respond to defamatory speech served as a central consideration in the creation of the public figure test.").

different sources, still does not allow for just anyone to have the kind of access that the *Gertz* Court had imagined was possessed solely of public figures. And it is true that not all courts have moved away from the access to media test altogether. Therefore, the access to media test remains a potentially confusing and damaging tool for the courts to wield in separating public figures from private individuals. In order to effectively make this distinction, the *Gertz* vision of the access to media prong must be revived to give it a new meaning and new life.

IV. DEVELOPMENTS IN MEDIA, SHIFTS IN THE MAINSTREAM CURRENT

At the time that *Gertz* was decided, the media consisted solely of print and broadcast outlets. This media makeup was taken for granted by the Court in its almost dismissive reference to the greater access enjoyed by public officials and public figures.[85] While it remains true that public officials and public figures are in the best position to garner the attention of large media outlets with minimal effort, this model fails to account for the massive changes that have taken place—and are still taking place—in the media world, and how those changes may impact the limited-purpose public figure.

A. New Definitions, New Media

In recent years, the communication world has undergone a "dramatic democratization"[86] and the media landscape has shifted greatly. With the advent of the blogosphere, followed closely by the rise of Facebook, YouTube, and Twitter, Internet media are no longer irrelevant or obscure.

Just a few short years ago, blogs were considered to be on the periphery of the media world, something less than real journalism. In 2006, blogs were referred to by one columnist as "the bustling, energetic Wild West of the new Internet media."[87] Even though, at that time, blogs had proven their significance by forcing Dan Rather's hand in revealing the truth about President George Bush's military record after he reported on President Bush's National Guard service based on what turned out to be forged documents,[88] blogs were still on the verge of being taken seriously.

However, bloggers no longer go relatively unnoticed. If there are rumors circulating in the blogosphere, they will often be responded to in

---

85. *See Gertz*, 418 U.S. at 344.

86. Perzanowski, *supra* note 84, at 833.

87. Ellen Goodman, *Bloggers Owe Carroll an Apology*, BOS. GLOBE, Apr. 7, 2006, at A17.

88. *See* Tim Goodman, *Apology from CBS on Bush Memos*, S.F. CHRON., Sept. 21, 2004, at A1.

the media. An example from the McCain-Palin campaign demonstrates this phenomenon. Despite the absence of any "mainstream" press coverage of Palin's sixteen-year-old daughter's pregnancy, a press release was issued by McCain's campaign addressing the pregnancy to dispel rumors that had been cropping up on blogs.[89] Now "prominent journalists, many of whom are bloggers themselves, promote blogs—or at least certain blogs, such as those run by mainstream media outlets—as legitimate media outlets."[90] Blogs have become normal features on news outlets' websites,[91] and in fact, it is commonly a marketing or corporate tool, without which professional competitors might see an organization as an outcast.[92]

In addition to blogs, Facebook and Twitter have recently taken on a legitimate role in the world of online media. More and more organizations are using Facebook and Twitter for their massive reach and their communication and marketing potential. Congressmen are taking tutorials on how to use Facebook to further relationships with constituents,[93] and there have been announcements of political candidacy on Twitter that are then reported in the print media.[94]

With so much integrated use of online services, it is clear that these tools are coming closer to the center of the media stage. However, it is not clear that courts are in tune with these changes, nor is it clear that they are prepared to accept the possibility that an otherwise private individual may have the capacity to reach thousands through her Facebook page, tweets, or blog, without necessarily assuming a place in the realm of public figures.

---

89. Michael D. Shear & Karl Vick, *No Surprises from Palin, McCain Team Says*, WASH. POST, Sept. 2, 2008, at A17 ("McCain advisers said that after talking to Palin, they decided to issue the statement about Bristol's pregnancy in the wake of repeated inquiries from reporters after liberal blogs raised questions . . . .").

90. Anthony Ciolli, *Bloggers as Public Figures*, 16 B.U. PUB. INT. L.J. 255, 257 (2006).

91. *See, e.g.*, *CNN Political Ticker*, CNN.COM, http://politicalticker.blogs.cnn.com/ (last visited Feb. 23, 2011); *Blog Directory*, NYTIMES.COM, http://www.nytimes.com/ref/topnews/blog-index.html (last visited Feb. 23, 2011).

92. Etan Horowitz, *Film Recalls Blogging's Simpler Times*, ORLANDO SENTINEL, Aug. 9, 2009, at G1; *see also* Rob Johnson, *Running the Show—Screen Shots: Product Placements Aren't Just for Big Companies Anymore*, WALL ST. J., Sept. 28, 2009, at R9 (explaining the potential for product placement and advertising on Facebook, Twitter, and blogs).

93. Ian Shapira, *Lawmakers Find a Friend in the Power of Facebook*, WASH. POST, Dec. 30, 2009, at C01 (discussing a lesson given by a Facebook representative to Republican congressman, Rep. Peter Roskam of Illinois, who was learning about the ways to use Facebook to provide more personal and timely information to his constituents, and to help his constituents feel more connected to him).

94. *See, e.g.*, Tom Infield, *Gerlach Declares GOP Run for Governor*, PHILA. INQUIRER, July 15, 2009, at B01.

B. "New" Media and the Impact on Defamation Law

The Supreme Court's basic assumptions as to media in the time of *New York Times* and *Gertz* reflect the nature of media in that time—"a simplistic and antiquated conception"[95] that hardly compares to how the media world looks today. The contexts in which blogs come up in courtrooms often involve reporters' privileges (that is, whether privileges that are granted to journalists should be extended to bloggers, as well[96]) and whether or not anonymous bloggers can be forced to reveal themselves when they have made defamatory statements.[97]

Courts are certainly not entirely unaware of the existence of this form of media, be it blogs or Facebook or MySpace.[98] Instances of abuse or harassment stemming from interactions on Facebook and MySpace are not infrequent;[99] child pornography and other cybercrimes force courts to look

---

95. Perzanowski, *supra* note 84, at 833.

96. *See, e.g.*, *In re* Miller, 397 F.3d 964, 979 (D.C. Cir. 2005) (Sentelle, J., concurring).

> [I]f we extend that privilege to the easily created blog . . . have we defeated legitimate investigative ends of grand juries in cases like the leak of intelligence involved in the present investigation? . . . [D]oes the privilege also protect the proprietor of a web log: the stereotypical "blogger" sitting in his pajamas at his personal computer posting on the World Wide Web his best product to inform whoever happens to browse his way? If not, why not?

Id. at 979, 981 (Sentelle, J., concurring); *see also* Lee v. Dep't of Justice, 401 F. Supp. 2d 123, 140 (D.D.C. 2005) ("The proliferation of communications media in the modern world makes it impossible to construct a reasonable or useful definition of who would be a 'reporter' eligible to claim protection from a newly minted common law privilege."). There is little exploration by courts as to whether bloggers are journalists, or something different entirely. *See* BidZirk, LLC v. Smith, No. 6:06-109-HMH, 2007 U.S. Dist. LEXIS 78481, at *16 (D.S.C. Oct. 22, 2007) ("[T]here is no published case deciding whether a blogger is a journalist.").

97. *See, e.g.*, McVicker v. King, 266 F.R.D. 92 (2010) (finding that First Amendment rights can be asserted by those posting on a blog, thereby leaving their anonymity intact); Mobilisa, Inc. v. Doe, 170 P.3d 712 (Ariz. Ct. App. 2007) (setting forth the test that must be met in order to compel the discovery of an anonymous Internet user in defamation cases); Krinsky v. Doe 6, 72 Cal. Rptr. 3d 231 (Ct. App. 2008) (holding that a subpoena to reveal an Internet poster's identity should have been quashed).

98. *See, e.g.*, Salter v. State, 906 N.E.2d 212, 220 (Ind. Ct. App. 2009) (acknowledging that Facebook and MySpace increase the risk of child pornography images appearing on the Internet); *In re* Forgione, 908 A.2d 593, 603 n.11 (Conn. Super. Ct. 2006) (acknowledging that students could access one another's personal information via "an Internet program or service known as 'The Facebook'").

99. *See, e.g.*, United States v. Beckett, No. 09-10579, 2010 U.S. App. LEXIS 4989 (11th Cir. Mar. 9, 2010) (dealing with an appeal from a conviction for child pornography charges that arose from the defendant's falsely created MySpace accounts, which were used to persuade minors to send nude photos over the Internet); United States v. McCloud, 590 F.3d 560 (8th Cir. 2009) (dealing with an appeal from a conviction for producing child pornography and in which the court described the defendant's interactions with victims over MySpace); United States v. Drew, 259 F.R.D. 449 (C.D. Cal. 2009) (describing interactions over MySpace through a fake profile set up by defendant, including the resulting suicide of the target of the MySpace interactions); United States v. Infante, No. 10-6144M, 2010 U.S.

to the Internet and develop at least a cursory understanding of its contents. Even jury instructions appropriately address the Internet services that might allow jurors to communicate with others.[100] However, it does not seem that courts are yet comfortable with defining the role that the Internet will play in defamation law as a component of the media—not just as courts grapple with how to appropriately address anonymous bloggers who are liable for defamation, but also how this arm of the media should be treated in considering the defamed individual's options for redress.

Media are rarely specifically defined in the defamation context, giving little guidance for what should be included in a court's assessment of just what media qualify for the access to media test.[101] Without taking that extra step to establish the types of media at play, courts are missing a major point of the *Gertz* test: the *Gertz* Court imagined this prong as a means of redress—redress cannot happen unless an audience that is the same or substantially similar can be accessed and exposed to such a rebuttal.

Dist. LEXIS 30730 (D. Ariz. Mar. 30, 2010) (describing the defendant's stalking of the victim, which included contact via Facebook).

100. *See, e.g.*, *In re* MAI-CIVIL, 2009 Mo. LEXIS 544, at *5–6 (Mo. Nov. 23, 2009) (including an instruction admonishing jurors that they are not to "use a cell phone, record, photograph, video, e-mail, blog, tweet, text, or post anything about this trial . . . to the Internet, 'facebook', 'myspace', 'twitter', or any other personal or public web site . . . ."); People v. Jamison, 899 N.Y.S. 2d 62 (N.Y. Sup. Ct. 2009) (instructing jury not to use Google Earth or to text or chat online about the case, in addition to instructing them not to communicate on social websites such as Facebook, MySpace, or Twitter).

101. Thomas D. Brooks, Note, *Catching Jellyfish in the Internet: The Public-Figure Doctrine and Defamation on Computer Bulletin Boards*, 21 RUTGERS COMPUTER & TECH. L.J. 461, 478 (1995) ("The Court has never offered anything near a working definition of 'the media.' Rather, its approach is reminiscent of that employed by Justice Stewart when faced with the task of defining pornography: the justices know it when they see it.") (citing Jacobellis v. Ohio, 378 U.S. 184, 197 (1964) (Stewart, J., concurring)); *see also, e.g.*, Chapman v. Journal Concepts, Inc., 528 F. Supp. 2d 1081 (D. Haw. 2007); Fiacco v. Sigma Alpha Epsilon Fraternity, 484 F. Supp. 2d 158 (D. Me. 2007) (summarizing that the position held by the plaintiff involved access to media, mentioning only articles published in a campus newspaper); Chafoulias v. Peterson, 668 N.W.2d 642, 654 (Minn. 2003) (concluding that, despite the fact that the initial defamatory story aired on a national ABC program, the plaintiff's ability to appear in a two-part story on a local NBC affiliate was sufficient to show he had "broad media access, allowing him to strategically place media appearances . . . ."). In *Chapman*, which concerned a plaintiff who was a surfer, the court summed up his media access:

> [T]he sheer volume of published materials quoting or referencing Plaintiff indicate that the surfing media was, and continues to be, interested in him . . . . Although the record on this matter is thin, it appears to the court that if Plaintiff wanted to rebut [the] article—whether through an interview, profile, or opinion piece—the surfing media would be receptive.

Chapman, 528 F. Supp. 2d at 1092. This is not atypical of a court's treatment of this prong of the test, wherein the court ignores any mention of the type of media in which those "interview, profile, or opinion" pieces might run. After acknowledging the necessity of assessing access, courts do not specifically explain what media would have satisfied the prong, nor the types of media involved in the instant case that do satisfy the prong.

V. ACCESS, ACCESS EVERYWHERE

In the current media environment, anyone with a computer can become a publisher, and while many bloggers remain in obscurity, bloggers and those well connected on social networking sites can successfully gain media attention.[102] When that occurs, a blogger who was otherwise a private individual may open herself up to the possibility of defamatory statements.

In order to appropriately protect the private blogger from the heightened standard of actual malice that she would be required to prove as a limited-purpose public figure, it is necessary to give weight to the other prongs of the test—that is, whether there is an isolated controversy, whether the plaintiff has voluntarily thrust herself into the controversy, and so on—before jumping straight to the access to media prong. In the absence of such an approach, courts will necessarily lapse back to the reasoning of the Court in *Rosenbloom*—one that was found to be unacceptable by the *Gertz* Court[103]—by weighing only the element of public interest in the controversy at hand and allowing that to uniformly create limited-purpose public figures.

Once the other factors of the test have been appropriately weighed, courts can turn to the access to media prong to differentiate plaintiffs who may not automatically seem to be a central figure in the controversy from those who clearly have thrust themselves into the controversy and have taken the lead in determining its outcome.[104] It is with this prong that courts

---

102. *See, e.g.*, Kyra Kyles, *Bravo to Ordinary Twitter Celebrity*, REDEYE (Aug. 6, 2010, 8:18 AM), http://www.chicagonow.com/blogs/kyles-files/2010/08/column-fodder-bravo-to-ordinary-twitter-celebrity.html. The blog discusses Twitter user Steven Holmes, a UK citizen who became the first person rapper Kanye West began following shortly after West started using Twitter. Holmes rejected the attention the celebrity's following incited, tweeting—presumably after the interviews he granted to local British media—"I won't be speaking to anybody else; surprisingly not everyone wants to be famous . . . . That's all I'm saying—peace out x." *Id.* The RedEye column noted Holmes' ability to "recognize, and rebuff, the ridiculous fame seemingly bestowed on anybody these days, from a baby singing 'Single Ladies' to a grown man squealing like a sow over double rainbows." *Id.*; *see also* Sarah Lyall, *A Tweet Read Across Britain Unleashes a Cascade of Vitriol on a User*, N.Y. TIMES, Nov. 2, 2009, at A8 (discussing a "tweet" made by a user called "brumplum" that launched a frenzied debate and called attention to the user, an otherwise unknown resident of Birmingham, England); Maureen Ryan, *An Unlikely New Source of Writing Talent: Blogs*, CHI. TRIB., Oct. 8, 2003, at C1 (discussing bloggers who had garnered wide following and readership, and their subsequent hiring potential); P.J. Huffstutter & Jerry Hirsch, *Blogging Moms Wooed by Firms*, L.A. TIMES, Nov. 15, 2009, at A1 (discussing a trend of food companies calling upon "mommy bloggers" to review their products).

103. Gertz v. Robert Welch, Inc., 418 U.S. 323, 346 (1974).

104. If this is the case, an individual will often be both attracting and creating media coverage through the very nature of her involvement in the controversy. This is when the access to media prong can appropriately be downplayed, since when evaluated, it will be found to be satisfied.

can gauge the individual's ability both to seek redress through the media and to access an audience through which the defamation can be rebutted.

A. Constant Contact Between Private Individuals

Communicating constantly through social networking and other Internet service providers has become so much a regular and routine practice of private individuals that there is not an assumption of receiving widespread attention from those communications.[105] In this age of social networking, virtually everyone who is active on the Internet has become a publisher to some extent[106]—this means there are millions of potential news outlets to be accessed everyday, with far fewer eyes on any individual outlet. However, even though "[m]illions of teenagers use MySpace, Facebook, and YouTube to display their interests and talents, . . . the posting of that information hardly makes them celebrities."[107]

Without an emphasis on the voluntariness and involvement in the controversy, it could be argued that anyone who can publish online should be considered a limited-purpose public figure.[108] Inaccurate assumptions about accessing online audiences may lead to widening the scope of limited-purpose public figures, as it may be taken for granted that communicating to audiences online does not necessarily equate to seeking

---

105. *Cf.* Ciolli, *supra* note 90, at 257 (arguing that a blogger must expect to receive attention when she puts her thoughts about a controversy on a publicly accessible website because of widespread readership of blogs). *But see* Lyall, *supra* note 102 (noting that the user "brumplum" stated on his blog that his seemingly casual and "mildly critical" tweet about British actor Stephen Fry had resulted in an unexpected surge of Twitter followership and media attention, thus demonstrating the unexpected attention that a private individual can spur without doing more than typing a quick tweet); *Nottingham 'Tweeter' Gets Followed Online by Kanye West*, NOTTINGHAM EVENING POST (U.K.), Aug. 6, 2010, at 3 (noting that the Twitter user whom Kanye West began following did not think the publicity of having a celebrity following him, a move which resulted in the user gaining 6,000 followers on the social networking site despite his otherwise relative obscurity on Twitter, was "worth it,").

106. *See* Perzanowski, *supra* note 84, at 835.

107. D.C. v. R.R., 106 Cal. Rptr. 3d 399, 428 (Ct. App. 2010). This is contrary to what was once thought about the ability to respond on the Internet; when first the possibility of posting immediately on message boards became an option, some thought that this would mean that anyone capable of creating such a posting could adequately respond. Thus, by the same argument, anyone who could access the Internet was a public figure. For this argument, see generally Mike Godwin, *The First Amendment in Cyberspace*, 4 TEMP. POL. & CIV. RTS. L. REV. 1 (1994); Michael Hadley, Note, *The* Gertz *Doctrine and Internet Defamation*, 84 VA. L. REV. 477 (1998). However, it has become clear more recently that the Internet is more often a place for private individuals to network broadly than for private individuals to take on a public persona by virtue of their networking.

108. *See* David Gordon, *Taking the First Amendment on the Road: A Rationale for Broad Protection for Freedom of Expression on the Information Superhighway*, 3 COMMLAW CONSPECTUS 135, 142 (1995).

"both influence and attention."[109] While certainly a person posting on the news feeds[110] of his 800 Facebook friends may be well-known within that group, that is hardly grounds to require him to prove *New York Times* actual malice the moment he is defamed; this is even more evident on Twitter, where a relatively unknown individual can drum up followers numbering in the thousands, many of whom may not even know the user's real name.[111] This becomes a dangerous gray area when defamation is at issue, because the plaintiff who cannot successfully show he is a private figure will be required to show actual malice—a burden that the Supreme Court never imagined would extend to truly private individuals.

Early views of the Internet did not take into account the possibility of this user-generated world that is the Web of today.[112] As one attorney noted in 1995, "[the Internet publisher] has greater access (than private figures) to the mass media and, thus, needs less libel protection, because he can rebut claims against him . . . . Through global, instantaneous communication, everybody has the ability to rebut everybody."[113] It is true that the individual has means on the Internet to widely access other individuals, and now almost any individual can be such an "Internet publisher"; but the assumption that "[t]he mere act of creating a blog draws public attention to the author and his or her views"[114] does not hold true in an era of such proliferation of user-generated content. The existence of so many sources of information reduces the number of eyes on any one source; so despite posting information on the Internet, an Internet user does not necessarily guarantee herself access to an audience of any significant proportion. Therefore, without properly balancing the generalization that Internet users can adequately rebut statements made about them against the other considerations of the limited-purpose public figure test, and without tailoring the test to reflect the nature of the media involved both in the defamation and in the potential for response, it is not clear how widely such a classification might reach.

As such, it is necessary for courts to approach this new version of access to media with caution. Simply concluding that "[b]y creating a blog,

---

109. Ciolli, *supra* note 90, at 271.

110. Facebook publishes a "News Feed" on the home page of all users, documenting the status messages and activity of the user's friends.

111. This is made clear by the plight of "brumplum," the Twitter user who had the misfortune of offending a popular actor, and therefore who now has over 1,200 followers. Lyall, *supra* note 102.

112. *See, e.g.*, Godwin, *supra* note 107, at 5.

113. Gordon, *supra* note 108, at 142 (alteration in original) (quoting David L. Marburger, a lawyer for a defendant in an Internet libel suit) (internal quotation marks omitted).

114. Ciolli, *supra* note 90, at 272.

especially a blog that enables comments or [W]eb syndication feeds, individuals seek both attention and influence in public debate, and thus fulfill one of the elements of a limited purpose public figure,"[115] the limited-purpose public figure test will know no bounds on the Internet. Private individuals who are actively involved on the Internet will be crossing liability lines unawares—or worse, if such a trend were to actually gain legal steam,[116] individuals might be deterred from sharing or networking broadly online. This could put a damper on user-generated content on the Internet, a tool that has become ubiquitous in today's culture, and which is continually changing to reflect the new ways it can be used to connect individuals more widely on an international scale.[117] If individuals no longer feel that they are free to connect and share with one another without exposing themselves to the risk of becoming public figures in defamation claims, this modern version of the marketplace of ideas could be chilled.

B. Gertz *in the Age of Social Networking*

When the Court was deciding *Gertz*, it did so with a singular

---

115. *Id.*

116. *Id.* at 269 (noting that as of the date of that publication, no blogger had sued another individual or entity for defamation, but such lawsuits are inevitable).

117. Social networking sites have contributed to coordinating political activism on a grand scale in recent years. This was especially apparent amid the January 2011 uprisings in Tunisia and Egypt, the organization of which was largely credited to Facebook by a number of media outlets. *See, e.g.*, Roger Cohen, *Facebook and Arab Dignity*, N.Y. TIMES, Jan. 24, 2011, *available at* http://www.nytimes.com/2011/01/25/opinion/25iht-edcohen25.html?_r=1&scp=1&sq=facebook%20and%20arab%20dignity%20cohen&st=cse (In discussing the successful Arab uprising that overthrew the government of Tunisia without an identifiable leader, Cohen notes, "Or rather, its leader was far away: Mark Zuckerberg, the founder of Facebook. Its vehicle was the youth of Tunisia, able to use Facebook for instant communication and so cyber-inspire their parents. . . . Facebook propelled insurrection from the interior to the Tunisian capital in 28 days."); Griff Witte, *Egyptian Opposition Calls for Massive Protest; Foreigners Flee*, WASH. POST, Feb. 1, 2011, at A1 (noting that while Facebook was initially an organizational tool, Internet access became scarce after several days of protests). Similarly, in Iran in 2009, protesters used Twitter to draw international attention to violence against protesters as the protests were happening. *See* David Zurawik, *Iran Protests Present a Revelation, Challenges in Newsgathering*, BALT. SUN, June 28, 2009, at 1E; Nazila Fathi, *Iran's Opposition Seeks More Help in Cyberwar with Government*, N.Y. TIMES, Mar. 19, 2010, at A6; *see also The Rage of Followers*, WASH. POST, Apr. 11, 2010, at G2 (questioning whether sites such as Twitter and YouTube allow for more power to challenge leaders—or at least for more global recognition of repressive leadership—in light of protests in Kyrgyzstan); Michael Wines, Sharon LaFraniere & Jonathan Ansfield, *China's Censors Tackle and Trip over the Internet*, N.Y. TIMES, Apr. 8, 2010, at A1 (describing a particular blogger who maintains six different blogs in order to try to outwit Chinese censors who attempt to block certain types of political speech on the Internet and noting how this particular blogger sees other Chinese Internet users growing incensed against the restrictions on their Internet speech and attempting to push the wall back).

understanding of the media landscape as it existed in 1974.[118] At that time, the media were entirely limited to print and broadcast media, often represented by large conglomerate news organizations.[119] The notion of the citizen journalist (a term that has been given to bloggers and other such individuals who take on the role of journalist, generally without affiliations with any news outlet[120]) or even the Internet were not so much as blips on the Court's radar screen. But the Court based its decision to include "access to media" as an element in establishing a plaintiff as a public figure for a particular reason, that of protecting the "good name" of the private individual.[121] That reasoning still has meaning today, despite the great shifts in the media landscape.

Today, more than thirty years after *Gertz*, millions of people get their news from the Internet[122]—whether from a blog or from a news organization's website, the Internet has become a widespread resource for accessing real and current news. As a realistic component of what comprises media in this era, this needs to be factored in to courts' considerations. When the *Gertz* Court spoke about accessing the media and the ease by which public figures were able to do so, it was addressing in simple terms what was a simple truth: those with a firm grasp on the public's attention through their position as public officials or widely known figures would have the opportunity to garner the press's attention to rebut statements made against them. The Court, seemingly without feeling the need to elaborate, accepted that it was these people who needed less protection from the courts because they had more opportunity to remedy

---

118. *See* Perzanowski, *supra* note 84, at 833.

119. *See* Marc A. Franklin, *Winners and Losers and Why: A Study of Defamation Litigation*, 5 AM. B. FOUND. RES. J. 455, 465 (1980) (defining the context for defamation cases and listing media defendants as those "engaging in newspaper, magazine, or book publishing or in broadcasting"); Max M. Kampelman, *Congress, the Media, and the President*, 32 PROC. ACAD. POL. SCI. 85, 90 (1975) ("There were in 1975 fewer than forty-five cities with two or more competing dailies and about 1,500 cities with a noncompetitive daily press. And each year more and more of these noncompetitive dailies are purchased by the big corporate chains.").

120. *See, e.g.*, Mark Glaser, *Your Guide to Citizen Journalism*, PBS (Sept. 27, 2006), http://www.pbs.org/mediashift/2006/09/your-guide-to-citizen-journalism270.html ("The idea behind citizen journalism is that people without professional journalism training can use the tools of modern technology and the global distribution of the Internet to create, augment or fact-check media on their own or in collaboration with others. . . . Because of the wide dispersion of so many excellent tools for capturing live events—from tiny digital cameras to videophones—the average citizen can now make news and distribute it globally, an act that was once the province of established journalists and media companies.").

121. Gertz v. Robert Welch, Inc., 418 U.S. 323, 341 (1974) (citing Rosenblatt v. Baer, 383 U.S. 75, 92 (1966) (Stewart, J., concurring)).

122. *See, e.g.*, Scott Kirsner, *On the Web, Audience Size Matters*, SAN JOSE MERCURY NEWS, May 27, 2007.

their grievances elsewhere.[123]

It is now, with the media evolution well under way and the continuing trend of more widely accessible online services shifting toward the center of the media stage, that this test does need elaboration. The aim in *Gertz* was to establish parameters as to who would be held to the higher standards invoked by requiring the *New York Times* actual malice test—and necessarily, to limit that group to those actually worthy of the protection, which that test affords. The Court imagined a plaintiff capable of redressing harms that may have resulted from defamatory statements in the media, and that a line would be drawn around those people capable of such access. Those on the other side of the line—private figures unable to access the audience privy to defamatory statements about them—would not be required to meet the heightened standard set forth by the Court.

The scenario may have been quite straightforward to the Court: perhaps it imagined the likes of Johnny Carson facing defamatory statements in the media (that is, in a newspaper or magazine, on the radio, or on television). In order to rebut what was said against him, he would have the capability of accessing a large audience by making a public statement, issuing a press release, holding a press conference, or otherwise addressing the allegations. (He could have, of course, also attempted a defamation claim in court, but would naturally have been required to prove actual malice.) The initial allegation and the subsequent response given by Carson would have drawn similar audiences and similar attention. It was because of this attention that the Court appropriately included this element in its public figure test; the litigation brought by those who have been defamed may only be a secondary concern if they are able to counteract the statement outside of court, and in doing so, to curb the statement's damage.

The *Gertz* Court's position, with such potential scenarios in mind, should now be recognized as one that aimed at encouraging public debate and the introduction of new ideas into the marketplace of ideas—and one that was extremely reluctant to chill any sort of speech that might result from self-censorship. The Court's goal was to protect those private individuals who did not have the means to adequately redress the defamatory words leveled against them because they did not have access to an audience that would effectively serve as a forum for rebuttal. In imagining this person, the Court had in mind someone who could not immediately turn to the same or similarly situated media outlets to address what had been said about him or her.

The test imagined by the Court in *Gertz*—and later in *Hutchinson*—would seem to construe the access to media element of the test by using a

123. Walker, *supra* note 9, at 975.

relatively narrow definition of access to media: not one that encompasses any and all opportunities to be heard by all varieties of audiences, but rather the opportunity to defend oneself to the audience (or a similar such audience) that initially received the damaging information. It is this same reasoning that should guide courts to a new conclusion as to what access to media means; in this age, there is little guarantee that a posting on a blog or social networking site will reach a similarly situated audience[124] that had exposure to the initial defamatory statement. Thus, a court that factors into its analysis the mere existence of a plaintiff's blog or the sheer number of Facebook friends who have access to statements made by the plaintiff online will not be carrying on the intent of the *Gertz* Court.

This is not to say that the audience sizes or compositions must be identical; instead, the point is that the defamed individual should have the opportunity to respond "effectively"[125] to statements made against him. In order to consider a response effective,[126] it must have some impact on the audience of or the effect of the initial defamation. This will simply not be true of a majority of online outlets, considering both the many-to-many mode of communication[127] and the very existence of such a vast number of sources of information available to the average Internet user. With fewer eyes on any particular online source, the defamation plaintiff is not in a position to effectively respond to allegedly defamatory statements by making a posting on just any site online.

It is the courts' responsibility to ensure that the correct lines are drawn between public and private plaintiffs in defamation cases. One of the tools that courts can use is the access to media test—but only if it is appropriately tailored to this era of communication. That means not simply accepting that any and all media outlets and networking sites are sufficient

---

124. Because of the international nature of the Internet, this could mean an audience similarly situated geographically, but it could also mean an audience of roughly equivalent size and composition that had (or could have had) initial exposure to the defamatory content. This Author tends to take the latter view when discussing "similarly situated." The same is true when the Author uses the description of "the same or substantially similar" with regard to the audience.

125. *See* Carr v. Forbes, Inc., 259 F.3d 273, 282 n.2 (4th Cir. 2001) (noting that "a court does not ask whether a defamation plaintiff has ever had access to a media outlet with the same size readership of the allegedly defamatory publication; such an inquiry would effectively prohibit widely read publications from ever commenting on local controversies. Our inquiry is rather whether the evidence demonstrates that the defamation plaintiff had access to channels of effective communication to respond to the allegedly defamatory statements.").

126. Effective is defined as "producing a decided, decisive, or desired effect." THE MERRIAM-WEBSTER'S COLLEGIATE DICTIONARY 397 (11th ed. 2004).

127. Perzanowski, *supra* note 84, at 834 n.9 ("Many-to-many communications media allow users to both contribute and receive information. Blogs, file sharing, and Wikis are among the current many-to-many applications.").

to show that effective channels of communication exist, but rather that the plaintiff have access to media such that he can effectively respond to the statements made against him in such a way as to have a public impact.

VI. CONCLUSION

In this era of mass communication possible with the click of a mouse (or the tap of a button, or screen, on a cell phone), courts cannot shy away from the difficult task of clarifying how the Internet interacts with the law. Defamation cases are certain to encounter these issues sooner rather than later, and when that happens courts will have choices to make. Are they to ignore the dozens of ways every individual can access the media? Are they to find that access sufficient to call anyone with a Facebook account a public figure? Or are they to appropriately consider the widespread use of networking online as an everyday activity of private individuals, placing the correct emphasis on how that individual became a part of the controversy at hand?

The courts should begin by considering the role of the individual within the controversy and how it is that he wound up in such a role (that is, whether she "thrust" herself or was "drawn" into the controversy[128]), and then to look at the nature of any access to media the plaintiff might have. By putting into place clear guidance that lower courts can use with consistency, "commentators will know in advance whether their statements will be protected."[129] This is to say, by understanding the role of access to media to be a lesser factor in the test as compared to the individual's participation in the controversy, the likelihood of confusion over what satisfies the test will be decreased. But at the same time, an understanding of what access to media means will ensure that courts are not tripping themselves up or merely paying lip service to the test. Rather than blindly accepting that any individual with the capability to blog may sufficiently find recourse through the Internet, courts should carefully and closely examine what the make-up of the audience was and how access to a sizeable and geographically similar audience may have tempered and served to mitigate the defamation. By analyzing this component of the plaintiff's status, the court will be giving the appropriate measure of importance to the ability of the individual to redress the harms done against him through the publication of potentially defaming statements.

In order to do so with accuracy and precision, courts must face the fact that the individuals that come before them alleging defamation are likely to be Internet users. It is a simple fact of today's culture that it is difficult to find a person not at least somewhat versed in the ways of the

128. Gertz v. Robert Welch, Inc., 418 U.S. 323, 352 (1974).

129. Walker, *supra* note 9, at 977.

Internet.[130] As such, courts should approach individuals who are well connected and established on social networking sites by properly balancing the prongs of the limited-purpose public figure test. After the court has established what exactly the individual's role in the controversy is, and how that person found herself in that role—that is, whether it was through a voluntary thrust or through involuntarily being drawn into the controversy—the court must then weigh the results of this consideration against the access to media that the individual does or does not have. The court should do so by beginning with considering the media context in which the allegedly defamatory material appeared. The court must consider that this individual's connections on Facebook or Twitter, or followers on her blog, will not be dispositive—and this is particularly true if the defamation took place in the traditional context of mainstream media. Was this a national radio or television broadcast, or a newspaper or magazine article published in a publication with widespread readership? It must then assess whether this person has garnered media coverage of their social networking; or whether the social networking they engage in is merely the private practice of a private individual wishing to stay current and connected with her friends. If it is in fact the latter, the court cannot mistake connections online for the greater requirements of media access—and it certainly cannot do so if the context in which the initial defamation appeared was such that the Internet connections the individual is able to make will do little by way of effectively responding to the much more widely publicized allegations.

In order to reach such conclusions, it is necessary for courts to embrace the current era of social networking. As time progresses, generations will continue to start their Facebook accounts at a younger age and become more savvy with Twitter, not to mention take advantage of sites and tools not yet in existence. As such, it is up to courts to track these changes with tailored decisions, reflecting the truth that individuals are only going to continue to be more connected online, without necessarily being any less in need of the protections imagined by *Gertz* as necessary for private individuals not equipped to successfully respond to defamation on their own.

With such calculated balancing and refined definitions to match the current Internet landscape, courts can successfully maintain the protections that the Supreme Court set forth for private individuals in order to

130. While the Author's eighty-two-year-old grandfather and noncomputer user would be an obvious exception to that generalization, a seventy-three-year-old great uncle of the Author recently recounted the telecommuting he does to continue his consulting work well into retirement.

safeguard that individual's own good name, notwithstanding a plaintiff's large pool of Facebook friends.

Restraining Amazon.com's Orwellian Potential: The Computer Fraud and Abuse Act as Consumer Rights Legislation

Alicia C. Sanders*

I. INTRODUCTION

Amazon.com revealed a capacity for irony when it remotely deleted certain copies of George Orwell's *1984* and *Animal Farm* from its Kindle e-book readers in 2009.[1] In response, two users filed a class action lawsuit

* J.D., Indiana University Maurer School of Law, December 2010; B.A. in Latin and Classical Studies, Purdue University, 2008. The Author would like to thank the members of the *Federal Communications Law Journal* who helped prepare this Note for publication.

1. *The Kindle Content Deletion Flap: Predictions on How Amazon.com Will Respond to the Newly-Filed Class Action*, JUSTIA.COM (Aug. 3, 2009, 3:43 PM),

against Amazon.com.[2] Among several causes of action, the plaintiffs claimed that Amazon.com had violated the Computer Fraud and Abuse Act of 1986 (CFAA) by causing harm to their Kindles without authorization.[3] The lawsuit is one example of the ways that the CFAA has grown since it was enacted.[4] The *Gawronski* lawsuit is a useful case study that shows why the expansion of the CFAA is a good thing for consumers, and why recent restrictions on the Act should not prevent lawsuits like the one Justin Gawronski brought against Amazon.com.

In recent years, the CFAA has been criticized as too expansive.[5] What started as a law to prevent hackers from harming federal computer systems has grown to encompass behavior that is not typically considered hacking. For example, the CFAA is now commonly used in private civil claims of employers against employees who use work computers for unauthorized purposes.[6] The CFAA has strayed far from its original purpose, causing a rise in federal litigation that would not otherwise exist.[7] Recent cases that curtailed employers' remedies for disloyal employees, along with one that declined to extend criminal penalties to a breach of a website's terms of service, mark the beginning of a move toward reining in the scope of the CFAA.[8] In many areas, the new judicial restraint may be justified.[9] But the civil causes of action arising under the CFAA deter some behavior that should be curtailed, like Amazon.com's unauthorized deletion of e-books.

A powerful CFAA can protect consumers from one-sided licensing deals like the purchase of e-books. One of the CFAA's unique benefits over alternative causes of action, like trespass to chattels, is that it creates uniform treatment for Internet-based contracts because the federal system has greater potential for uniformity than state law. The CFAA also has the conceptual advantage of conceiving of e-book ownership as a bargained-for set of rights in a file, not as personal property in the same way that physical

---

http://blawgsearch.justia.com/blawgpost/2009/08/03/kindle-content-deletion-flap- [hereinafter *Content Deletion Flap*].

2. *Id.*; Complaint, Gawronski v. Amazon.com, Inc., No. C09-1084-JCC (W.D. Wash. July 30, 2009), *available at* http://docs.justia.com/cases/federal/district-courts/washington/wawdce/2:2009cv01084/161529/1/.

3. Complaint, *supra* note 2, at paras. 50–57.

4. *See generally* Sarah Boyer, *Computer Fraud and Abuse Act: Abusing Federal Jurisdiction?*, 6 RUTGERS. J.L. & PUB. POL'Y 661 (2009).

5. *See, e.g.*, *id.* at 662.

6. *Id.* at 670.

7. *See id.* at 662–63.

8. Jacqui Cheng, *Disloyal Employees Are Not Hackers, Says Court*, ARS TECHNICA (Sept. 18, 2009, 1:19 PM), http://arstechnica.com/tech-policy/news/2009/09/disloyal-employees-are-not-hackers-says-court.ars?utm_source=rss&utm_medium=rss&utm_campaign=rss; United States v. Drew, 259 F.R.D. 449 (C.D. Cal. 2009).

9. *See, e.g.*, Cheng, *supra* note 8.

books are property.[10] This concept more accurately reflects the reality of the e-book market than do alternative causes of action. Instead of categorically restricting the CFAA to cover only hacking, the CFAA should continue to apply to devices like the Kindle. Any future judicial or statutory restraints on the statute should not constrain e-book purchasers' ability to use the statute to protect themselves from the licensors of e-book files. In addition, a revision of the CFAA expressly creating a cause of action for tethered e-book readers should be added.

II. THE *GAWRONSKI* SETTLEMENT

A. The Amazon Kindle

The Amazon Kindle is a handheld wireless device that displays electronic books that have been purchased from Amazon.com's online Kindle Store.[11] Amazon's Whispernet, the network that tethers the e-reader to Amazon.com and allows downloading of e-books, is accessible from any Kindle without extra fees.[12] Amazon.com also created a free software download for PC that displays Kindle e-books for those who want to read e-books on a traditional computer screen.[13]

The Amazon Kindle has become one of the most popular devices in consumer electronics.[14] Amazon.com announced that the Kindle was its most popular gift item during the 2009 holiday season and that sales of e-books surpassed sales of traditional paper books for the first time in the few days following December 25, 2009.[15] Kindle e-book files prevent users from sharing or transferring the files in ways that violate the Amazon Kindle User Agreement.[16] The Kindle's popularity has caused several

---

10. *Cf.* Michael Seringhaus, *E-Book Transactions: Amazon "Kindles" the Copy Ownership Debate*, 12 YALE J.L. & TECH. 147 (2009) (conceiving of the issue as a dichotomy between e-books as sales or e-books as licenses, preferring, in the end, sales).

11. *Kindle Wireless Reading Device*, AMAZON.COM, http://www.amazon.com/Kindle-Wireless-Reading-Display-Generation/dp/B002Y27P3M/ref=sa_menu_kdp3w3 (last visited Feb. 22, 2011); *Amazon.com: Kindle Store*, AMAZON.COM, http://www.amazon.com/kindle-store-ebooks-newspapers-blogs/b/ref=topnav_storetab_kinh?ie=UTF8&node=133141011 (last visited Feb. 22, 2011).

12. *See Kindle (Latest Generation) License Agreement and Terms of Use*, AMAZON.COM, http://www.amazon.com/gp/help/customer/display.html?nodeId=200505590 (last visited Feb. 22, 2011).

13. *Kindle Wireless Reading Device*, *supra* note 11.

14. Katie Allen, *Amazon e-Book Sales Overtake Print for First Time*, GUARDIAN.CO.UK (Dec. 28, 2009, 7:29 PM), http://www.guardian.co.uk/business/2009/dec/28/amazon-ebook-kindle-sales-surge.

15. *Id.*

16. *Amazon Kindle: License Agreement and Terms of Use*, AMAZON.COM, http://www.amazon.com/gp/help/customer/display.html?nodeId=200144530 (last visited Feb. 22, 2011) [hereinafter Terms of Use].

competitors to release their own e-readers, most notably the Sony Reader and the Barnes & Noble NOOK.[17] Competing models and stores operate in essentially the same way as the Kindle.

B. The Gawronski *Complaint and Settlement*

In July 2009, Kindle owners booted up their e-readers only to see that their copies of George Orwell's *1984* and *Animal Farm* had been remotely deleted from their systems.[18] Certain works of Ayn Rand went missing as well.[19] Other than providing a refund for the missing books, Amazon.com refused to explain the deletions, but later admitted that it had removed the e-books because of copyright problems.[20] According to a statement later made by Amazon.com, the particular e-books had come from Canada, where copyright protection lasts for the life of the author plus fifty years.[21] Orwell's books are public domain in Canada; however, they were still under copyright in the United States, where copyright protection lasts 95 years after the publication date for works published before 1978.[22] If the copyright owner had brought suit against Amazon.com, the company could have faced anywhere between $750 and $150,000 in statutory damages for each infringing work.[23] Once Amazon.com realized its mistake, it immediately stopped selling the e-books in its Kindle Store and retrieved the copies it had already sold.[24] Amazon.com employees remotely accessed the Kindles of each person who had bought the offending e-books and deleted the files, immediately reimbursing the purchase price of the book.[25]

In response, two plaintiffs, Justin Gawronski and Antoine Bruguier, brought a class action lawsuit against Amazon.com in the United States District Court in the Western District of Washington on July 30, 2009.[26] The plaintiffs claimed that Amazon.com harmed them by taking away their property—the e-book files—and also claimed that they had lost valuable electronic marginal notes (such as "remember this paragraph for your thesis"), which were useless without the corresponding files.[27] The

17. *See NOOK*, BARNES & NOBLE, http://www.barnesandnoble.com/nook/index.asp (last visited Feb. 22, 2011); *see also Reader*, SONY, http://ebookstore.sony.com/reader/ (last visited Feb. 22, 2011).

18. Complaint, *supra* note 2, at paras. 1–2.

19. *Id.* at para. 17.

20. Michael D. Scott, *Here Today, Gone Tomorrow . . . Thanks to DRM*, CYBERSPACE LAW, Sept. 2009.

21. *Id.*

22. *See* 17 U.S.C. § 304(a).

23. 17 U.S.C. § 504(c)(1–2).

24. Scott, *supra* note 20.

25. *Id.*

26. Complaint, *supra* note 2, at para. 84.

27. David Johnson, *Kindle Class Action Settlement:* Gawronski v. Amazon *Suit*

complaint alleged that Amazon.com had breached its own terms of service by deleting the files.[28] The complaint also claimed trespass to chattels and a violation of the Computer Fraud and Abuse Act.[29] The plaintiffs identified three classes in the lawsuit: the Kindle Class, which includes all owners of the Kindle, Kindle 2, or Kindle DX; the Big Brother Class, which includes all individuals who had an e-book deleted by Amazon.com; and the Big Brother Work-Product Subclass, which includes anyone who had made marginal notes on an e-book that Amazon.com had deleted.[30]

A few weeks before the settlement on September 3, 2009, Amazon.com had announced that affected users could either replace the book with an identical copy that did not violate the Copyright Act or receive thirty dollars from Amazon.com.[31] Amazon.com reached a settlement with the plaintiffs on September 25, 2009.[32] The settlement was with the two named plaintiffs only, Gawronski and Bruguier, not the entire affected class—the remaining class members therefore still have standing to sue Amazon.com on related grounds.[33] In the settlement, Amazon.com agreed to restore marginal notes for all affected users who had made notes on their previous copies of the deleted books.[34] Amazon.com also agreed that, for all books purchased under the original Terms of Use, which granted

> the Kindle purchaser the "non-exclusive right to keep a permanent copy" of each purchased Work and to "view, use and display [such Works] an unlimited number of times, solely on the [Devices] . . . and solely for [the purchasers'] personal, non-commercial use," it will not remotely delete or modify these books from Kindle devices purchased or being used in the U.S.[35]

There were, however, some exceptions to the settlement:

> Under the settlement, Amazon retains the right to continue to unilaterally delete books from Kindle devices if: (a) the user consents, (b) the user requests a refund or fails to pay for the book, (c) a judicial or regulatory order requires deletion, or (d) deletion or modification is

---

Regarding Amazon.com's Removal of Orwell Works from Kindle Devices Settles, but Leaves Many Questions, DIGITAL MEDIA LAWYER BLOG (Sept. 29, 2009), http://www.digitalmedialawyerblog.com/2009/09/gawronski_v_amazoncom_the_clas_1.html [hereinafter Johnson, *Class Action Settlement*]; *see also* Complaint, *supra* note 2, at para. 54; Scott, *supra* note 20, at para. 8.

28. Complaint, *supra* note 2, at para. 49.
29. *Id.* at paras. 50–57.
30. *Id.* at paras. 53–54.
31. Johnson, *Class Action Settlement*, *supra* note 27.
32. *Id.*
33. *Id.*
34. *Id.*
35. *Id.* (quoting Terms of Use, *supra* note 16).

> necessary to protect the user, the Kindle device or Amazon's network. In addition, Amazon agreed to pay a $150,000 fee to the plaintiffs' counsel, which in turn agreed to "donate a portion of that fee to a charitable organization that promotes literacy, children's issues, secondary or post-secondary education, health or job placement."[36]

Faced with the risk of paying damages to a large class of its customers, Amazon.com also suffered a media backlash.[37] Blogs and news articles (and the complaint itself) pointed to the Big Brother-like deletion.[38] The invasiveness of the deletion seemed to irk Kindle users more than a simple breach of contract would have, especially since Amazon.com had already refunded the purchase price of the books by the time the story had attracted much attention.[39] The realization that Kindle e-books were unlike traditional paper books seemed to be sinking in; the Kindle is not a private bookshelf unreachable by the seller, but is in fact a device constantly tethered to an online seller.[40] A tethered device allows the seller much greater access to its goods than that to which many people are accustomed. Devices like the Kindle, assuming their popularity will continue as their prices drop and their availability rises, raise new legal issues for book buyers and are changing not only the way people read books, but also the way people own books. Applying the CFAA to unauthorized use of tethered devices could help regulate behavior and protect people's rights as book owners.

C. Historical Background: Expansion and Restriction of the CFAA

The Computer Fraud and Abuse Act of 1986[41] was originally designed to impose criminal penalties on computer hackers who caused harm to federally owned computers, but the statute has been amended multiple times since 1986.[42] Today, the CFAA prohibits unauthorized access to "protected computers," which includes any computer connected to the Internet or any computer that has moved in interstate commerce.[43] Perhaps the most significant amendment to the CFAA inserted a provision that created a private cause of action in addition to the already-existing criminal penalties.[44] This private cause of action is the one used in the *Gawronski* suit.

---

36. *Id.* (quoting Terms of Use, *supra* note 16).
37. *See, e.g.*, *Content Deletion Flap*, *supra* note 1.
38. *See, e.g.*, *id.*; Complaint, *supra* note 2, at para. 3.
39. *See* Johnson, *Class Action Settlement*, *supra* note 27.
40. Scott, *supra* note 20.
41. 18 U.S.C. § 1030 (2006).
42. Boyer, *supra* note 4, at 665–66.
43. 18 U.S.C. § 1030(e)(2).
44. *Id.* at § 1030(g).

A cause of action brought under the CFAA requires a showing of damage and loss.[45] Damage is "any impairment to the integrity or availability of data, a program, a system, or information," while loss includes any reasonable cost to the plaintiff, "including the cost of responding to an offense, conducting a damage assessment, and restoring" the damaged system to its prior condition.[46] Loss also can include "any revenue lost, cost incurred, or other consequential damages incurred because of interruption of service."[47] This broad definition of loss is significant because anyone bringing a civil action may obtain damages only if he or she can show one of the following: loss of at least $5,000 per year, medical harm, physical injury, threat to public health or safety, or damage to a government computer.[48] Since the calculation of loss can include damage assessments, which are easy to incur after the fact for the sole purpose of bringing suit, this requirement is easy to meet.

The civil provisions of the CFAA have been used primarily by employers attempting to block employees and former employees from using company data to start competing companies or sell company data to competitors.[49] Many of these suits could have been handled under traditional contract law in state court, but the plaintiffs have used the CFAA as a way to move into federal court.[50] Recently, however, courts have tried to reduce the scope of the CFAA. For example, in *LVRC Holdings LLC v. Brekka*, the Ninth Circuit declined to hold that a former employee was acting "without authorization" when he e-mailed his employer's financial statements to his home computer and later used those statements to start a competing company.[51] According to the court, because the employee was using his own password and credentials to log in to the company computer, he was acting with authorization and could not be held liable under the CFAA.[52]

The CFAA's criminal cause of action has also been restricted as part of the general trend. The defendant in the highly publicized cyberbullying case *United States v. Drew*, also known as the "MySpace Mom" case, was acquitted in 2009 via a judgment notwithstanding the verdict.[53] The

---

45. *Id.*

46. *Id.* at § 1030(e)(8), (11).

47. *Id.* at § 1030(e)(11).

48. *Id.* at § 1030(c)(4)(A)(i); § 1030(g).

49. Boyer, *supra* note 4, at 670.

50. *See id.*

51. LVRC Holdings LLC v. Brekka, 581 F.3d 1127, 1129 (9th Cir. 2009); Cheng, *supra* note 8.

52. *LVRC Holdings*, 581 F.3d at 1133, 1135; Cheng, *supra* note 8.

53. United States v. Drew, 259 F.R.D. 449 (C.D. Cal. 2009). Lori Drew created a false profile on the social networking website MySpace.com. *Id.* at 452. Pretending to be a

prosecution in *Drew* argued that an intentional breach of a website's user agreement, which prohibited the creation of a false profile, amounted to a criminal violation of the CFAA.[54] The court refused to extend criminal penalties to a violation of a website's terms of service, reasoning that the prosecution's theory of the case far surpassed the original purpose of the CFAA—namely, to prevent hackers from damaging computer systems.[55]

Like the prosecutors in the *Drew* case, the *Gawronski* plaintiffs were also asking a court to apply the CFAA to a defendant who is not a typical hacker. However, Amazon.com as a defendant caused damage that was more like the anticipated harm from a hacker than the harm Drew caused. While the CFAA has been rightly restricted in some contexts, like cyber-bullying, the restrictions need not extend to any activity that was not strictly contemplated by the original drafters of the statute. In fact, the CFAA gives some plaintiffs benefits that other causes of action do not.

III. THE ADVANTAGES OF THE CFAA OVER OTHER REMEDIES

The plaintiffs in *Gawronski* listed several causes of action in addition to the CFAA, including breach of contract, violation of a Washington consumer protection law, and trespass to chattels.[56] Critics of the modern CFAA have said that it is being used to cover too many situations, and that it amounts to an abuse of federal jurisdiction when plaintiffs use it as an excuse to get out of state court and into federal court.[57] As this Note will show, the CFAA is a useful cause of action in cases like *Gawronski*, more so than the other remedies available to plaintiffs.

A. Analysis of Gawronski's CFAA Claim on the Merits

First, it is important to show as a threshold matter that the plaintiffs would have been likely to succeed on their CFAA claim. In essence, Amazon.com would have likely lost on the merits of the CFAA cause of action because the plaintiffs could have proved that Amazon.com acted without authorization when it deleted the e-books. The Terms of Use guaranteeing a permanent copy would be proof of the bounds of

---

teenage boy, Drew communicated with her daughter's classmate, Megan Meier, who had not been getting along with Drew's daughter. *Id.* Drew made Meier believe that the fictitious boy liked her, and then made her believe that the boy had lost interest in her by saying "the world would be a better place without her . . . ." *Id.* (internal quotations omitted). Meier committed suicide shortly afterward. *Id.* Technically, Drew had not committed a crime other than possibly a violation of the CFAA. *Id.* at 451.

54. *Id.* at 451, 467.

55. *Id.* at 461, 465.

56. Complaint, *supra* note 2, at paras. 50–83.

57. *See, e.g.*, Boyer, *supra* note 4, at 670–71.

Amazon.com's authorization to access users' Kindles.[58] Although the plaintiffs would have had to prove that they had incurred a loss of $5,000, they (and future potential plaintiffs) could meet this requirement because the CFAA has such a broad definition for "loss." In particular, the complaint alleges the following:

> Amazon violated 18 U.S.C. § 1030(a)(5)(A)(i) by knowingly causing the transmission of a command to delete content to Plaintiffs' Kindles, which are protected computers as defined in 18 U.S.C. § 1030(e)(2)(B) because they are used in interstate commerce and/or communication. By deleting content from Plaintiffs' Kindles by way of remote deletion, Amazon intentionally caused damage without authorization to Plaintiffs' Kindles. . . . Amazon violated 18 U.S.C. § 1030(a)(5)(A)(iii) by intentionally accessing Plaintiffs' Kindles, protected computers, without authorization, and as a result, caused damage to Plaintiffs' Kindles by remotely deleting content stored on them.[59]

Amazon.com has publicly admitted to removing particular e-books from users' devices, as asserted by the plaintiffs, so the plaintiffs would have been able to prove at least that much of a CFAA claim.[60] One potential sticking point for the court may have been whether a Kindle was in fact a computer as defined in 18 U.S.C. § 1030(e)(1), which defines a computer as "an electronic . . . device performing logical, arithmetic, or storage functions . . . but such term does not include an automated typewriter or typesetter, a portable hand held calculator, or other similar device."[61] Amazon.com may have been able to argue, therefore, that Kindle is more like a calculator than a computer, given that it does not have a full computer's functions, but that argument likely would not have succeeded. Because Kindles are sophisticated enough to have a web browser, limited word processing functions for marginal notes, and the ability to download and store files, the Kindle is more like a computer than a calculator.

For the damages requirement, the CFAA requires plaintiffs to prove a loss of at least $5,000 to recover under the Act.[62] The plaintiffs' argument is split here between the two specific plaintiffs and two different classes of plaintiffs:

> Plaintiffs and Big Brother Class members suffered damages even though Amazon refunded the cost of the purchased content to Plaintiff and Big Brother Class members because: 1) they now have to replace the deleted content at a higher cost; and, 2) their Kindles were reduced in value by way of the deletion in that an electronic reading device that enables Amazon to remotely delete content is less valuable than one not subject to unconsented remote deletions of content.

---

58. Terms of Use, *supra* note 16.
59. Complaint, *supra* note 2, at paras. 51–52.
60. Johnson, *Class Action Settlement*, *supra* note 27.
61. 18 U.S.C. § 1030(e)(1) (2006).
62. *Id.* at § 1030(a)(5)(B)(i).

> Plaintiff Gawronski and the Big Brother Work-Product Subclass suffered damages because they created content on their Kindles within the purchased content that Amazon deleted. The remote deletions rendered their work-product useless and worthless because their work-product necessarily was linked to the deleted purchased content. As a result, Plaintiff Gawronski and the Big Brother Work-Product Subclass must expend further resources and effort in order to recreate their now useless work-product.
>
> As a result of these takings, Amazon's conduct has caused a loss to one or more persons during any one-year period aggregating at least $5,000 in value in real economic damages.[63]

Loss is defined in the statute as,

> any reasonable cost to any victim, including the cost of responding to an offense, conducting a damage assessment, and restoring the data . . . or information to its condition prior to the offense, and any revenue lost, cost incurred, or other consequential damages incurred because of interruption of service . . . .[64]

The plaintiffs and the Big Brother Class claimed damages because they had to replace the deleted content at a higher cost because, presumably, the refund is less than the purchase price they would pay at an American-owned e-store where the copyright on Orwell's work is still valid. This difference in price fits clearly within the CFAA's definition of loss, as it represents the cost incurred by restoring the data to its prior state, but the loss in value of their devices is harder to evaluate. The difference in value between a Kindle that ensures permanent copies and one that does not make that guarantee is a subjective value, or is at least difficult to determine. Recalling that only aggregate loss matters under the statute, the $5,000 threshold seems within reach for a lawsuit with such a large class. The aggregate value of the e-books would likely succeed $5,000 in purchase price alone. The marginal notes would likewise have value, even though it would be subjective and difficult to measure, but even a nominal amount spread across a class would likely reach $5,000. The *Gawronski* plaintiffs would have, in all likelihood, been able to make a solid prima facie case for a CFAA violation.

B. Comparison of the CFAA Claim and Other Remedies

The CFAA has several advantages over other available remedies. First, because the federal system has greater potential for uniformity than state law does, the CFAA allows for more uniform treatment of Internet-based contracts. Compared to trespass to chattels, the CFAA more accurately reflects the reality of the e-book market—the CFAA has the conceptual advantage of conceiving of e-book ownership as a bargained-for

63. Complaint, *supra* note 2, at paras. 53–55.
64. 18 U.S.C. § 1030(e)(11).

set of rights in a file, not as personal property in the form of physical books. A too-broad limitation on the CFAA would deprive plaintiffs of the remedies the CFAA affords them. A revision of the CFAA expressly creating a cause of action for Internet-based contracts, like the one Amazon.com has with its Kindle, should be added. Such a revision would be a suitable middle-ground that would allow for restrictions to the CFAA in areas like employment law but would preserve the protections for circumstances like those in the *Gawronski* complaint.

1. Breach of Contract and State Law

The Amazon.com Terms of Use is a contract that governs the user's rights in his or her e-books.[65] The *Gawronski* complaint alleged breach of contract because in the Kindle's Terms of Use, Amazon.com had granted Kindle users a permanent copy of their e-books.[66] Such a contract claim, without the accompanying CFAA cause of action, would have been tried in a Washington court. The breach of contract claim, whether in state or federal court, would likely have had the same result. Because the language in the Terms of Use did not provide Amazon.com a right to delete purchased books from the devices, Amazon.com's conduct amounted to a breach of contract. Both federal and state courts would have come to the same conclusion.

Expecting users to resolve their disputes regarding Internet contracts for e-books in state court creates the problem of lack of predictability. Uniformity is one advantage that the CFAA lends to lawsuits related to e-book readers. Because the CFAA allows plaintiffs who have a dispute involving unauthorized access to their computers to move into federal court, the CFAA makes it easier for both users and companies who sell tethered devices to understand the rules of their dealings. Specifically, the CFAA could provide a clear, uniform rule that no unauthorized access would be permitted to any tethered device. Such a federal law would prevent the company from taking advantage of any specialized state rules that may be difficult for an individual user to discover—for example, a state rule that inferred authorization so long as the users were compensated for a deleted file. That type of state rule, if given time to develop in a particular jurisdiction, could add a hidden term into an otherwise transparent contract. The problem would be complicated if there were many different e-book stores online, so that a user could potentially be agreeing to contracts governed by several different state contract laws. While such a system would still be workable, having a strong federal rule governing Internet transactions makes it easier for an Internet user to

65. Terms of Use, *supra* note 16.

66. Complaint, *supra* note 2, at paras. 46–49.

understand exactly what laws apply to a transaction. Admittedly, circuit splits could cause unpredictability, but not to as great an extent as with state law. In any event, plaintiffs could move into federal court anyway using diversity jurisdiction, since most users would not live in the same states as the companies. For Internet contracts especially, then, federal laws may be the best way to regulate companies' behavior.

The same reasoning applies to the cause of action for violation of Washington state law.[67] The *Gawronski* complaint alleges a violation of the Washington Consumer Protection Act (CPA), which provides that " [u]nfair methods of competition and unfair or deceptive acts or practices in the conduct of any trade or commerce are . . . unlawful."[68] The existence of such a law is not guaranteed in all jurisdictions, and a company might draft its contract so as to be governed by the laws of a less consumer-friendly state. The availability of a federal cause of action ensures that an Internet user has a better idea of what remedies are available to him or her. The user will be able to draw upon various state remedies, but no matter what, the federal law will still apply and be available as well.

Of course, the CFAA is effective in preventing abuse by e-book companies and other companies that offer tethered devices only to the extent that accessing and deleting files on the tethered devices is unauthorized by the contract. In the *Gawronski* case, Amazon.com's Terms of Use unambiguously offered a permanent copy. But there is nothing that prevents Amazon.com or other tethered-device companies from contracting around the problem in the future. Although Amazon.com has since announced that it will not delete e-books in the same manner again, it has reserved the right to delete e-books if it deems the deletion necessary to protect its network.[69]

Essentially, Amazon.com's promise not to delete e-books from users' Kindles applies only to circumstances similar to those in the *Gawronski* case. In its settlement, Amazon.com reserved its right to delete e-books if it deems the deletion necessary.[70] Furthermore, Amazon.com can change its Terms of Use in the future to allow deletion of e-books that have copyright problems, as could other competing companies like Barnes & Noble.As a result, the CFAA's main weakness in the e-book reader scenario is that companies can easily draft contracts around the problem by writing express

---

67. *Id.* at paras. 77–83.

68. WASH. REV. CODE § 19.86.020 (West 1961).

69. Peter Kafka, *Amazon: We Won't Delete Your Kindle Books Unless We Need to Delete Your Books*, ALL THINGS DIGITAL (Oct. 1, 2009, 8:10 AM), http://mediamemo.allthingsd.com/20091001/amazon-we-wont-delete-your-kindle-books-unless-we-need-to-delete-your-books/?mod=ATD_sphere.

70. Johnson, *Class Action Settlement*, *supra* note 27 (citing Terms of Use, *supra* note 16).

authorization into the contracts. To the consumer, purchasing an e-book seems to be largely the same experience as purchasing a physical book; the final product—certain words on a page or screen—is the same regardless of how it is displayed. But purchasing e-books in an online store is different from buying physical books at a store. In a physical bookstore, there is no written contract, only an oral agreement at the point of sale. Consumer rights in that book are governed by the UCC and federal copyright law.[71] Most stores allow returns, and the likelihood of litigation is very low. Furthermore, owners of physical books are able to resell their books in a secondary market under the "first sale" doctrine.[72] In contrast, Kindle customers are not owners of books at all, but licensees bound by the Kindle Terms of Use.[73] Most Internet users do not read the full terms of use in online contracts, even though those agreements are, for the most part, binding. Because the user is less likely to read and understand the contract, it is easy for companies to draft the contracts in ways unduly favorable to them. Many companies, when drafting the contracts, have the ability to select and designate a venue or governing state rules that the user is not in any position to bargain for or against.

This Note proposes that the CFAA should be expanded in part, especially if it is curtailed in other areas, like employer-employee disputes. For example, Congress could add a definition for tethered devices so that it includes devices like the Kindle and the NOOK. The amendment could simultaneously create a separate cause of action for deleting files contained on the device without express consent. The CFAA could require renewed consent for each file and ban a blanket authorization by contract.

Such an amendment to the statute would have certain advantages. Most important, it would protect users' property rights in their e-books. Cementing property rights in an electronic medium is especially important because most electronic media are dispensed over the Internet by large companies like Amazon.com. A typical Internet user has almost no bargaining power compared to Amazon.com or Barnes & Noble. It would be all too easy for large, web-based companies to create a "Big Brother" custom where books can be deleted at the whim of the sellers, so long as such a blanket provision was in their terms of service. Preempting the

---

71. Seringhaus, *supra* note 10, at 150.

72. *Id.* at 160.

73. *Id.* at 149–51. Seringhaus argues that e-book transactions, though stylized as licenses, should be reinterpreted by courts as sales. *Id.* This issue has not gone before courts. Amazon.com would certainly prefer to keep the transactions categorized as licenses because that categorization preserves its ability to delete e-books or block users from purchased content as it deems necessary. Either way, the CFAA would still apply to Kindle transactions because the CFAA prohibits unauthorized access to the Kindle device, rather than the e-books stored on the devices.

dangers of new technology is in the spirit of the original CFAA, which was created because of fears related to computer hackers. Today, people might more likely be afraid of the abuses of powerful companies.

This dystopian possibility is real enough in consumers' eyes, hence the media backlash from Amazon.com's actions. If Amazon.com had accidentally sold a different, less dystopian book—say, a Harlequin romance—without the proper copyright permission and then deleted it, there is a significant chance that the deletion would not have made news at all. Amazon.com had compensated the users, after all, and the most loss any person experienced was missing marginal notes. The unlucky coincidence that it was George Orwell's work that was deleted touched on a raw nerve that elevated the problem from economics to a matter of principle.

2. Trespass to Chattels

The increasing use of computers in the 1990s led to an expansion of the common law doctrine of trespass to chattels to create civil liability for causing harm through unauthorized use of computers.[74] Originally, trespass to chattels (commonly known as "conversion's little brother") was an infrequently used doctrine that provided a remedy for interference with chattels that caused harm, but not enough harm to constitute conversion.[75] Unlike trespass to real property, a prima facie case for trespass to chattels requires actual harm to the chattel; unlike conversion, a person liable for trespass to chattels owes only actual damages, not the full value of the chattel.[76]

To prove a trespass to chattels case, the plaintiff must show that the defendant intentionally dispossessed another person of the chattel, or used or intermeddled with a chattel in another's possession.[77] "Intermeddling" means intentionally causing physical contact.[78] The plaintiff must prove that he or she suffered damage because of one of the following: the plaintiff was dispossessed of the chattel; the chattel's condition, quality, or value was impaired; the plaintiff was deprived of the use of the chattel for a substantial time period; or the plaintiff suffered bodily harm or harm to some person or thing in which the plaintiff has a legally protected interest.[79] Finally, the intermeddling must be more than nominal.[80]

---

74. *See* Shyamkrishna Balganesh, *Common Law Property Metaphors on the Internet: The Real Problem with the Doctrine of Cybertrespass*, 12 MICH. TELECOMM. & TECH. L. REV. 265, 266–69 (2006).

75. *Id.* at 279.

76. *Id.* at 279–80.

77. RESTATEMENT (SECOND) OF TORTS § 217.

78. *Id.* at cmt. e.

79. *Id.* at § 218.

The elements of a trespass to chattels claim as applied to computers began to change with the case *eBay, Inc. v. Bidder's Edge, Inc*., in which the court considered the risk of future potential harm when analyzing the severity of the intermeddling.[81] Defendant Bidder's Edge was using a software robot to continuously search eBay to populate an auction aggregation site.[82] eBay asked Bidder's Edge to stop the continuous searching, and eBay also attempted to block the robots, but Bidder's Edge used proxy servers to bypass the IP address blocks.[83] *Bidder's Edge* is significant for two reasons: first, the court found that although websites are open to the public, Bidder's Edge acted without authorization because eBay's physical servers were private property—chattels—with which Bidder's Edge had interfered; and second, the court found that Bidder's Edge had proximately damaged eBay's chattels by using some of the server's resources as bandwidth.[84] The court admitted that Bidder's Edge did not harm or noticeably slow eBay's system, but more than nominal damage would have occurred if other companies had decided to join Bidder's Edge in continually crawling eBay's website.[85] This holding expanded the common law trespass to chattels doctrine because now nominal damage was sufficient, so long as the plaintiff could invoke the fear of encouraging others to copy the defendant.[86] Because the court considered possible future harm, the threshold for damage in trespass to chattels cases was, by that time, almost nonexistent.

This low threshold for damage allowed the *Gawronski* plaintiffs to add trespass to chattels as one of their causes of action. Even if they could not prove later on that they had suffered $5,000 worth of damage—as required to recover under the CFAA—they could at least advance a trespass theory. Furthermore, there is a possibility that trespass to chattels as redefined in *Bidder's Edge* could include more than just the value of the single missing e-book and the loss of the marginal notes. The future potential harm calculation from *Bidder's Edge* could give plaintiffs more than actual damages; it could include the value of loss from future potential deletions. However, this extra benefit is unlikely because the *Gawronski* facts are based on a loss of a particular file, not from use of system resources.

The tort case for the *Gawronski* plaintiffs would likely succeed. The

---

80. *Id.* at cmt. e.
81. eBay, Inc. v. Bidder's Edge, Inc., 100 F. Supp. 2d 1058 (N.D. Cal. 2000).
82. *Id.* at 1060–62.
83. *Id.* at 1062–63.
84. *Id.* at 1071–72.
85. *Id.*
86. *See id.*

plaintiffs could easily prove that Amazon.com intentionally dispossessed the plaintiffs of the chattel (the e-book). Under that theory, though, the status of the marginal notes is not as certain as it would be under a breach of contract or a CFAA claim. The value of the marginal notes could be either a user modification to the e-book that increased the chattel's value, or it could be a separate chattel all on its own as a separate user file. Under the latter theory, the marginal notes would have been "intermeddled with" rather than dispossessed, because the notes remained on the Kindle but were useless without context. Either way, the plaintiffs likely could establish a prima facie case for trespass to chattels.

Trespass to chattels as a remedy for unauthorized deletions from a tethered device is susceptible to the same problems as state-based contract claims. The common law nature of tort remedies means that different jurisdictions may treat trespass to chattels differently. One court may decide that the common law requirement of actual damage to a device should apply to tethered devices. How much damage must be done to a device, and what exactly constitutes damage, are also variables subject to reinterpretation by state courts. For example, one court could reasonably hold that deletion of a single file with immediate, reasonable compensation is merely nominal damage, while another could reasonably conclude that the same action is actual damage to a chattel. The distinction could turn on what exactly the court determines to be the chattel. If the chattel is the device, then a compensated deletion could seem minor, but if the chattel is determined to be the e-book file itself, then the damage seems more significant. There could be uncertainties in a single jurisdiction where the issue has not been decided, and also uncertainties between jurisdictions when an Internet user does not have the resources or bargaining power to predict or control the eventual venue of a dispute.

As discussed above, a further amendment to the CFAA that expands the statute to explicitly cover tethered devices like e-readers would bring the statute more in line with current technology. The CFAA was not designed with Kindles or other tethered devices in mind, but fixing an outdated statute by amending it to cover new technology seems a better solution than stripping the CFAA of its current uses and focusing solely on hacking. While hacking remains a concern, new technology has emerged such that consumer rights are now a matter pressing on the public's consciousness. An expanded statute would be able to better protect consumers who face technologies that offer limited rights in media.

Similarly, trespass to chattels was designed to cover personal property before computer files existed, so its underlying rationale does not necessarily apply to virtual property. Courts' reinterpretation of the old doctrine, as in *Bidder's Edge*, shows that new technology has changed the

meaning of chattel. Or, more accurately, courts have changed the doctrine to remedy inequities for which there was no appropriate modern relief. Rather than twisting an almost-forgotten tort to protect consumers of a new device, amending the CFAA is a better, more thoughtful response because drafting legislation gives Congress more control over how the law will be applied. Allowing new doctrines to develop in the common law could lead to different results in different jurisdictions. A differing jurisdictional approach has little to do with the average Internet user's expectations that the law will apply similarly to contracts with Amazon.com and with Barnes & Noble.

Significantly, the CFAA has a conceptual advantage over trespass to chattels, because it better mirrors how the market currently treats e-books. This is because trespass to chattels does not quite conceptually fit what is happening when someone accesses a computer without authorization and deletes a file. Trespass to chattels, as mentioned, requires the chattel itself to be harmed, and it is not even clear whether a file will be considered a chattel—as a physical book would be. *Intel Corporation v. Hamidi* made it clear that harm to a particular chattel is required for the tort.[87] In *Hamidi*, the harm a former employee did to a company was harm to the company, not to the physical computers.[88] In the context of e-books, the only thing that is inarguably a chattel is the device itself, not the file. Deleting a file does not necessarily physically damage a computer.

The problem with what Amazon.com did to its Kindle users was that it accessed the Kindles and changed content without the users' consent. The concept behind the CFAA is that two parties assign each other a set of permissions, and the parties must abide by those permissions. Any unauthorized access that does enough harm may give rise to a civil cause of action. The key is that the CFAA treats e-books as a nexus of different rights: the company reserves the right to remove the e-book if it is not paid for, and the user retains the right to a permanent copy so long as it is paid for. Users may think of the books as their "personal property" in the same way that they may think of books on their shelves as personal property, thus subject to the doctrine of trespass to chattels. But tethered devices are *sui generis* because they operate by giving users permission to view a file, not by actually allowing a user to physically possess the product.

The "permissions" treatment of e-books under the CFAA is more in line with the legal landscape of e-books. This parallel makes the CFAA a useful remedy because it conceptually fits the problem. While the *Gawronski* plaintiffs certainly had a strong chance of succeeding on a trespass to chattels claim, the CFAA more accurately reflects the legal

87. Intel Corp. v. Hamidi, 71 P.3d 296, 300 (Cal. 2003).
88. *Id.*

status of e-books. Any revision of the CFAA should explicitly account for e-books as something valuable to protect.

IV. CONCLUSION

Although the dispute quickly settled, the *Gawronski* lawsuit remains a useful case study that shows why applying the CFAA to situations involving e-books is important to protect consumers. It is enlightening to examine this case because the CFAA has been criticized as too expansive. The original intent of the CFAA was to prevent hackers from harming federal computer systems, but now it is applied to many different situations, including employment suits. Recently, courts have begun to restrict the CFAA to return the statute to its original purpose—but courts should be careful not to go too far. If the CFAA were appropriately amended, it could not only avoid such restrictive intrepretations by courts, but it also has the potential to protect consumers from one-sided licensing deals like those found in the current e-books market.

As it currently exists, the CFAA provides several advantages to consumers that other causes of action do not. First, the CFAA provides a way for consumers to access federal courts, which can ensure a more uniform treatment of Internet-based contracts than does state law. The CFAA also has the conceptual advantage of conceiving of e-book ownership as a bargained-for set of rights in a file, not personal property in the same way that physical books are property. This concept more accurately reflects the reality of the e-book market than common law approaches.

To take further advantage of these benefits, a revision of the CFAA expressly creating a cause of action for tethered e-book readers should be added. Such an amendment would prevent companies from attempting to contract around the CFAA. Furthermore, allowing plaintiffs to use the CFAA in e-book suits would ensure that the interpretation of contracts governing e-book purchases would not unreasonably favor the rights of the seller over those of the e-book purchaser.

Technology is always changing, and the law must stay ahead of the curve, or, at the very least, try to keep up. The *Gawronski* complaint and other cases show that the CFAA, much more recently developed than trespass to chattels, is much more useful to consumers of modern technology. However, even the CFAA significantly predates the advent of tethered devices such as e-book readers, and could therefore use some amending to better apply to instances like Amazon.com's Orwellian deletions and similar problems that may arise in the near future.

Television for All: Increasing Television Accessibility for the Visually Impaired Through the FCC's Ability to Regulate Video Description Technology

Joshua S. Robare*

* J.D. Candidate, Indiana University Maurer School of Law, May 2011; B.A. in Mathematics and Political Science, Alma College 2008.

I. INTRODUCTION

Many people take for granted the relatively simple action of sitting down at the end of the day and turning on the television. They can relax and let wave after wave of sounds and images wash over them, relieving their stress and tension. Regardless of whether the dial is set to sports or a soap opera, news or nonsense, drama or comedy, television is something that has become part of the fabric of almost every person's life. However, there are a significant number of people in the United States who are unable to enjoy this activity. The U.S. judicial system has created a "have and have-not" dichotomy when it comes to persons with disabilities enjoying television. As a result of the D.C. Circuit's 2002 decision in *Motion Picture Association of America, Inc. v. Federal Communications Commission*, the FCC is allowed to regulate closed captioning, forcing television manufacturers and broadcasters to implement technology that will allow deaf Americans to enjoy television more fully.[1] In the same decision, the court found that the FCC did not have power to promulgate regulations regarding video descriptions[2] that would allow blind and seeing-impaired Americans to have a more complete television experience, similar to those without a disability.[3]

The Survey of Income and Program Participation is a national survey that collects data on a regular basis to identify the percentage of the American population with hearing loss or deafness.[4] This survey has found that "1 in 20 Americans are currently deaf or hard of hearing. In round numbers, nearly 10,000,000 persons are hard of hearing and close to 1,000,000 are functionally deaf."[5] Americans who suffer from hearing loss

1. Motion Picture Ass'n of Am., Inc. v. FCC, 309 F.3d 796 (D.C. Cir. 2002).

2. Video descriptions help the seeing impaired have a more complete entertainment experience by articulating the action taking place on screen during breaks in a program's natural audio track; they describe key visual elements and action that cannot be picked up by listening to the dialogue alone. JACLYN PACKER & CORINNE KIRCHNER, WHO'S WATCHING? A PROFILE OF THE BLIND AND VISUALLY IMPAIRED AUDIENCE FOR TELEVISION AND VIDEO vii (1997), *available at* http://www.afb.org/Section.asp?SectionID=3&TopicID=135&DocumentID=1232#intro. Important elements such as the movement of a character on the show, what a scene looks like, and nuanced character interactions would all be captured by video descriptions.

3. *Motion Picture Ass'n*, 309 F.3d at 807.

4. Ross E. Mitchell, *How Many Deaf People Are There in the United States? Estimates from the Survey of Income and Program Participation*, 11 J. OF DEAF STUD. & DEAF EDUC., 112, 112 (2006).

5. *Id.*

or complete deafness have become the "haves" when it comes to the FCC's ability to provide a satisfactory television experience; since 1993, the FCC has taken steps to make sure that closed captioning[6] is available to as many Americans as possible.[7] The ability of the FCC to help those with hearing problems is in stark contrast to its ability to help those with seeing problems through the use of video descriptions. Allowing the FCC to regulate video descriptions would help the 25.2 million Americans who have reported problems seeing, many of whom are unable to see at all.[8]

This Note argues that the time has come to take action and increase availability of video descriptions. Part II of this Note examines the court's decision in *Motion Picture Association of America*. It considers both the views of the visually impaired community and the entertainment industry leading up to the court's decision. Part II further examines the major justifications that the court used in reaching its decision. Part III begins by exploring why the lack of video description technology is a problem. As a result of the decision in *Motion Picture Association of America*, closed captioning and video description have been placed in juxtaposition to one another. This Section explores the divergence in treatment between the two and whether those differences justify their disparity in treatment under the current regulatory scheme. The Section ends by looking at the changes available for video description technology as a result of the digital transition and how the change affects the ease of implementing the technology. Part IV of this Note explores two possible solutions to the problem. The first solution requires the government to provide brief financial support to the video description industry in an effort to make it self-sustaining. The second solution suggests passing legislation similar to the proposed Twenty-First Century Communications and Video Accessibility Act, which aims to restore the FCC's ability to regulate video descriptions.

II. REACHING THE DECISION

Several important factors led to the decision in *Motion Picture Association of America* The 1996 amendments to the Communications Act of 1934 started a chain reaction of events within the FCC. It was not until the decision in *Motion Picture Association of* America that key questions about video descriptions were answered. The court had to look not only at

---

6. Closed captioning displays the words being spoken on screen as text so persons with hearing disabilities can read what actors are saying and still enjoy a television program.

7. *FCC Consumer Facts: Closed Captioning*, FED. COMM. COMMISSION, http://www.fcc.gov/cgb/consumerfacts/closedcaption.html (last visited Feb. 22, 2011).

8. *Facts and Figures on Adults with Vision Loss*, AM. FOUND. FOR THE BLIND, http://www.afb.org/Section.asp?SectionID=15&TopicID=413&DocumentID=4900 (last visited Feb. 22, 2011).

how video descriptions were made, but also at the inherent power that the FCC was granted by Congress to carry out its duties.

A. The Effects of the 1996 Telecommunications Act

The holding in *Motion Picture Association of America* was largely influenced by the 1996 Telecommunications Act. The Act, which amended the Communications Act of 1934 changed the FCC's control over programming accessibility by adding provisions about both closed captioning and video descriptions.[9] The first five subsections, refer to the FCC's powers relating to closed captioning; only the last two deal with video descriptions.[10] The provisions relating to closed captioning required the FCC to make a full report to Congress, create regulations specifying actions that the television industry needed to make to implement closed captioning technology, and create a timeline specifying when the new technology needed to be in place.[11] The last two subsections dealing with video descriptions were extremely brief in comparison to their closed captioning counterparts.[12] The Act merely defined the term video description and called on the FCC to make a report and present it to Congress.[13]

Examining the congressional record of the Act does little to clear up whether Congress intended to grant the FCC equal power to regulate closed captioning and video descriptions. With regard to video descriptions, the House version of the bill included the following language:

> The report shall assess appropriate methods for phasing video descriptions into the marketplace, technical and quality standards for video descriptions, a definition of programming for which video descriptions would apply, and other technical and legal issues. Following the completion of this inquiry the Commission may adopt regulations it deems necessary to promote the accessibility of video programming to persons with visual impairments.[14]

The last sentence of this excerpt would seem to support the contention that Congress did not intend for there to be disparate treatment of closed captioning and video description, but instead wanted the FCC to be able to create and enforce rules and regulations regarding both. This is further supported by the concluding lines in the congressional record on the topic, which read: "It is the goal of the House to ensure that all Americans

---

9. Telecommunications Act of 1996, § 713(a)–(g), Pub. L. No. 104-104, 110 Stat. 56 (codified at 47 U.S.C. § 613(a)–(g)).

10. *Id.*

11. *Id.*; *see also* Motion Picture Ass'n of Am. v. FCC, 309 F.3d 796, 799 (D.C. Cir. 2002).

12. *See* 47 U.S.C. § 613(f)–(g) (1996).

13. *Id.*

14. 142 CONG. REC. 1441, 1955 (1996).

ultimately have access to video services and programs, particularly as video programming becomes an increasingly important part of the home, school and workplace."[15] When the House and Senate were working to reach the final version of the bill, the conference committee excluded the language about the FCC's power to create and enforce regulations regarding video descriptions,[16] although the record is unclear as to why.

Despite the statutory differences, the FCC initially attempted to treat video descriptions and closed captioning the same. The FCC was acting under the belief that Congress had passed the bill hoping to bring universal access to television, regardless of disability.[17] After the passage of the Act, the FCC required cable operators, broadcasters, satellite distributors, and other multichannel video programming distributors to close caption their television programs.[18] The FCC created a transition schedule that required an increasing amount of programming to include closed captioning each year.[19]

The FCC also began creating requirements and timetables for video descriptions. These requirements stated that broadcasters affiliated with the ABC, CBS, Fox, and NBC would be required to provide video descriptions for a minimum of fifty hours per calendar quarter of prime-time or children's programming.[20] The requirements additionally applied to other television providers who had 50,000 or more subscribers.[21]

Forecasting the challenges it would face in court, the FCC itself was divided on whether it had the authority to make the changes to video description requirements. It was a close three-to-two vote by the FCC Commissioners in favor of creating and enforcing the regulations.[22] In his dissenting opinion, Commissioner Michael K. Powell said that the FCC lacked authority because, "Congress spoke to video description in section 713(f), and purposely limited the Commission to studying the issue and reporting to Congress . . . ."[23] Commissioner Powell specifically looked at

---

15. *Id.*

16. *See id.* at 1956.

17. The FCC voted three to two to adopt rules requiring certain video programmers to supplement certain programming with video descriptions. Motion Picture Ass'n of Am., Inc. v. FCC, 309 F.3d 796, 800 (D.C. Cir. 2002). The FCC concluded that it possessed the statutory authority to adopt these rules. Implementation of Video Description of Video Programming, *Report and Order*, 15 F.C.C.R. 15230, paras. 57–61 (2007) [hereinafter *Report and Order*].

18. *FCC Consumer Facts*, *supra* note 7.

19. *Id.*

20. *See Report and Order*, *supra* note 17, at para. 6.

21. *See id.*

22. *Motion Picture Ass'n*, 309 F.3d at 800.

23. Press Statement, Comm'r Michael K. Powell, Comm'r of the FCC, Dissenting in Part, Implementation of Video Description of Video Programming 1 (Jul. 21, 2000),

the actions of the conference committee in striking the provisions regarding the FCC's authority to pass video description regulations as making it "abundantly clear that Congress specifically considered granting discretionary authority to the FCC to promulgate video description rules and elected not to do so."[24]

The majority of the Commissioners did not find the fact that Congress took out the clause as dispositive of its intent to prevent the FCC from making rules regarding video descriptions. The majority stated:

> While this history indicates that section 713 [of the Telecommunications Act of 1996] should not be construed to authorize a Commission rulemaking, the history does not indicate that section 713 should be construed to prohibit such a rulemaking, given our otherwise broad powers to make rules, as expressed in sections 4(i) and 303(r) of the Act. Had Congress intended to limit our general authority, it could have expressly done so, as it has elsewhere in the Act.[25]

The majority further relied on the Supreme Court's earlier categorization of the Telecommunications Act of 1996 as an amendment to the 1934 Act rather than freestanding legislation.[26] Thus, it argued, the FCC's authority in the original legislation was not supplanted and the FCC could still make regulations that may be necessary in the public interest.[27]

The arguments espoused by both the majority and minority FCC Commissioners were reargued when the matter was litigated in front of the court in *Motion Picture Association of America*. The arguments of the dissenting Commissioners helped shape the main points of the Motion Picture Association of America and heavily influenced the outcome of the case.

B. Initial Reception to the Video Description Regulations—the Battle Begins

When the FCC opened up the proposed video description regulations for comment, the new provisions received a mixed reception. The American Council of the Blind applauded the FCC for these efforts and also offered its expertise.[28] The Council believed that the regulations were necessary and could also be accomplished with minimal financial burden

---

available at http://www.fcc.gov/Speeches/Powell/Statements/2000/stmkp015.html.

24. *Id.* at 2.

25. *Report and Order*, *supra* note 17, at para. 58. For the Supreme Court's categorization, see *AT&T Corp. v. Iowa Utilities Board*, 525 U.S. 366, 377–78 (1999).

26. *Report and Order*, *supra* note 17, at para. 59.

27. *Id.* at para. 60.

28. Letter of American Council of the Blind, Implementation of Video Description of Video Programming, FCC MM Docket No. 99-339 (rel. Feb. 1, 2000).

on the television industry.[29] Television providers, such as DIRECTV, felt that the FCC lacked statutory authority and were worried that new regulations would impair their ability to stay competitive with cable providers.[30] Among DIRECTV's main concerns was the financial burden that would be placed on it.[31] The new laws required the use of secondary audio channels that only approximately one third of DIRECTV's channels supported.[32] It was not long before the voices of dissent turned into legal challenges against the FCC's ability to mandate video descriptions.

The Motion Picture Association of America (MPAA) was among a handful of organizations that challenged the FCC's authority to regulate video descriptions. The MPAA argued that the FCC did not have the power under the Telecommunications Act of 1996 to regulate video descriptions, and no other existing provisions granted it such power.[33] At the core of the MPAA's argument was the belief that the FCC did not have unlimited authority to act as it saw fit with respect to all aspects of television transmissions.[34]

The court in *Motion Picture Association of America* considered the two main arguments the FCC had relied on its *Report and Order*. The FCC's first argument was that its authority to regulate video description came from the same set of provisions in the Telecommunications Act of 1996 that gave it the power to regulate closed captioning.[35] Its second argument was that its power to regulate came from a combination of section 1, section 2(a), and section 4(i) of the Communications Act of 1934. Taken together they argued that the FCC possessed the ability to regulate video descriptions inherently.[36]

After comparing the closed captioning and video description provisions of the Telecommunications Act of 1996, the court found the FCC's first argument unpersuasive.[37] Instead, the court found it persuasive that Congress decided not to include language about the power to regulate video description despite choosing to do so for closed captioning. The court stated:

> The difference in the language employed in [the sections relating to closed captioning] makes it clear that subsection (f) is not intended to

---

29. *Id.* at 7.

30. Comments of DIRECTV, Inc. at 2, Implementation of Video Description of Video Programming, FCC MM Docket No. 99-339 (rel. Feb. 25, 2000).

31. *Id.* at 5–8.

32. *Id.* at 2–3.

33. Motion Picture Ass'n of Am., Inc. v. FCC, 309 F.3d 796, 798 (D.C. Cir. 2002).

34. *Id.* at 798.

35. *Id.* at 802–03.

36. *Id.*

37. *Id.* at 802.

> provide a mandate for video description requirements. Subsection (f) neither parallels the closed captioning mandate contained in subsection (b) nor suggests that Congress provided the FCC with discretionary authority to adopt video description rules.[38]

Section 713(b) of the 1996 Act says that the FCC shall create the necessary regulations, and those regulations shall ensure that "video programming first published or exhibited after the effective date of such regulations is fully accessible through the provision of closed captions . . . ."[39] In contrast, the language of section 713(f) is nowhere near as empowering. It allows the FCC to "commence an inquiry to examine the use of video descriptions on video programming . . . ."[40] The section mentions the creation of a report and the conducting of an inquiry, while never specifically mentioning any other action.[41]

The *Motion Picture Association of America* court subsequently rejected the second argument made by the FCC in its *Report and Order*[42], where the FCC relied on the enabling provisions of the 1934 Communications Act: "The Commission may perform any and all acts, make such rules and regulations, and issue such orders, not inconsistent with this chapter, as may be necessary in the execution of its functions."[43] The FCC believed that this statutory authority was enough to give it the discretion to regulate video descriptions. The court discarded the FCC's argument that nothing in the Amendments to the Telecommunications Act prohibited it from making regulations regarding video descriptions—the Act simply did not mention a positive right to create them.[44] It was the FCC's position that because of these provisions—and because closed captioning and video descriptions were so similar—its power to regulate one indicates the power to regulate the other.[45]

The court found that allowing the FCC to mandate video descriptions should not be allowed because unlike closed captioning, "[v]ideo description is not a regulation of television transmission that only incidentally and minimally affects program content; it is a direct and significant regulation of program content. The rules require programmers to create a second script."[46] The court believed that closed captioning

---

38. *Id.*

39. Telecommunications Act of 1996, § 713(b), Pub. L. No. 104-104, 110 Stat. 126 (codified at 47 U.S.C. § 613(b)).

40. Telecommunications Act of 1996, § 713(f) (codified at 47 U.S.C. § 613(f)).

42. *See id.*

43. *Report and Order*, *supra* note 17, at para. 54.

43. Communications Act of 1934, § 4(i), ch. 652, 48 Stat. 1064 (codified at 47 U.S.C. § 154(i)) (2006).

44. Motion Picture Ass'n of Am., Inc. v. FCC, 309 F.3d 796, 801–02 (D.C. Cir. 2002).

45. *See id.* at 803.

46. *Id.*

requirements were simplistic because all that was necessary was the creation of a transcript of what the actors were saying on screen.[47] The statutory provisions would be easy for a studio to implement because they require only that a studio recreate the script containing all of the words that were spoken on screen.[48]

In contrast, the court found the process needed to create video description technology easily distinguishable from closed captioning because video description would require the creation of a new script, hiring of additional actors, and review by a producer to make sure that the content fit with the feel of the show.[49] The court felt that all of these additional actions added up to a change in program content and imposing an additional financial burden on television studios.[50] Since video description regulation would impact program content, the court held that it fell outside the purview of the FCC,[51] which was created to "regulat[e] interstate and foreign commerce in communication by wire and radio so as to make available, so far as possible, to all the people of the United States, without discrimination on the basis of race, color, religion, national origin, or sex . . . ."[52] The court interpreted the phrase "all the people of the United States" to refer only to geographic location and not those with disabilities.[53]

The court rejected the FCC's 47 U.S.C. § 154(i) argument by analogizing it to the "necessary and proper" clause in the Constitution.[54] The court decided that it was not a standalone clause and must be read in conjunction with all other parts of the code;[55] the FCC cannot promulgate regulations without express authority from another source.[56]

The decision did leave open the possibility that with congressional approval, the FCC would be able to pass regulations mandating video descriptions regardless of the effect they would have on content.[57] From the holding, one could infer that it would take an express act of Congress to

---

47. *Id.*

48. *Id.*

49. *Id.*

50. *See id.* (explaining that video descriptions require a producer to evaluate the program, a new script, and new actors as opposed to closed captioning which is simply a straight translation of the dialog into text which already exists in the form of the script).

51. *See id.* at 804.

52. Communications Act of 1934, § 1, ch. 652, 48 Stat. 1064 (codified as amended at 47 U.S.C. § 151) (2006) (establishing the FCC).

53. *Motion Picture Ass'n*, 309 F.3d at 804.

54. *Id.* at 806.

55. *Id.*

56. *Id.*

57. Sarah M. Preis, *To Regulate or Not to Regulate: The FCC's Authority to Regulate Online Copyright Infringement Under the Communications Act*, 2008 U. CHI. LEGAL F. 535, 546–47 (2008).

overcome the current interpretation of the 1996 Telecommunications Act. The court found that "[a]fter originally entertaining the possibility of providing the FCC with authority to adopt video description rules, Congress declined to do so. This silence surely cannot be read as ambiguity resulting in delegated authority to the FCC to promulgate the disputed regulations."[58] Congress would need to reverse its position on the importance of video description and pass new legislation giving the FCC discretion similar to what it has for closed captioning.

III. CONFLICT AND CHANGES

Even before the decision came down in *Motion Picture Association of America*, video descriptions were a contentious topic. Different factions within the visually impaired community could not agree on what regulations needed to be created and how extensive they should be. To understand the need for action to be taken to remedy the current status of video description technology and the ability of the FCC to regulate it, it is important to understand the barrier that the lack of video descriptions poses to the safety and socialization of the visually impaired community. Changes in television technology after the court's decision and the transition to digital television could serve as a catalyst for change in the legal landscape. Digital television might be able to assuage many of the problems that conflicting parties had over the idea of video description regulations.

A. Why Is This a Problem?

It is December in Michigan and you are home for the night. You are sitting on the couch with your feet bundled up in cozy slippers, a mug of hot chocolate in your hands. As you begin to watch your favorite program, you hear the annoying "beep, beep, beep" and look down to read a winter storm warning scrolling across the bottom of the screen. As annoyed as you are about the obnoxious beeping sound that interrupted your sitcom, you are grateful to know that maybe tomorrow would not be the best day to plan on driving and that you need to make back-up plans for the kids in case school is canceled. However, if you are blind, you have no idea of what the warning accompanying the beeping says.

One of the reasons the initial regulations lacked overwhelming support from the seeing-impaired community was that it did not solve one of its major concerns. Some considered the more pressing issue to be access to safety information, which was scrolled across the screen in times of emergency. In its comment to the FCC about video description

58. *Motion Picture Ass'n*, 309 F.3d at 806.

regulations, the National Federation of the Blind condemned:

> the lack of access to emergency weather and news information scrolled across the bottom of the screen; the lack of access to the identities of talking heads in national and local news broadcasts; the lack of access to sports scores for [their] local team; or the lack of access to printed information during commercials some of which are health-related and display vital phone numbers.[59]

The Federation expressed its concern that equal access to this information would not be provided unless mandated by the FCC.[60] The Federation was also concerned that the initial attempt at regulation was focused solely on what the blind community would enjoy, instead of on what it needed.[61]

The comments of the National Federation of the Blind differed from those of the American Federation for the Blind. The American Federation for the Blind pointed out that equal access to all television was important for the seeing impaired of all ages for myriad reasons:

> Whether the viewing experience is educational or entertaining, people who are blind or visually impaired are usually denied access to the full message, unless, of course, video programming is described. For children, such disenfranchisement may mean immediate exclusion from social interaction with their sighted peers. Without video description, blind children and adults alike are denied the opportunity to learn things such as the nuances of body language, the significance of costume or dress, and much more—important concepts which a sighted child or adult learns easily through visual observation.[62]

The foundations of these arguments are easy to comprehend. Everyday people talk about what they watched on television the night before. Bonding over favorite television programs or touching news stories is a regular occurrence for people of all ages across the social spectrum. Without being able to see action on the screen, the visually impaired lose out on the chance to form bonds with those around them.

The positions of the National Federation of the Blind and the American Federation for the Blind both help to illustrate why there is a need for video description services to be regulated by the FCC. Video description services are needed to ensure universal access to important information that is presented nonaudibly during broadcasts. This information is needed for both health and safety reasons, but because of the cost of the technology, it is unlikely that it would be implemented unless it is mandated. Ensuring the safety of others during disasters and inclement

---

59. Comments of the National Federation of the Blind at 1, Implementation of Video Description of Video Programming, FCC MM Docket No. 99-339 (rel. Feb. 23, 2000).

60. *Id.* at 2.

61. *Id.*

62. Comments of Alan Dinsmore on Behalf of American Foundation for the Blind at 2, Implementation of Video Description of Video Programming, FCC MM Docket No. 99-339 (rel. Feb. 24, 2000) [hereinafter American Foundation for the Blind Feb. 2000 Comments].

weather is morally important and video descriptions provide the government with an opportunity to do that.

People can also use broadcast television for a number of other things—it provides an opportunity to fit in socially, and to take part in normal human activities. Although those who are seeing impaired can still participate in "water cooler" conversation, they cannot fully participate because they cannot fully experience television. The descriptions can also help children socialize normally by picking up visual cues transcribed in video descriptions. These socialization cues are facets of human interaction that children would otherwise have missed. Action must be taken through FCC regulations or other remedies to fix these problems.

B. The Effect of Video Descriptions on the Television Industry

When the proposed regulations relating to video descriptions opened for comment, many advocacy groups and businesses with a stake in the television industry commented on the positive and negative effects the regulations could have. The comments made by these groups illustrated the effects that video description regulations would have on the entertainment industry as a whole, and specifically on the television industry. Citizens with other disabilities, as well as networks and television studios, were all concerned with the overall impact of the regulations.

The group TDI (formally Telecommunications for the Deaf and Hard of Hearing, Inc.)[63] supported the proposed actions of the FCC, but was concerned that they did not go far enough.[64] TDI felt that previous reports and studies by the FCC had not yielded significant progress in television access to the blind in the preceding five years.[65] It hoped that the FCC would increase the scope of its proposals and decide that all television had to have video descriptions.[66] TDI believed that the regulations were an adequate first step for the FCC to be taking, but hoped that video description regulations would extend further in the future.[67]

The National Cable Television Association (NCTA), in its comment, joined with others in criticizing the FCC, arguing that it was overstepping

---

63. TELECOMM. FOR DEAF & HARD OF HEARING, INC., http://www.tdi-online.org/ (last visited Feb. 22, 2011). "TDI is a national consumer organization that seeks to represent the interest of the twenty nine million Americans who are deaf, hard of hearing, late deafened and deaf-blind." Reply Comments of Telecommunications for the Deaf, Inc. at 2, Implementation of Video Description of Video Programming, FCC MM Docket No. 99-339 (rel. Feb. 25, 2000) [hereinafter TDI Reply Comments].

64. TDI Reply Comments, *supra* note 63, at 2–3.

65. *Id.* at 3.

66. *Id.*

67. *Id.* at 2–3.

its mandate.[68] The NCTA additionally decried the expansive cost that cable providers would face.[69] It did not believe that the necessary infrastructure had been developed to provide any type of meaningful access to video descriptions.[70] In addition to the lack of infrastructure and cost, the NCTA was concerned about the time pressure that it would place its members under.[71]

Television networks like A&E Television[72] were similarly concerned with the effects of the regulations. They cited their concern over the FCC's lack of statutory authority to create the regulations and also the increased cost to noncable networks like themselves.[73] A&E stated:

> Video description is a developing service that faces many obstacles before it can become successful, and the industry has had only limited experience with the service. Moreover, the proposed rules would impose a disproportionate burden on cable networks, the economics of which are vastly different from the large broadcast networks.[74]

A&E viewed the efforts as morally praiseworthy but not something that was worthy of a mandate.[75]

Another comment came from the Narrative Television Network (NTN).[76] It reiterated the importance of implementing the regulations and stated its belief that the timetables proposed by the FCC would be adequate.[77] NTN said that "[v]isually impaired people, including those who own and operate NTN, have been waiting for many years to be able to enjoy the many benefits of accessible television and movie programming."[78]

These comments illustrate the wide total effect that video descriptions

---

68. Reply Comments of National Cable Television Association at 2–6, Implementation of Video Description of Video Programming, FCC MM Docket No. 99-339 (rel. Mar. 29, 2000) [hereinafter National Cable Television Mar. 2000 Reply Comments].

69. *Id.* at 7–9.

70. *Id.* at 6–7.

71. *Id.* at 7–9.

72. "A&E Television Networks ("AETN") . . . [is] an independent cable programmer offering the A&E Network, The History Channel, The BIOGRAPHY® Channel and History Channel International." Comments of A&E Television Network at 5, 16, Implementation of Video Description of Video Programming, FCC MM Docket No. 99-339 (rel. Feb. 25, 2000) [hereinafter A&E Comments].

73. *Id.* at 5–14, 16–19.

74. *Id.* at iii.

75. *See id.* at 2–3.

76. The Narrative Television Network (NTN) was "founded in 1988 by [the] blind and visually impaired" and has been a leader in making television programming and movies accessible to the visually impaired. Comment of Narrative Television Network at 2, Implementation of Video Description of Video Programming, FCC MM Docket No. 99-339 (rel. Feb. 22, 2000).

77. *Id.* at 4–5.

78. *Id.*

would have on the many branches of the television industry. Smaller networks would be forced to come up with a way to fund the video description process. However, even A&E recognized that this was a praiseworthy goal.[79] If a method of funding could be found and Congress gave the FCC the statutory authority to make video description regulations, the main arguments of the opposition would be alleviated.

C. Showdown: Video Description Versus Closed Captioning

The technology used to create closed captioning for television is vastly different from the technology required for video descriptions. These differences helped to frame the battle that took place in 2002 when the FCC lost the ability it believed it possessed to regulate video description implementation. Not only is the technology different but also video descriptions require additional costs that closed captioning does not. However, with technology changes over the past eight years, technology might not have been a factor if the same battle took place today. Many of the comments to the initial legislation included concerns over the cost of video description technology—but with that concern assuaged, one obstacle in the path of new legislation may have been removed.

Closed captioning allows viewers to read dialog that actors and commentators are saying on the screen. The "closed" in closed captioning means that the captions are not visible to everyone, and can be turned on or off.[80] Captioning has been used since 1948 when the film *America the Beautiful* was captioned.[81] Captioning for television was first publicly previewed in 1971, and the FCC set aside channels for it in 1976.[82] The process requires an operator to translate what is being said into text. Closed captioning is usually done before a show airs, but technology now allows a translator to work live, and type the transcription as it happens. Before the transition to digital television, closed captioning was accomplished using EIA-608.[83] Technology originally allowed broadcasting of the closed

79. *See* A&E Comments, *supra* note 72, at 2–3.

80. *See FCC Consumer Facts*, *supra* note 7.

81. *Captioned Movie Access Advocacy–Timeline*, NAT'L ASS'N OF DEAF, http://www.nad.org/issues/technology/movie-captioning/timeline (last visited Feb. 22, 2011).

82. Mary Bellis, *Closed Captioning*, ABOUT.COM: INVENTORS, http://inventors.about.com/library/inventors/blclosedcaptioning.htm (last visited Feb. 22, 2011).

83. *See* Sarkis Abrahamian, *EIA-608 and EIA-708 Closed Captioning*, EVERTZ–RESOURCES & PRESENTATIONS, http://www.evertz.com/resources/eia_608_708_cc.pdf (last visited Feb. 22, 2011). EIA-608 is named after the Electronic Industries Alliance which is a professional organization that created the technology. The Alliance ceased operations on December 31, 2010. EIA, http://www.ecaus.org/eia/site/index.html (last visited Feb. 22, 2011).

captions on one designated caption channel and was usually devoted to English translations.[84] This has recently expanded to allow multiple captioning channels to be used, so that captions can be created in multiple languages.[85]

Closed captioning technology has been required on all televisions larger than thirteen inches since the passage of the Television Decoder Circuitry Act of 1990.[86] Because it has been so widely mandated, there has been a significant incentive for television broadcasters to find cost-effective ways of captioning. Through the Telecommunications Act of 1996, the FCC mandated an eight-year phase-in for captioning of programs airing for the first time.[87] "As of January 1, 2006, all 'new' English language programming . . . first published or exhibited on or after January 1, 1998, and digital programming first aired on or after July 1, 2002, must be captioned, with some exceptions."[88] The FCC also requires that old programs be captioned as well—those that were created and broadcast before the creation of the Act.[89]

Closed captions are sent over the normal broadcast signal. Before the digital transition, signals were sent at a slow rate, allowing only sixty symbols to be sent per second.[90] This low signal rate meant that captions could be transmitted in color, but would still appear in black and white on the bottom of the screen. The text would be able to appear in up to four rows. [91]

Closed captioning technology has advanced with the transition to digital television have allowing for many advances. The change in technology has allowed the captions to shift from only appearing in the top or bottom third of the screen to appearing anywhere on the screen, which allows viewers to be able to easily discern who is talking on screen.[92] The change also allows closed captioning to be displayed in a number of new languages because it

---

84. *See* Scott Allen, *A Brief History of Closed Captioning*, MENTAL_FLOSS, (Sept. 3, 2009, 10:51 AM), https://www.mentalfloss.com/blogs/archives/33518; *see also TechFacts: Information About Captioning for Video Professionals, Volume 3–Closed Captioning: The State of the Art*, MEDIA ACCESS GROUP WGBH, http://main.wgbh.org/wgbh/pages/mag/resources/archive/techfacts/cctechfacts3.html (last visited Feb. 22, 2011) [hereinafter *TechFacts*].

85. *Id.*

86. Abrahamian, *supra* note 83.

87. *See* Closed Captioning & Video Description of Video Programming, *Report and Order*, 13 F.C.C.R 3272, para. 12 (1997).

88. *FCC Consumer Facts*, *supra* note 7.

89. *Id.*

90. *TechFacts*, *supra* note 84.

91. *Id.*

92. *Id.*

allows for the use of new characters.[93] Television shows in Chinese, Thai, Japanese, Korean, and Arabic can all be captioned now.[94]

Video description technology has not existed for nearly as long as closed captioning. It was first invented and used in 1990 by WGBH, a public television station in Boston.[95] The recorded descriptions of key visual elements were broadcast over a third audio channel.[96] Although the technology is relatively new, the idea has been around for a long time.[97]

The process of making the script for video descriptions is much more involved than that of closed captioning. Instead of involving just one translator, video description is a team effort. First, a group of describers watch the program and write down the key visual elements, then they turn these elements into a script.[98] Next, they have to edit and time each of the elements in order to fit them into the natural pauses of a program.[99] Then, a post-production supervisor reviews the script and edits it for continuity, clarity, and style.[100] Finally, the script has to be recorded and matched with the video to complete the whole track.[101]

Prior to the digital transition, television providers conveyed video descriptions to viewers by using secondary audio programming (SAP).[102] SAP is also used for a number of things in addition to video descriptions, such as presenting the same program in a different language.[103] Like closed captioning, SAP works only when activated.[104] Most televisions manufactured after 1995 have SAP technology capabilities.[105] It is also possible to get a portable SAP receiver if your television is not equipped

---

93. *What Are "708" and "608"?*, CPC: HOME E-CAPTIONING, http://www.cpcweb.com/hdtv/708.htm (last visited Feb. 22, 2011).

94. *Id.*

95. B.J. Cronin & S.R. King, *The Development of the Descriptive Video Services*, NAT'L CENTER TO IMPROVE PRAC. SPECIAL EDUC. THROUGH TECH., MEDIA AND MATERIALS, http://www2.edc.org/NCIP/library/v&c/Cronin.htm (last visited Feb. 22, 2011).

96. *Id.*

97. *Id.*

98. "In the 1960s, some attempts were made to fill in the gaps for Star Trek programs through audio cassettes. In the 1970s, a former radio broadcaster began describing movies over a Philadelphia radio station. In 1981, Margaret Pfanstiehl began describing live theatrical performances in Washington, DC." *Id.*

99. *DVS®:FAQ: What Is the Process of Descriptive Video Service?*, MEDIA ACCESS GROUP WBGH, http://main.wgbh.org/wgbh/pages/mag/services/description/dvs-faq.html (last visited Feb. 22, 2011).

100. *Id.*

101. *Id.*

102. Closed Captioning and Video Description of Video Programming, *Report*, 11 F.C.C.R. 19214, para. 94 (1996).

103. *Information About Secondary Audio Programming*, ACCESS DOME, http://www.accessdome.com/com-sap/sap.general.asp (last visited Feb. 22, 2011).

104. *Id.*

105. *Id.*

with one.[106]

The cost of video description can vary depending on how complicated the project is and how much extra work must go into creating the video descriptions. For a television station broadcasting a two-hour feature film, the cost can range from $8,000 to $12,000.[107] For hour-long television programs, the cost is only around $3,400.[108] Most of these costs are incurred post production, long after production of the movie or television show has been completed.[109] When commenting on the proposed FCC regulations for video descriptions, before they were found to be outside of the FCC's purview, the American Foundation for the Blind suggested that cost could be reduced if video descriptions were rolled into the regular production budgets of television shows and movies.[110] Studios would not have to create an additional script, hire new writers, or hire new producers, because they would be able to use the same ones that were already working on the principle production.

Today only a handful of shows are broadcast with video descriptions available to viewers. Many of these programs are on PBS,[111] but there are also a tiny number on the major network stations. Four of CBS's top shows—*NCIS*, *NCIS: LA*, *Criminal Minds*, and *CSI: Crime Scene Investigation*—are broadcast with video descriptions,[112] and on Fox, the only show with video descriptions available is *The Simpsons*.[113] NBC and ABC do not offer any shows with video descriptions.[114]

D. The Transition to Digital Television's Effect on Video Descriptions

On June 12, 2009, the transition to digital television was completed and all television stations are now broadcasting in digital format.[115] This

---

106. *Id*.

107. American Foundation for the Blind Feb. 2000 Comments, *supra* note 62, at 4.

108. *Id*.

109. *Id.*

110. *Id.* at 4–5.

111. *PBS September/October/November/December 2010/January and February 2011*, MEDIA ACCESS GROUP WGBH, http://main.wgbh.org/wgbh/pages/mag/services/description/ontv/pbs-schedule.html (last visited Feb. 22, 2011).

112. *DVS® on CBS*, MEDIA ACCESS GROUP WGBH, http://main.wgbh.org/wgbh/pages/mag/services/description/ontv/cbs-schedule.html (last visited Feb. 22, 2011).

113. *Fox Schedule*, MEDIA ACCESS GROUP WGBH, http://main.wgbh.org/wgbh/pages/mag/services/description/ontv/fox-schedule.html (last visited Feb. 22, 2011).

114. Kim McAvoy, *Stations Must Bear Cost of Service for Blind*, TV NEWS CHECK (Sept. 1, 2010), http://www.tvnewscheck.com/article/2010/09/01/44899/stations-must-bear-cost-of-service-for-blind.

115. *FCC Consumer Advisory: Video Descriptions and the Digital Television Transition*,

transition has had a significant impact on the accessibility of current video description services and the implementation of future video description services.

The transition to digital television has increased the number of audio channels that can be used to broadcast video descriptions.[116] Where there used to be only one or two channels available to broadcast alternative information, there are now six. Before the digital transition, broadcast stations had to choose between including video descriptions and broadcasting in alternative languages; that problem no longer exists. The FCC explained the difference in encoding:

> Because digital television encodes audio in a different manner than the encoding used in analog television, digital television does not utilize a SAP channel to transmit video descriptions. The digital television standards provide for two types of main audio service and six types of associated services, including associated services for people with vision disabilities.[117]

The change is good for television stations because now they can broadcast in multiple languages and also reserve an alternative audio channel for video descriptions. The networks will not have to alienate any of their consumers by excluding the medium in which the consumers would want to enjoy a program.

Despite its benefits, the transition to digital television has caused some problems, especially for those who were already relying on video People with older televisions encountered a problem during the transition because digital televisions encode audio differently than analog televisions.[118] Without purchasing a converter box their televisions had no way to process the new digital audio signal. Not all converters on the market are able to make the conversion,[119] leaving some seeing impaired people with no way to use the video description services. The government created a coupon program to alleviate some of the costs faced by those unable to make the transition.[120] Similar problems are faced by those members of the hearing impaired community who are dependent upon closed captioning.[121] The problem, however, is greatly diminished for

---

FED. COMM. COMMISSION, http://www.fcc.gov/cgb/consumerfacts/dtvvideodescription.html (last visited Feb. 22, 2011).

116. *Id.*

117. *Id.*

118. *Id.*

119. *Id.*

120. *Id.*

121. *See FCC Consumer Advocacy: Closed Captioning and Digital-to-Analog Converter Boxes for Viewing Free Over-the-Air Programming on Analog Televisions*, FED. COMM. COMMISSION, http://www.fcc.gov/cgb/consumerfacts/CC_converters.html (last visited Feb. 22, 2011).

members of that community since all televisions since 1993 larger than thirteen inches can display closed captions.[122] Those who received closed captions through their televisions are still able to do so after the transition.[123] Only those with televisions smaller than thirteen inches or televisions made before 1993 have to put full faith in the converter boxes.

An additional problem caused by the conversion is the requirement that the visually impaired learn how to access the video description services in a new way. Customers will have to figure out how to access the additional audio streams through a button on the remote or through a menu on the television,[124] either of which poses obvious challenges for the seeing impaired. It might be a challenge for people with disabilities to figure out how to do this, but it would seemingly present no larger of a problem than figuring out how to access video descriptions to begin with. This is not a difficulty faced by members of the deaf community who have to figure out the new way to access closed captioning, since they can view the on-screen menus.

The digital transition carries with it a unique opportunity to stimulate the video description market or impose mandatory regulations. The transition has made access to additional audio channels easy. Broadcasters can broadcast video descriptions in addition to alternative languages. Digital technology is also in high demand, and the government can take this opportunity to impose requirements for that technology.

IV. SOLUTIONS TO THE CURRENT SITUATION

As a result of the digital video transition and other technological advances, it is an ideal time for the regulation of video description technology. There are two different paths that the government could take to ensure that television programs and emergency information will be accessible to the millions of blind or seeing impaired in the United States. The first option is to increase the financing of video description services. This financing would provide an incentive for major studios to implement the technology and the system would eventually become self-sustaining. The second approach is to pass federal regulations that would place video description technology on equal footing with closed captioning.

A. Stimulating the Video Description Market

In August 2009, FCC Commissioner Michael Copps held a town hall meeting discussing the digital transition and the FCC's efforts to increase

122. *Id.*

123. *Id.*

124. *FCC Consumer Advocacy: Video Descriptions and the Digital Television Transition*, *supra* note 115.

access to television for people with disabilities.[125] Although video description technology was not the focus of the meeting, the subject came up during a question about funding. Karen Peltz Strauss, the Deputy Chief of the Consumer and Governmental Affairs Bureau at the FCC, who oversees the FCC's disability and consumer access programs and policies,[126] said that one of the biggest remaining concerns with video description technology was the cost.[127] In order to successfully increase access to video description technology, efforts need to be made to lower costs for networks and studios.

According to WGBH, the pioneer of video description technology, no commercial television program has offered video descriptions without public funding until recently.[128] Both WGBH and NTN receive major funding from the Department of Education.[129] In 2005, the Department of Education provided a grant to NTN in the amount of $800,000.[130] The purpose of the funding was to help the network describe an additional 750 hours of educational television for children.[131] WGBH also received a grant for $800,000 in 2005 from the Department of Education.[132] Although these amounts seem substantial when compared to existing funding of video descriptions, the amount would have to dramatically increase to support all major networks.

By increasing the amount of funding granted to organizations like WGBH and NTN, the government could offset the start-up costs and learning curve that major networks would encounter trying to start their own video describing programs from scratch. Allowing networks to initially outsource the video description process to those with experience (such as WGBH and NTN who would be receiving government funding) would expand the number of shows with video descriptions, help the

---

125. Kevin Taglang, *FCC Townhall Addresses Broadband Opportunities for Individuals with Disabilities*, BENTON FOUND. (Aug. 20, 2009), http://benton.org/node/27266.

126. Press Release, FCC, FCC Chairman Genachowski Names Karen Peltz Strauss as Deputy Chief in Consumer Bureau (Mar. 12, 2010), http://hraunfoss.fcc.gov/edocs_public/attachmatch/DOC-296832A1.pdf.

127. *See* Taglang, *supra* note 125.

128. Comments of CPB WGBH National Center for Accessible Media at 33, Implementation of Video Description of Video Programming, FCC MM Docket No. 99-339 (rel. Dec. 1, 1999) [hereinafter CPB WGBH National Center for Accessible Media Dec. 1999 Comments].

129. Jaclyn Packer, *Video Description in North America*, *in* 237 COLLOQUE INSERM: NEW TECHNOLOGIES IN THE EDUCATION OF THE VISUALLY HANDICAPPED 103, 105–106 (Dominique Burger ed., 1996); U.S. DEP'T OF EDUC., 2005 ANN. REP. TO CONGRESS ON THE *INDIVIDUALS WITH DISABILITIES EDUCATION ACT*, PART D, 51 (2005), http://www2.ed.gov/about/reports/annual/osep/2005/part-d/idea-part-d-2005.pdf.

130. U.S. DEP'T OF EDUC., *supra* note 129, at 52.

131. *Id.*

132. *Id.* at 51.

service expand to additional markets, and increase the demand for the services. Over time, mainstream studios, such as CBS, ABC, and NBC, would be able to develop their own full-time video description services in-house. Studios would no longer be wary of developing video descriptions because the market for their consumption would have been established. They would also be able to cut costs by doing the descriptions themselves because they could integrate the descriptions with the production process.

In addition, the visually impaired would feel more socially connected to others, as this would expand their cultural knowledge base. They would be able to better take part in water cooler conversations the next day at school and work. Similar, children with visual impairments would not have to feel left out because they missed the big show that was on the night before. More people watching would translate into additional revenues for studios. Studios would be able to further tap into the 25.2 million Americans who report vision loss.[133] An increase in the number of viewers would increase the amount of money they could charge advertisers for ad space, and increase their profits.

The increase in the number of secondary audio channels available on digital televisions will serve to benefit television stations in a number of ways. First, stations no longer have to choose between providing video descriptions and broadcasting a program in different languages. Second, stations can now broadcast emergency information on one of the secondary audio channels, instead of requiring visually impaired viewers to search for the information from another source. Networks that provide such a service would in turn receive increased loyalty from members of the visually impaired community.

Although the cost of descriptive technology could be high, there is also a huge opportunity for profit. Since video description technology is not currently widely utilized by studios, it would be economically advantageous to compete in that market. The concerns that currently exist about entering the market would no longer be warranted because there would be a guarantee that the technology would be used. There is a potential gain of between five and twenty-one billion dollars in revenue for the cable industry.[134] Some of this gain would have to be used to offset the

---

133. *Facts and Figures on Adults with Vision Loss*, AM. FOUND. FOR BLIND, http://www.afb.org/Section.asp?SectionID=15&TopicID=413&DocumentID=4900 (last visited Feb. 22, 2011).

134. Reply Comments of Helen Harris and Descriptive Theatre Vision at 2, Implementation of Video Description of Video Programming, FCC MM Docket No. 99-339 (rel. Mar. 27, 2000) ("[J]ust take [for] example the revenue generated from cable. If a subscriber pays $60 a month for cable service, that equals over $700 a year. Our figures show that there are 30 million people who can avail themselves of description. If those 30 million would subscribe, that would be $21 billion dollars additional revenue. Cut that in

additional expenses incurred by show producers and broadcasters, but that expense would not be significant enough to prevent net gain by the industry.

Market stimulation can be seen in the deal that was formed between WGBH and Sony Pictures Home Entertainment.[135] According to the press release, "Descriptive Video Service provides carefully crafted narration of key visual elements inserted into natural pauses in dialogue. Key visual elements are those which viewers with vision loss would ordinarily miss and include actions, costumes, gestures, facial expressions, scene changes and onscreen text."[136] Recognizing that there are over twelve million movie fans with vision loss,[137] this partnership opens up the home movie market to a wider range of people. Mainstream movies such as *Up*, *Zombieland*, *Julie & Julia*, and *Couples Retreat* were released on DVD and included descriptive narration.[138] Through similar partnerships, television studios could expand their audiences. Knowing that television programs would include video descriptions as a secondary option within a broadcast would garner more consumers from the seeing impaired community.

Once television programs include video descriptions, further opportunities will exist for studios to profit. Just like other consumers, members of the blind and seeing impaired community would purchase their favorite transcribed television shows on DVD. Video description would also have the potential to increase profits once shows were sold into syndication. Television programs with video descriptions included would have a built-in following that networks could rely upon.

The federal government would not need to continue financing video description technology forever. The government would only need to provide enough capital to get video description technology off the ground and increase awareness of its availability.[139] This would stimulate the market and enable it to become self-sustaining; the initial capital would serve to "prime" the video description "pump."[140] In its comments, WGBH

---

half, and it's $10 billion dollars. Cut that in half again, and it's still $5 billion dollars additional revenue that the vision impaired could contribute to the income of someone participating in description.").

135. Press Release, Media Access Group at WGBH, Sony Pictures Home Entertainment Partners with WGBH Media Access Group to Deliver Descriptive Video Service® on Home Video Titles (Sept. 3, 2009), http://main.wgbh.org/wgbh/pages/mag/about/news/sony.html.

136. *Id.*

137. *Id.*

138. *Accessible DVDs*, MEDIA ACCESS GROUP WGBH, http://main.wgbh.org/wgbh/pages/mag/resources/accessible-dvds.html (last visited Feb. 22, 2011).

139. *See* CPB WGBH National Center for Accessible Media Dec. 1999 Comments, *supra* note 128, at 30–31 (describing how such an approach has worked in the past).

140. This has been evidenced by the success of several public broadcast stations that

cited the various public broadcast producers that have utilized video description technology for several years with funding from federal grants and now no longer require such support.[141] In its original comment to the FCC proposed regulations, WGBH stated, "[w]hile major PBS stations in all of the top 20 markets carry DVS, so do many smaller member stations, some in the bottom 20 markets. Clearly if small and perennially hard-pressed public television stations can uncover the resources to add SAP-broadcast capability, so can most commercial stations."[142]

The transition to digital television has served to diminish the cost of implementing video description technology, one of the chief concerns of the parties that opposed the FCC's initial creation of the regulations.[143] Under the old analog system of video description, it was costly to mix the video descriptions with the regular audio.[144] The capabilities of new digital receivers reduce that cost. Under the old analog format, viewers had to pick between either the regular broadcast audio or the alternative audio. This all-or-nothing approach existed in part because the channels were typically used for broadcasting in a different language. Today, as a result of the digital transition, broadcasters can transmit multiple streams of video on a single channel at one time.[145] Where there was once only one option under the old format, broadcasters now have more audio channels to provide the service.[146]

Therefore, while finding enough initial funding poses a significant barrier to the implementation of video description technology, there are clear financial benefits in doing so. Stations that use video descriptions would realize an increase in revenue and could also realize an increase in viewership of their described shows, both of which would please commercial sponsors. The development and implementation of the technology would also increase the profits of the companies that create them.

---

similarly received federal grants, as described by WGBH in its comments. *Id.* at 30.

141. *Id.* at 25–26, 30–31.

142. *Id.* at 14 (internal citations omitted).

143. *See, e.g.*, National Cable Television Mar. 2000 Reply Comments, *supra* note 68, at 7–9.

144. *See* CPB WGBH National Center for Accessible Media Dec. 1999 Comments, *supra* note 128, at 34.

145. Peter H. Putnam, *The Basics of Digital Television*, AV SCI. F. (Mar. 24, 2004), http://www.avsforum.com/hdtvfaq/HDTV-FAQ.htm.

146. *FCC Consumer Advisory: Video Descriptions and the Digital Television Transition*, *supra* note 115.

B. Federal Regulation Mandating Implementation of Video Description Technology

Through the introduction of new legislation, the federal government could firmly establish that the FCC has the power to regulate video descriptions. This solution would address both aspects of the problem by mandating access to emergency information, as well as requiring closed captioning of television programs.

Now is the perfect time to reassess the FCC's authority to regulate video descriptions. Representative Edward Markey, a Democrat from Massachusetts, has introduced a bill in the U.S House of Representatives titled the Twenty-First Century Communications and Video Accessibility Act (Twenty-First Century Act).[147] The bill is cosponsored by fifty-three other representatives.[148] Representative Markey is the chairman of the House Subcommittee on Telecommunications and the Internet.[149] In promoting the bill, Representative Markey said, "'Now we're full-blown into this digital era, and we, in general, need to upgrade the laws that ensure that there is accessibility for all the people who use these new technologies.'"[150] The legislation illustrates that this is truly a bipartisan issue.[151] As of the writing of this Note, the Twenty-First Century Act had passed the House with a roll call vote resulting in 348 Ayes, 23 Nays, and 61 Present/Not Voting.[152] Despite passing in the House of Representatives, the Twenty-First Century Act still would have to go through several legislative steps to become law.

The Twenty-First Century Act is comprehensive and addresses many of the challenges faced by those with disabilities relating to new and changing technology. In addition to addressing these many issues, the Act firmly establishes the right of the FCC to regulate video descriptions.[153] By granting the FCC that power, the Act ensures that the needs of the blind and seeing impaired can be addressed as technology continues to advance.

Beyond giving the FCC the power to regulate broadcasters, the Twenty-First Century Act takes a number of other important steps to help the blind and seeing impaired community, including efforts to make

---

147. *H.R. 3101: Twenty-First Century Communications and Video Accessibility Act of 2010*, GOVTRACK.US, http://www.govtrack.us/congress/bill.xpd?bill=h111-3101#at (last visited Feb. 22, 2011).

148. *Id.*

149. Kim Hart, *Access Denied: The Blind or Deaf Can Feel Left Behind as the Tools of Technology Advance*, WASH. POST, June 19, 2008, at D01.

150. *Id.*

151. *See id.*

152. *H.R. 3101: Twenty-First Century Communications and Video Accessibility Act of 2010*, *supra* note 147.

153. H.R. 3101, 111th Cong. § 202(a) (2010).

television and other video technology easier to use. The Act authorizes the FCC to investigate ways to make onscreen television menus and other interfaces easier for those with disabilities to use.[154] Current regulations require that televisions with screens larger than thirteen inches must be able to broadcast closed captioning; this Act would further require those televisions to support video descriptions.[155]

For video descriptions, the Act basically turns back the clock to before the decision in *Motion Picture Association of America, Inc. v. Federal Communications Commission*. The Act "authorizes the FCC to promulgate additional rules to (1) ensure that video description services can be transmitted and provided over digital TV technologies, (2) require non-visual access to on-screen emergency warnings and similar televised information and (3) increase the amount of video description required."[156] Mandating that emergency information be broadcast aurally addresses one of the biggest concerns faced by the seeing impaired community—this ensures that members of this community will have increased access to safety information that will prove invaluable in times of emergency.

Passage of the Twenty-First Century Act would be taking a huge leap in solving all of the problems resulting from the lack of video description technology in television today. Although many specific details would still have to be addressed—such as the timetable for implementation—the Act would build upon the successful model of closed captioning to ensure success.

Even if the Twenty-First Century Act is not passed, it is still an ideal time to reconsider the results in *Motion Picture Association of America* and the repercussions it has had for the seeing impaired community. The transition to digital television presents the perfect opportunity to implement a change that would increase the safety and quality of life for the seeing impaired. Even without a congressional act, financing can be secured to stimulate a change in practices of major television studios.

V. CONCLUSION

The decision in *Motion Picture Association of America, Inc. v. Federal Communications Commission* had far-reaching consequences that have significantly impacted the lives of seeing impaired Americans. When the FCC lost the power to mandate implementation of video descriptions, members of the seeing impaired community lost the ability to enjoy things

154. H.R. 3101, 111th Cong. § 204(a) (2010).

155. H.R. 3101, 111th Cong. § 203(a) (2010).

156. *21st Century Communications and Video Accessibility Act*, NAT'L ASS'N OF DEAF, http://www.nad.org/issues/civil-rights/communications-act/21st-century-act (last visited Feb. 22, 2011).

most people take for granted. Because of the cost of creating video descriptions a majority of shows on television do not have them. Similarly, because networks are not required to have the technology in place, people with see impairments are not informed of vital emergency information that scrolls across television screen.

Some of the concerns expressed by the Supreme Court would no longer be a barrier to wide implementation of video description technology. Technological advances have made it easier and cheaper than ever for television studios to use video descriptions in their programs. The digital transition has transformed the broadcast television landscape opening up options to broadcasters that were not available even a few years ago.

As a result of the switch from analog to digital television, there are now two solutions to this issue. The first solution would be to financially stimulate the video description market—the government could help create video description services for television programs that would eventually become self-sustaining. Small public broadcasters having been describing video for years with help from federal grants. Over time they have increased the efficiency and lowering the cost of the process. Networks would be able to rely on their knowledge base on knowhow as they were launching their own video description services.

The second solution would be to create federal regulations mandating video descriptions. New regulations passed would not only serve to allow greater enjoyment of television programs, but would also allow for increased social integration, and access to vital emergency information. The *21st Century Communications and Video Accessibility Act* has been introduced and passed in the House of Representatives, although it has not yet become law. Either of these two courses of action has the potential to prevent the damage caused by the court's decision in *Motion Picture Association of America* from continuing to disadvantage the visually impaired.

FEDERAL COMMUNICATIONS LAW JOURNAL
ADVERTISING RATES

	Per Issue	Per Volume*
Full Page	$700	$2000

Specifics:

1) The FCLJ accepts ads from law-related business only;
2) Spaces are available on a first come, first served basis;
3) Ad size: 4 ¾" x 8";
4) Ads must be in black and white, camera-ready format. Portable Document Files (.pdf) are preferable, but other electronic formats are acceptable as long as they follow these guidelines
    a. Black and white line art/text must be at least 600 dpi
    b. Grayscale photos must be at least 266 dpi
    c. Both types need to be saved at the finished size with the required resolution
5) Hard copies are acceptable as long as they are produced at 1200 dpi on Hammermill 28/70 Photo White, or comparable paper (glossy or matte). Ads submitted in this format will be charged a camera fee of $20.

* The *Federal Communications Law Journal* is published three times a year with spine dates of December, March, and May. All per volume ads will include a link on the *FCLJ* website.

Volume 64 Deadlines:

	Space reservation	Artwork & Copy
Issue 1	Sept. 7, 2011	Oct. 5, 2011
Issue 2	Jan. 20, 2012	Feb. 15, 2012
Issue 3	Mar. 1, 2012	Mar. 25, 2012

Contact:
Executive Editor
Federal Communications Law Journal
Indiana University Maurer School of Law
211 South Indiana Avenue
Bloomington, IN 47405-1001
Phone (812) 855-5952
Fax (812) 855-0555
E-mail: fclj@indiana.edu