



The attached material is posted on regulation2point0.org with the permission of the Korea Information Society Development Institute (KISDI).

The Convergence of Broadcasting and Telephony: Legal and Regulatory Implications*

Christopher S. Yoo**
University of Pennsylvania

Abstract

The convergence of broadcasting and telephony induced by the digitization of communications technologies is forcing policymakers to rethink their basic approach to regulating these industries. Now that voice and video are becoming available through every transmission technology, policymakers can no longer define the scope of regulatory obligations in terms of the mode of transmission. In addition, jurisdictions that employ separate agencies to regulate broadcasting and telephony must reform their institutional structures to bring both within the ambit of a single regulatory agency. The emergence of intermodal competition will also place pressure on both telephone-style regulation, which protects against monopoly pricing and vertical exclusion, as well as broadcast-style regulation, which focuses on content and ownership structure. It will also force regulators to rethink social policies such as universal service and public broadcasting. At the same time, it is possible that convergence will be incomplete and that end users will maintain more than one network connection, which would reduce the danger of anticompetitive activity and allow policymakers to stop short of forcing every connection to be everything to everyone. Lastly, the increase in traffic volumes associated with the advent of Internet video may require the deployment of multicast protocols, content delivery networks, and more aggressive traffic management, all of which potentially implicate the debate over network neutrality currently taking place in the U.S.

Introduction

For decades, each type of electronic communications represented a technological and economic universe unto itself. Voice communications were available exclusively over wires provided by local telephone companies. Video programming was transmitted exclusively over the electromagnetic spectrum by local broadcasters. U.S. policymakers developed elaborate regulatory regimes based both on the technological and economic characteristics of the transmission medium on the one hand and the nature of the communications being transmitted on the other (Yoo, 2002, pp. 285-290).

For example, the high fixed costs associated with

the establishment of local telephone networks led Congress to regulate telephone companies as natural monopolies, enforced through statutory provisions initially enacted as Title II of the Communications Act of 1934. Telephone companies were thus subjected to rate regulation to protect against supracompetitive pricing as well as an array of access requirements and structural restrictions to guard against vertical exclusion. Because telephony was used exclusively to convey person-to-person communications, it had few implications for free speech and thus was largely free from content regulation.

Broadcast regulation, initially enacted in Title III of the Communications Act of 1934, was also influenced by the economic characteristics of the underlying

* While my analysis is framed in terms of U.S. law, the insights are fairly general and should be of interest to scholars from all jurisdictions.

** Professor of Law and Communication and Founding Director of the Center for Technology, Innovation, and Competition, University of Pennsylvania, USA. E-mail: csyoo@law.upenn.edu

transmission medium, in this case spectrum-based technologies' ability to create interference. Because broadcasting involved mass communications rather than just person-to-person communications, it was also subject to a wide range of content restrictions intended to promote certain forms of desirable content, such as political speech and children's educational programming, and for a time even prevented broadcast stations from abandoning important radio formats (*FCC v. WNCN Listeners Guild*, 1981). Other content regulation sought to curb certain forms of undesirable speech, such as indecency or advertisements about gambling. The broader social implications of broadcast speech led policymakers to subject broadcasters to a broad array of ownership restrictions that were more restrictive than could be justified by purely economic concerns (*FCC v. National Citizens Committee for Broadcasting*, 1978). At the same time, broadcasting's broader free speech implications led Congress to prohibit any regulations that would turn broadcasters into common carriers (47 U.S.C. § 153(10)).

Because each type of communications was available exclusively through a single mode of transmission, it was initially unproblematic that U.S. law made the transmission technology the key determinant of the regulatory regime to be imposed. Over time, different transmission technologies developed the capability to transmit different types of communications. For example, FM radio broadcasters began to use the subcarrier bands usually employed to provide stereophonic sound to provide services unrelated to broadcasting, including some forms of person-to-person communications. Television broadcasters similarly began employing the space between television frames (known as the vertical blanking interval) to transmit a wide variety of data, computer software, and paging services. Even though these services represented person-to-person communications, the fact that the regulatory regime turned largely by the means of transmission meant that these services nonetheless would have been subject to broadcast-style content restrictions had the FCC not specifically exempted these services from those requirements (Huber et al., 1999, § 10.8.2). The problems with this approach are illustrated by the service known as

teletext, which was employed a portion of the television broadcast spectrum to transmit an electronic newspaper. Even though U.S. law generally regards newspapers as constitutionally immune from any content requirements or political access rules (*Miami Herald Publishing Co. v. Tornillo*, 1974), because teletext uses the spectrum as its means of transmission, it would have been subject to the political broadcasting rules had the Federal Communications Commission (FCC) not specifically exempted it from those rules (*Telecommunications Research & Action Center v. FCC*, 1986).

Ultimately, a different technology would emerge as a more important source of spectrum-based person-to-person communications: wireless telephony. At the same time, wireless providers also were able to provide video and forms of mass communications. The result was a hybrid platform capable of providing both person-to-person and mass communications. The FCC responded by exempting these data services from most of the content requirements traditionally associated with broadcasting (Huber et al., 1999, § 10.3.4).

Conversely, cable television made it possible to deliver video programming via copper wires. As such, cable television presented something of a regulatory quandary. Economically, cable possessed the natural monopoly characteristics associated with telephony. At the same time, it transmitted mass communications and thus raised the broader social implications associated with broadcasting. The FCC struggled to determine the proper regulatory category for cable television. After initially concluding that cable fell outside of its regulatory jurisdiction, the FCC soon reversed course and ruled that it had ancillary jurisdiction over cable because of its tangential impact on broadcasting and began to impose broadcast-style regulation to cable. Because cable uses wires to convey communications in much the same manner as telephony, regulatory authorities also subjected cable to telephone-style economic regulation, such as rate regulation and access requirements traditionally associated with natural monopoly. The U.S. Supreme Court upheld the FCC's assertion of jurisdiction over cable as ancillary to its authority to regulate broadcasting (*United States v. Southwestern Cable Co.*, 1968), but struck down certain economic regulations as

inconsistent with free speech principles and the policies reflected in the Communications Act (*FCC v. Midwest Video Corp.*, 1979). In the end, the FCC's jurisdiction over cable was not resolved until Congress enacted major legislation in 1984 that added a new title to the Communications Act creating a regulatory category known as "cable services" and prescribing the regulatory regime that would be applied to that category (Yoo, 2002, pp. 286-288).

Policymakers faced a similar problem when confronting the broadband Internet. For example, early courts struggled over the proper regulatory classification of cable modem service, with some concluding that it was a cable service (*MediaOne Group, Inc. v. County of Henrico*, 2000), and with others concluding it was a telecommunications service (*AT&T v. City of Portland*, 2000). Interestingly, either conclusion would have resulted in subjecting broadband to a regulatory regime created for a completely different technology and economic environment.

At this point, the problems that convergence poses for regimes that make the mode of transmission the primary determinant of the scope of regulatory obligations should be apparent. Moreover, the fact that each type of communications was available over multiple media meant that regulators could no longer treat each medium as posing separate and independent regulatory issues. Instead, policymakers had to take into account potential interactions among different technologies. In addition, this nascent form of convergence tended to break down natural monopolies by enabling previously distinct systems of production to engage in a form of intermodal competition, such as the emerging rivalry between wireline and wireless telephony.

The digitization of all media will eventually lead to complete convergence, in which video and voice are both simply different applications available through the same broadband pipe. As this occurs, all regulatory distinctions based on the medium used to transmit particular forms of communications will completely break down. In addition, the FCC has placed all data services in a new regulatory category largely free from statutory restrictions called "information services" regardless of whether they are provided via cable

television, telephone, wireless, or power line networks (Spulber & Yoo, 2009, ch. 10). This conclusion leaves the FCC free to tailor the regulatory regime to the Internet's key economic and technical characteristics. Although the final principles that will govern information services are still a work in progress, it is clear that the impending convergence between voice and video will require a number of adjustments to both the institutions as well as the substantive principles of communications policy.

Implications of convergence for regulatory structure

One of the most striking implications of the convergence of voice and video is the need to vest regulatory authority over both types of communications in a single agency. This does not represent a significant obstacle in jurisdictions where a single agency governs both broadcasting and telephony. It poses a bigger challenge in countries where broadcasting and telephony have traditionally been overseen by separate agencies.

Because convergence opens technologies that were once regarded as natural monopolies to intermodal competition, many commentators have suggested that it also implies an increasingly narrow role for sector-specific regulation and a greater role for conventional antitrust (Geradin & Kerf, 2003). Weighing against that is the U.S. Supreme Court's recognition of antitrust courts' limited ability to engage in longstanding and ongoing oversight of the access mandates and other traditional attempts to curb anticompetitive practices through behavioral rather than structural relief (*Verizon Communications, Inc. v. Law Offices of Curtis V. Trinko, LLP*, 2004; see generally Spulber & Yoo, 2009, ch. 9, reviewing the application of antitrust law to the telecommunications sector).

Implications of convergence for telephone-style economic regulation

Convergence also implies that economic regulation

based on concerns about natural monopoly should begin to disappear. The fact that transmission technologies that once were separate now compete in the same product markets creates intermodal competition. Although such competition is likely to remain oligopolistic, it may nonetheless be sufficiently vigorous to protect consumers better than the imperfect process of agency regulation.

Few commentators and industry observers will rue the passing of these regulatory regimes. The traditional mechanism for ensuring natural monopolists do not charge excessive prices is known as cost-of-service ratemaking, which sets prices based on the provider needs to cover its operating costs plus its capital investments. Because capital investments require providers to make an up-front cash investment that will not be recouped until later years, providers are allowed to earn a rate of return on their capital investment (as measured by the “rate base”), but not operating costs.

Cost-of-service ratemaking is fraught with a wide range of well recognized problems (Spulber & Yoo, 2009, pp. 135-151). As an initial matter, determining the appropriate rate of return has proven extraordinarily difficult, because such a determination depends on identifying other ventures bearing similar risk. Unfortunately, the only enterprises with similar risks are themselves regulated enterprises and thus cannot serve as truly independent benchmarks. This determination is complicated by the fact that small differences in rates of return can have dramatic effects on the total revenue that the carrier is allowed to generate.

Determining the proper rate base has also proven to be a significant regulatory challenge. Ratemaking authorities initially calculated the rate base in terms of the replacement cost of the network, which allowed the rate base to reflect subsequent increases and decreases in value (*Smyth v. Ames*, 1898). The problem was that replacement cost was difficult to administer. Determining how much particular configurations of network elements would cost on the current market raised difficult problems of proof and typically devolved into a battle between expert witnesses. In addition, changes in demand and technology would often render particular facilities obsolete or would imply a very difficult network

configuration. The fair value approach left unclear whether the rate base would be based on the reproduction cost of the network as actually configured or of a hypothetical network configured in the most efficient manner.

The speculative nature of determining replacement cost led Justice Brandeis to propose basing the rate base on historical cost. While recognizing that historical cost was less economically suited to reflecting changes in value, it had the virtue of being easier to administer (*Missouri ex rel. Southwestern Bell Telephone Co. v. Public Service Commission*, 1923 (Brandeis, J., concurring in the judgment)). In later cases, the Supreme Court declined to mandate any particular approach to calculating the rate base, instead opting to defer to the methodology chosen by the agency so long as it is reasonable (*Federal Power Commission v. Hope Natural Gas*, 1944). Since that time, courts have sustained a wide array of different methodologies for calculating the rate base (Spulber & Yoo, 2009, p. 128).

Cost-of-service ratemaking also induces a number of systematic inefficiencies. As an initial matter, cost-plus pricing regimes give firms little incentive to economize. In response, agencies reviewing rates excluded any capital investments that were not “prudent” when made (Spulber & Yoo, 2009, p. 129). Using an *ex post* decisionmaking process to determine whether a particular action was reasonable *ex ante* is susceptible to hindsight bias in ways that can penalize firms for investments that failed for reasons that were impossible to foresee at the time.

Firms, moreover, are allowed to earn a rate of return on capital expenses, but not on operating expenses. Most firms usually have the choice to use production processes that are more or less capital intensive. The ratemaking methodology discussed above thus introduces a bias in favor of capital-intensive solutions even when other solutions would be more efficient (Averch & Johnson, 1962). Some regulators have attempted to eliminate these distortions by introducing price cap regulation, in which the prices firms are allowed to charge do not depend on actual costs, but instead are based on some base year and then adjusted upwards in subsequent years for inflation and downwards for

changes in productivity. Determining initial price levels and price adjustments in subsequent years has proven to be extremely difficult. The empirical literature is divided on whether price caps lead to lower or higher rates (Spulber & Yoo, 2009, p. 129).

The process of establishing rates can also dull competitive forces in other ways. Rate filings require companies to give their competitors advance notice of any changes in prices and strategy. In addition, collusion is easier to maintain when products are undifferentiated and when the prices charged are visible. The ratemaking process serves both of these functions and even places the regulatory agency in a position to punish any deviations from the cartel price. Even absent overt collusion, standardizing products and increasing price transparency facilitates non-cooperative oligopolistic behavior as well (Spulber & Yoo, 2009, pp. 255-256).

Ratemaking can also foreclose welfare-enhancing forms of price discrimination. When fixed cost is large relative to variable cost, as is traditionally the case in local telephone service, the average cost curve lies above the marginal cost curve over the entire industry output. Thus, any price that allows the network owner to cover its costs necessarily creates some degree of deadweight loss. Discriminatory pricing regimes, such as Ramsey pricing, can ameliorate this deadweight loss by allocating a larger proportion of the fixed costs to those customers whose demand is most inelastic and allocating a smaller proportion to those customers who are most price sensitive (Baumol & Bradford, 1970). In fact, price discrimination can theoretically lead to efficient outcomes if fixed costs are allocated in perfect inverse proportion to elasticity of demand.

Ratemaking also increases transaction costs, both directly, by requiring firms to participate in the ratemaking process, as well as indirectly, by providing incentives for competitors to challenge rates even when those challenges are unlikely to succeed on the merits. To the extent that retail access requires that all customers pay uniform rates for uniform services, it also limits network owners' ability to customize their offerings to the needs of particular customers.

In addition, concerns about natural monopoly also led policymakers to institute an elaborate set of access

and unbundling requirements to guard against vertical exclusion (Spulber & Yoo, 2009, pp. 143-151). Such access mandates have little chance of success when the product being regulated and the access interfaces are complex. As a result, access and unbundling mandates require regulators to oversee numerous dimensions of product quality as well as price. As the U.S. Supreme Court has noted, disputes over access to telecommunications networks "are highly technical" and "likely to be extremely numerous, given the incessant, complex, and constantly changing interaction of competitive and incumbent [local telephone companies] implementing the sharing and interconnection obligations." Moreover, when interfaces are complex, policing access "can be difficult because the means of illicit exclusion, like the means of legitimate competition, are myriad" (*Verizon Communications, Inc. v. Law Offices of Curtis V. Trinko, LLP*, 2004; *accord National Cable & Telecommunications Association v. Brand X Internet Services*, 2005).

Equally importantly, concerns about vertical exclusion are based on the presumption that the network owner has a monopoly. The emergence of intermodal competition is rendering this presumption increasingly untenable. Moreover, a growing group of empirical studies indicate that access requirements deter new investment in last-mile technologies. Mandating access may thus be inappropriate in cases where competition is feasible but has yet to emerge. Although some have suggested that unbundling requirements can be used to facilitate new entry by allowing entrants to avoid having to build out their entire network at once, empirical studies suggest that unbundling requirements have deterred investment in new telephone and broadband technologies and have failed to find any evidence indicating that unbundling has served as a stepping stone on the way to full facilities-based competition (see references cited in Yoo, 2007, p. 29).

It thus comes as no surprise that advanced data networks, such as fiber, digital subscriber lines, and cable modem systems, have never been subject to rate regulation. In addition, the FCC has largely exempted them from all unbundling and access requirements (Spulber & Yoo, 2009, p. 343).

Implications of convergence for broadcast-style content regulation

The convergence of broadcasting and telephony also has important implications for the content restrictions associated with broadcast-style regulation. These regulations are premised on the importance of broadcasting in general (and television in particular) as a source of news and information as well as the fact that the potential for interference placed a natural limit on the number of people who can speak at any one time. This relative scarcity of broadcast serves as the main justification for requiring broadcasters to devote these limited resources to conveying preferred speech (such as political broadcasting and children's television) and for restricting them from conveying dispreferred speech (such as indecency and advertisements about gambling) (Yoo, 2003, pp. 260-266).

The advent of the Internet as an important medium of communications has undermined television's claim as most individuals' primary source of news and information. At the same time, because data networks are not subject to interference, there is no natural limitation on the number of people who can speak that could justify compelling speakers to represent the views of others or from conveying speech that the government deems to be of low value (*Reno v. ACLU*, 1997). Although the U.S. Supreme Court has recognized that bottleneck control over a limited physical resource might serve as a basis for more limited regulatory intervention (*Turner Broadcasting System, Inc. v. FCC*, 1994), the rise of intermodal competition has made it unlikely that any court would find that exception to be applicable (Yoo, in press-b).

Implications of convergence for structural regulation

Convergence is also requiring the FCC to rethink its media ownership restrictions. Now that the same type of communications are available through multiple transmission technologies, the FCC can no longer calculate media concentration simply by focusing on

the number of available outlets for any particular communications technology. Instead, it must integrate all of the various technologies into a single metric that gives appropriate weight to each. Creating such a metric in a principled way has proven quite difficult, with the Court of Appeals striking down the FCC's most recent attempt to issue new media ownership rules primarily for the lack of consistency in its methodology for determining how much weight to give to different media (*Prometheus Radio Project v. FCC*, 2004). More importantly, the FCC has recognized that telephony has become sufficiently competitive to justify removing all of the important structural restrictions that had previously been imposed on the industry (Spulber & Yoo, 2009, pp. 246-248). The growing availability of the same types of mass media content via different transmission media has increased intermodal competition to the point where structural regulation of video and other related technologies has become increasingly difficult to justify as well (Yoo, 2002).

Implications of convergence for social policies

The fact that all forms of communications are available through the same data network also has implications for social policies. The emergence of competition erodes regulators' ability to use cross subsidies to promote social goals. Such cross subsidy systems presume the existence of a monopoly provider. To cite one example, U.S. telecommunications policy has historically allowed long distance rates to exceed cost and mandated that providers use the excess revenue to set prices for local service below cost. Business users were similarly overcharged, with the excess returns used to lower the cost of residential service. Such cross subsidies cannot survive the emergence competition, since the new entrants will naturally be attracted to the overpriced segments. The resulting competition will place downward pressure on prices for long distance and business services to the point where there are no longer any excess returns to redistribute to local and residential services.

The migration away from traditional telephony is also placing greater pressure on the funding mechanism for explicit universal service subsidies. For example, after initially exempting voice over Internet protocol (VoIP) from having to contribute to the universal service fund (FCC, 2004), the agency later reversed its position (*Vonage Holdings Corp. v. FCC*, 2007). It is far from clear that adding VoIP to the number of providers having to contribute to the universal service fund is sufficient to solve the problem. People have begun to replace telephone calls with a much broader array of communications services. Such diversification raises questions about which forms of communication should contribute to the universal service fund as well as which forms of communication should be the fund's beneficiaries.

At the same time, convergence represents something of a mixed blessing for public broadcasters, providing greater competition from alternative sources as well as greater potential opportunities for dissemination. Jurisdictions can respond to these developments in different ways. They can use the advent of competition as a justification for curtailing support for public broadcasting. They can attempt to quarantine convergence by preventing content initially transmitted via terrestrial broadcast from being available on the Internet, as happens in Japan. Or they can permit public broadcasters to expand into other types of content, as the BBC has done. That said, the BBC's aggressive strategy with respect to the Internet has prompted unsubsidized competitors to complain that the BBC is using its privileged position to compete on an unlevel playing field.

Implications of convergence for freedom of speech

Convergence should also have a significant impact on U.S. jurisprudence on free speech. Traditionally, the U.S. Supreme Court has employed a wide range of exceptions to subject different transmission technologies to varying levels of scrutiny under the First Amendment. The rationales underlying these exceptions have long

been the target of devastating criticism challenging their analytical coherence (Yoo, 2003). Moreover, the courts have essentially found these exceptions inapplicable to the Internet. In short, the Internet's ability to convey all major forms of communications effectively renders incoherent any regime that attempts to base the level of First Amendment scrutiny on the technological means of transmission. The extension of basic free speech principles to electronic media previously held to a lower level of scrutiny will place additional pressure on the regulatory regimes discussed above (Yoo, in press-b).

The myth of the one screen

In recognizing the growing importance of convergence, one need not assume that all modes of transmission will become completely fungible. For example, each of the various transmission media provides different amounts of bandwidth, which causes variations in their ability to support different applications. In addition, technologies such as cable modem service and wireless are subject to local congestion to a greater degree, which make the bandwidth that their subscribers receive more sensitive to the downloading activities of their immediate neighbors (Yoo, 2008). Perhaps most importantly, wireless technologies provide mobility, a functionality unavailable to other modes of transmission.

These technical differences among different transmission media suggest that some applications may not run equally well on all technological platforms. Indeed, we may well see network providers begin to differentiate the services they offer (Spulber & Yoo, 2009, pp. 387-388). Moreover, most individuals will likely continue to subscribe to separate wireline and wireless services, in part to ensure connectivity should one link fail and in part to avail themselves of the advantages of both mobility and greater bandwidth. The fact that most individuals will "multihome" (i.e., maintain more than one network connection) provides an important source of competitive pressure. The redundancy and robustness implicit in multihoming also makes it less important that every person be able to reach the entire universe of

network content through any particularly connection (Yoo, in press-a).

The potential impact of video-induced traffic growth

The first two applications that drove network growth were e-mail and the World Wide Web, both of which are technologies native to the Internet rather than convergent technologies. Many industry observers believe that the next killer application will be video. As of now, the video that has migrated to data networks is only beginning to match the quality of traditional television programming, and most people continue to receive a majority of their video programming via more traditional sources, such as broadcasting, cable, and now direct broadcast satellites (DBS). Some commentators predict that viewers will begin to abandon these traditional sources and that the migration of this demand for video to the Internet will create an “exaflood” that will cause a dramatic increase in network traffic (see Swanson & Gilder, 2007, for the initial statement).

Any acceleration in growth could have potentially dramatic effects. Since 2003, network traffic has been growing at a rate of roughly 50% to 60% each year. Commentators have observed that the Internet seems to obey a version of Moore’s Law, in that transmission capacity, processing power, and storage are improving at a sufficient rate to sustain roughly 50% annual growth in network traffic without requiring substantial increases in capital spending (Minnesota Internet Traffic Studies, 2007). Any video-induced increase in growth beyond current rates threatens to disrupt this delicate balance.

The engineering community has long explored alternative solutions. One is based around a technology called “multicasting.” Because the Internet evolved around person-to-person communications like e-mail and file transfers, it is built around unicast protocols, which establish a virtual connection between two distinct points along which senders and receivers exchange content and acknowledgements. This means

that a content provider wishing to use the current unicast architecture to transmit video programming to one million viewers would have to set up one million separate connections, retransmit the same set of packets one million times, and be prepared to receive acknowledgements for each packet from all one million viewers. Not only is the overhead of managing all of these connections potentially prohibitive. The volume associated with sending of duplicate packets and multiple acknowledgements can overwhelm the network.

To avoid these problems, the network engineering community has been experimenting with a large number of reliable multicast protocols to create a distribution hierarchy by designating particular routers located in the web as acknowledgement points. Instead of keeping track of how to reach each individual receiver, multicast protocols reduce overhead by allowing content providers to rely on the acknowledgement points to keep track of the recipients and to calculate the necessary routes. Multicast protocols also reduce total network traffic. Instead of sending a separate stream of packets to each recipient, the content provider simply initiates data streams to the smaller number of acknowledgement points. This centralization of distribution greatly reduces network traffic by allowing a single stream from the sender to serve multiple recipients. These acknowledgment points then bear the responsibility of forwarding the traffic on to the receivers, receive any acknowledgements from those receivers, and resend any lost data. Placing these functions closer to the end user can dramatically reduce the total burden on the network as well (Comer, 2006, ch. 16).

Another option is repositioning content in locations closer to the end user. Many network providers periodically retrieve popular webpages in a cache in their local facility. Thus instead of sending one hundred requests for the same webpage through the network, caching allows the network provider to send a single query, store the response in its local cache, and then serve all of the one hundred requests without posing any further burden on the network. Some companies called content delivery networks (CDNs) do this on a commercial, third-party basis. For example, the largest CDN (known

as Akamai) stores content in and distributes content from over fourteen thousand caches around the world. Prerecorded video content can be handled in this manner, as occurred when another leading CDN known as Limelight cached video content during the Beijing Olympics. The problem is that this solution only works for prerecorded video. It does not work well for live video or interactive video (such as video conferencing, virtual worlds, and graphics-intensive real-time games). Moreover, live video is particularly vulnerable to a phenomenon known as “flash crowds,” in which large numbers of end users simultaneously flock to one location.

Still another solution to increased traffic is simply to add more capacity. When links in a network become congested, the routers in a network place packets in a queue, which delays the data stream. Queue space is limited, moreover, and as the buffer becomes full, the router begins to drop packets, which then must be resent and becomes a further source of delay. These delays, which are usually for less than a second, were tolerable for traditional Internet applications, such as e-mail and web browsing, in which delays of a fraction of a second were barely noticeable. It is more of a problem for voice and video, in which a delay of as little as 300 milliseconds can render a service unusable.

Many commentators argue that adding capacity is always the preferred solution (Lessig, 2001, p. 47). The problem is that adding capacity is not always an option. Consider wireless networks: the capacity available to wireless networks is limited by the amount of spectrum allocated to them. Although they could expand capacity by increasing the number of microwave towers that they use and reducing the power at which they operate, there is a natural limit to how much additional capacity can be added in this manner.

Furthermore, capacity generally cannot be expanded instantaneously. The process of business planning and regulatory approval can take up to eighteen months. In the meantime, it is inevitable that network providers’ predictions of how much capacity will be needed in particular geographic areas will be imperfect from time to time, either because they did not anticipate demographic shifts in where people live or because they did

not anticipate the emergence of some new Internet application that alters the patterns of network usage. Given that no one can see the future perfectly, it is inevitable that bandwidth shortages will occur from time to time (Yoo, 2005, pp. 70-71). Although backbones and other providers operating in the core of the network are likely to be able to accommodate such growth, the greater cost and lead-times needed to expand last-mile networks pose greater challenges (Odlyzko, 2003).

Because of these limitations, the network engineering community has long been pursuing alternative solutions. Rather than simply increasing their capital spending, networks can address the problems of congestion simply by giving a higher priority to traffic that is more sensitive to delay and a lower priority to traffic that is less sensitive. Wireless broadband providers have been perhaps the most aggressive in this regard. The physics of wave propagation dictate that the amount of bandwidth available to a wireless subscriber will vary as that subscriber walks across a room. If the subscriber passes through a low bandwidth location, wireless networks typically give higher priority to voice traffic and holds the subscriber’s e-mail until he or she reaches a location with higher available bandwidth. Simply put, the particular quality of service needed from the network varies from application to application. Prioritizing time-sensitive traffic benefits end users by helping provide each application with the quality of service it requires.

More generally, the network engineering community has long been experimenting with mechanisms to provide applications with different levels of quality of service. The first initiative, known as Integrated Services (IntServ), allows end users to reserve bandwidth in advance. The downside was that implementing IntServ would require substantial changes to the router infrastructure. The second initiative, known as Differentiated Services (DiffServ), takes advantage of the fact that the existing architecture contains a mechanism for signaling the type of service needed by each packet to enable routers to give a higher priority to packets flagged with a higher priority level. Many companies have begun to use DiffServ in their internal networks to ensure that delay-sensitive traffic is delivered in a timely manner.

For example, Comcast is using DiffServ to prevent delays in its voice service, and AT&T is using DiffServ to ensure that there are no delays in its video service. Another solution, known as MultiProtocol Label Switching (MPLS), adds a label to the front of each packet and routes on the basis of the label (Comer, 2006, pp. 327-336, 510-515).

These techniques give network providers seeking to protect their customers from this potential upsurge in traffic another choice. Instead of accelerating the rate at which they are adding bandwidth, they can instead simply prioritize voice and video traffic over traffic associated with other applications. The potential benefits to consumers are eloquently demonstrated by a small UK service provider known as PlusNet, which segregates all traffic into seven different classes of priority. The resulting reductions in cost and improvements in performance have allowed PlusNet to win numerous awards for consumer value and customer satisfaction.

The migration of video into the data network thus has implications for the debate over network neutrality, which has been the leading issue in U.S. Internet policy over the past several years. Although the details of particular proposals vary widely, many network neutrality proponents oppose prioritizing traffic based on its source, destination, or the application with which it is associated. Others would object to any efforts to charge more for traffic based on the type of service being provided. The FCC recently initiated a regulatory proceeding to address whether such discrimination should be permissible (FCC, 2009).

In working out this problem, policymakers should bear in mind that adding capacity and network management represent alternative solutions to the same problem and that the preferred solution should vary from place to place, from technology to technology, and over time based on the relative costs of each solution. Although adding bandwidth may often prove to be the better solution, as noted earlier, the inevitable delays in adding capacity may sometimes make it unavailable as an option. In addition, it is conceivable that for technologies such as wireless, where capacity is more constrained, network management may well prove the better solution on a contemporary basis.

A comparison of the new video offerings from Verizon and AT&T provides a nice illustration of the tradeoff. Verizon is investing \$23 billion to build its fiber-based FiOS network, which holds the promise of providing up to 100 Mbps in service. AT&T's strategy places greater leverage on the existing infrastructure by deploying a VDSL2-based network called U-verse. U-verse provides smaller amounts of bandwidth, ranging from 20 to 32 Mbps depending on a particular customer's precise location, but at the much lower cost of \$6 billion. The problem is that U-verse does not have enough bandwidth to provide video in the same manner as cable companies and FiOS. Thus instead of sending all of the available channels all the time regardless of whether anyone is watching, U-verse uses a switched digital video technology that only transmits the particular channel that the subscriber is watching at any particular time. In addition, it avoids the delays that can render video programming unwatchable by using DiffServ to give higher priority to the traffic associated with its own proprietary video offerings. In many ways, AT&T's practices represent precisely the type of content that gives network neutrality proponents pause. It prioritizes a single application (video) from a single source (AT&T) and runs the risk of allowing AT&T to gain a competitive advantage by favoring its own content over others. And yet, these practices are what has allowed AT&T to avoid having to spend an additional \$17 billion needed to deploy fiber-based solutions like FiOS.

Given the magnitude of the looming problem and the tightening of the capital markets associated with the recent recession, policymakers should avoid regulations that make higher capital investments the only solution to the problem of video-induced traffic growth and should instead permit networks to use prioritization to employ the capacity that already exists more efficiently. Placing regulatory restrictions on network management would not only degrade the service of existing customers. Increasing the amount of capacity needed to support a particular number of customers would increase the per capita expense of building new networks. This de facto increase in costs would both limit broadband deployment in developing countries and make it harder to serve

rural and other low-density populations in developed countries.

Not only do these techniques prioritize or give special treatment to certain applications and certain sources. Many of these techniques also require introducing greater intelligence into the core of the network. And many of these techniques depend on individual content providers' and subscribers' willingness to pay. Any network neutrality proposal must be carefully crafted not to foreclose any of these highly creative solutions.

I have long argued that policymakers should be careful not to adhere to any potential vision of the network architecture. Instead, they should ensure that regulatory structures are flexible enough to permit network providers to experiment with a wide range of potential solutions, while at the same time retaining the authority to step in should problems emerge. This is best accomplished by a case-by-case approach that examines practices after the fact to determine whether a concrete harm has been shown rather than through rules that categorically ban particular practices in advance (Yoo, 2004, 2005, 2006, 2008, 2009). The FCC's recent decisions have endorsed the case-by-case approach I have proposed (FCC, 2008, 2009). That said, it remains important that this case-by-case approach not be implemented in a manner that suppresses experimentation. The network engineering community is developing a wide array of solutions to these problems. Regulatory policy must be careful to ensure that end users continue to have the opportunity to benefit from the tremendous and ongoing creativity of this vital and important community.

Conclusion

The shift to a world in which every type of communications is available only through every means of transmission will require reconceptualizing almost every aspect of the regulatory regime. In many jurisdictions, it will require reform of the basic structure of regulatory institutions. In all jurisdictions, it will pose significant challenges to the existing structures of economic, content, structural, and social regulation, as well for

principles of free speech.

That said, many believe that even bigger changes are visible on the horizon. Specifically, the impending migration of all video content to packet-based networks has the potential to force commentators and policymakers to reevaluate the basic architecture of the Internet itself. Understanding how to regulate in this environment will require policymakers to develop technological expertise to match the economic and jurisprudential tools on which they have long relied. Indeed, it may challenge basic understandings about the nature of the network that may be even more profound than the challenges to traditional telephone- and broadcast-style regulation that convergence has already raised.

References

- AT&T Corp. v. City of Portland, 216 F.3d 871 (9th Cir. 2000).
- Averch, H., & Johnson, L. L. (1962). Behavior of the firm under regulatory constraint. *American Economic Review*, 52, 1052-1069.
- Baumol, W., & Bradford, D. (1970). Optimal departures from marginal cost pricing. *American Economic Review*, 60, 265-283.
- Comer, D. E. (2006). *Internetworking with TCP/IP* (5th ed., vol. 1). Upper Saddle River, N.J.: Pearson Prentice Hall.
- Communications Act of 1934, 47 U.S.C. § 153.
- FCC v. Midwest Video Corp., 440 U.S. 689 (1979).
- FCC v. National Citizens Committee for Broadcasting, 436 U.S. 775 (1978).
- FCC v. WNCN Listeners Guild, 450 U.S. 582 (1981).
- Federal Communications Commission. (2004). Petition for declaratory ruling that Pulver.com's Free World Dialup is neither telecommunications nor a telecommunications service, Memorandum opinion and order. *Federal Communications Commission Record*, 19, 3307-3324.
- Federal Communications Commission. (2008). Formal complaint of Free Press and Public Knowledge against Comcast Corporation for secretly degrading peer-to-peer applications, Memorandum opinion and order. *Federal Communications Commission Record*, 23, 13028-13061.
- Federal Communications Commission. (2009). Preserving the open internet, Notice of proposed rulemaking. Retrieved from: http://hraunfoss.fcc.gov/edocs_public/attachmatch/FCC-09-93A1.pdf
- Federal Power Commission v. Hope Natural Gas, 320 U.S.

- 591 (1944).
- Geradin, D., & Kerf, M. (2003). *Controlling market power in telecommunications: Antitrust v. sector-specific regulation*. Oxford, England: Oxford University Press.
- Huber, P. W., Kellogg, M. K., & Thorne, J. (1999). *Federal telecommunications law* (2nd ed.). New York, NY: Aspen Law & Business.
- Lessig, L. (2001). *The future of ideas*. New York, NY: Vintage Books.
- MediaOne Group, Inc. v. County of Henrico, 97 F. Supp. 2d 712 (E.D. Va. 2000), *aff'd*, 257 F.3d 356 (4th Cir. 2001).
- Miami Herald Publishing Co. v. Tornillo, 418 U.S. 241 (1974).
- Minnesota Internet Traffic Studies. (2007). Internet Growth Trends & Moore's Law, <http://www.dtc.umn.edu/mints/igrowth.html>
- Missouri ex rel. Southwestern Bell Telephone Co. v. Public Service Commission, 262 U.S. 276 (1923).
- National Cable & Telecommunications Association v. Brand X Internet Services, 545 U.S. 967 (2005).
- Odlyzko, A. M. (2003). Internet TV: Implications for the long distance network. In E. Noam et al. eds., *Internet Television* (pp. 9-18). Mahwah, NJ: Lawrence Erlbaum.
- Prometheus Radio Project v. FCC, 373 F.3d 372 (3d Cir. 2004).
- Reno v. ACLU, 521 U.S. 844 (1997).
- Smyth v. Ames, 169 U.S. 466 (1898).
- Spulber, D. F., & Yoo, C. S. (2009). *Networks in telecommunications: Economics and law*. New York, NY: Cambridge University Press.
- Swanson, B., & Gilder, G. (2007, February 20). The coming exaflood. *Wall Street Journal*.
- Telecommunications Research & Action Ctr. v. FCC, 801 F.2d 501 (D.C. Cir. 1986).
- Turner Broadcasting System, Inc. v. FCC, 512 U.S. 622 (1994).
- United States v. Southwestern Cable Co., 392 U.S. 157 (1968).
- Verizon Communications, Inc. v. Law Offices of Curtis V. Trinko, LLP, 540 U.S. 398 (2004).
- Vonage Holdings Corp. v. FCC, 489 F.3d 1232 (D.C. Cir. 2007).
- Yoo, C. S. (2002). Vertical integration and media regulation in the new economy. *Yale Journal on Regulation*, 19, 171-300.
- Yoo, C. S. (2003). The rise and demise of the technology-specific approach to the First Amendment. *Georgetown Law Journal*, 91, 245-356.
- Yoo, C. S. (2004). Would mandating network neutrality help or hurt broadband competition?: A comment on the end-to-end debate. *Journal on Telecommunications and High Technology Law*, 3, 23-68.
- Yoo, C. S. (2005). Beyond network neutrality. *Harvard Journal of Law and Technology*, 19, 1-77.
- Yoo, C. S. (2006). Network neutrality and the economics of congestion. *Georgetown Law Journal*, 94, 1847-1908.
- Yoo, C. S. (2007). What can antitrust contribute to the network neutrality debate? *International Journal of Communication*, 1, 493-530.
- Yoo, C. S. (2008). Network neutrality, consumers, and innovation. *University of Chicago Legal Forum*, 179-262.
- Yoo, C. S. (2009). Network neutrality after Comcast: Toward a case-by-case approach to reasonable network management. In Randolph J. May ed., *New Directions in Communications Policy* (pp. 55-83). Durham, NC: Carolina Academic Press.
- Yoo, C. S. (in press-a). Recent innovations in the Internet's architecture that challenge the status quo, *Journal on Telecommunications and High Technology Law*, 8. Retrieved from: <http://ssrn.com/abstract=1472074>
- Yoo, C. S. (in press-b). Free speech and the myth of the Internet as an unintermediated experience. *George Washington Law Review*, 78. Retrieved from: <http://ssrn.com/abstract=1475382>

Article history:

Received September 14, 2009

Received in revised form November 2, 2009

Accepted November 17, 2009