UPDATED AS OF 4/28/2015

FOR WORKSHOP PARTICIPANTS ONLY; NOT FOR PUBLIC DISTRIBUTION

Silicon Flatirons
A Center for Law, Technology, and Entrepreneurship at the University of Colorado

AND

Northwestern Law
Searle Center on Law, Regulation, and Economic Growth

Cable Academic Workshop

May 5-6, 2015
Chicago, Illinois

Reading Packet

Broadband Policy

Spectrum Policy

Competition Policy

Internet Competition Issues

The Evolving Multichannel Video Programming Marketpalce

Private Event Webpage

Broadband Adoption and Use:
What have we learned?

John B. Horrigan, PhD
Pew Research Center
JHorrigan@PewResearch.Org

This paper serves as an overview of more detailed work that will address the question: Five years after delivery of the National Broadband Plan (NBP), what have we learned about how to increase broadband adoption and use, one of the NBP's key priorities?

The paper will unfold as follows:

1) Reflections on how stakeholders saw the issue five years ago.
2) Discussion of what has been learned since then – which is quite a bit.
    - This portion of the paper will present original research on one of the most significant broadband adoption programs – Comcast's Internet Essentials
3) Implications for the future
4) The paper offers a model for stakeholders to promote broadband adoption and use, a model based on research and lessons learned from the past several years.

Five years ago

The best way to sum up the broadband adoption debate five years ago is that, collectively, the policy community was long on goals and aspirations, but short on proven models to guide policymakers. The Commerce Department's Broadband Technology Opportunities Program (BTOP) was in its nascent stages and its impacts would not be felt for several years. A handful of non-profits had engaged with the issue and initiatives such as Computers for Youth and many showed promise. But we didn't have a "plug and play" model that could serve as the basis for recommendations for the NBP.

Since 2010, our overall progress on home broadband adoption has been steady, but not spectacular. Commerce Department data put home broadband adoption at 68% in

2010, a number that grew to 73% in 2013 – the latest government data available. For low-income Americans, broadband adoption is still a problem. For those with household incomes under $25K annually, broadband adoption has grown from 42% in 2010 to 47% in 2013. Now, it is true that mobile access has had a huge impact on the usage landscape since then. But, for reasons having mainly to do with data caps on wireless plans for smartphones, a home broadband subscription is the anchor for people's online access. It remains an important metric for policymakers and other stakeholders.

What we've learned

Probably the biggest lesson about broadband adoption since 2010 is that the problem is solvable – but perhaps not in the way some people assumed a few years ago. Research has shown, for instance, that barriers to broadband adoption are plural. Non-adopters typically cite two or three reasons (e.g., cost, digital skills, lack of relevance) why they don't subscribe to broadband. That means focusing on one lever – say, price of a monthly subscription – is not likely to move the dial much. Recent research conducted by Jay Schwarz at the FCC and colleagues at Connected Nation (entitled "The Willingness to Pay for Broadband of Non-Adopters") found that two-thirds of non-adopters said they would not subscribe to broadband even if the price were zero.

The key lesson here is that solving the broadband adoption problem is not about devising the right price-based mechanism to change people's behavior. Rather, it is about building capacity at the local level – at institutions non-adopters trust. The problem calls for drawing people into broadband use by showing them what the Internet can do for them and giving them the skills to trust it and use it.

Models for the future

If encouraging trust and cultivating skills can draw people online, what is the right program design to do those things? Models that have arisen over the past several years offer guidance. Some are from the programs BTOP has funded. The private sector has played a key role too, most prominently Comcast's Internet Essentials, but also Google's "digital inclusion" efforts in Google Fiber cities. These different models add up to a general model for broadband adoption & use. It's captured in three words:

- Partnerships
- Engagement
- Training.

This is the PET model for promoting broadband adoption, use, and digital readiness. Elaborating on each element:

Partnerships: Several years ago, NTIA sought to understand "best practice" in its BTOP-funded projects that focused on sustainable broadband adoption and public computing centers. The resulting "Broadband Adoption Toolkit" shows the importance of developing partnerships with established and trusted neighborhood institutions to promote adoption. Communities should develop a broadband adoption plans to meet citizens' needs. Private sector partners are crucial to boosting public awareness of programs, offering training, and supply discount computer and home Internet subscriptions. Partnerships, in other words, were indispensable to illuminating for people the value of having broadband access.

Engagement: Programs have to "meet people where they are." That is, they have to appeal to people to change established routines that do not rely on the Internet to do things. This is a place-based strategy that calls for taking advantage of existing community institutions to promote digital engagement. Pew Research Center research demonstrates how libraries often fill this role. Over half of all library users say that the availability of computers and Internet access at libraries is important to them and half say libraries' resources for finding jobs or applying for them is important.

Comcast's Internet Essentials program is very relevant to the digital engagement equation. IE aims at low-income families with children and offers training, a $150 computer, and a $10 per month home Internet subscription plan. By aiming at something likely to be relevant to the target population – their kids' education – IE clearly is addressing a need for its target population.

Research I have done surveying IE customers shows that the service helps IE households meet expectations of connectivity. In "The Essentials of Connectivity", published last year, I found that IE households overwhelmingly say that their kids' schools expect that they have broadband at home. Notably, strong majorities also say that other institutions – such as banks, health care providers, and government – also expect that people have access at home. These institutions basically assume the Internet's relevance to users and that delivery of services digitally will unfold seamlessly. But that's a risky assumption. The "Essentials of Connectivity" research has a different lesson: It shows that programmatic intervention can help overcome the relevance barrier for certain non-adopting groups and promote digital engagement.

Training: This is final part of the model and, on its face, a bit of a no-brainer. If a key barrier to broadband adoption is dearth of skills with computers and the Internet, then investing in training is a common-sense strategy. But it's no sure bet that such investments will pay off. Yet there is a growing body of evidence that such training has payoffs. My recent work on Comcast Internet Essentials – called "Deepening Ties," released last week – looks at IE customers as they've traveled along their broadband adoption paths through time. The research design is longitudinal, in that I surveyed over 700 IE households at two points in time – the first survey in January 2014 soon after they got service and the second again eight months later in September 2014. This has not been done before in broadband adoption research. The findings are striking. Those who had formal training on how to use the computer or the Internet were more likely to use the Internet to:

- Stay in touch with family and friends
- Look for or apply for a job
- Access government services

The differences were substantial – as much 15 percentage point compared to those who had not had training. The longitudinal design offers strength in drawing inferences on the findings. The design allows for impacts on behavioral and attitudinal measures to be controlled for respondents' baseline levels in digital skills in the first (January 2014) survey, growth in digital skills over time, as well as past online experience (i.e., whether respondents had *ever* had home Internet service before becoming IE customers).

The following table shows differences in results for those who had formal training on how to use the Internet and computers versus those who did not.

Comparing activities for those with training and those without

	Got training (IE, library, community center, other)	No training
Do school work	86%	81%
Stay in touch with family, friends, and neighbors	65	53
Access entertainment like videos, movies and online games	56	52
Look for or apply for a job	55	38
Get access to government services	44	30
Get access to banking and financial services	44	44
Look for or start a business	18	14

What was the source of the training people had? Trusted community institutions – libraries, community centers, and IE training programs that were often run through community institutions. The "Deepening Ties" research also found that there is more to be done when it comes to training. Only one-third (31%) of IE households had some sort of computer or Internet training – suggesting that encouraging more people to take advantage of training would have payoff.

It is worth noting that these findings hold for formal training, but the survey also asked about informal types of training, such as whether someone had received training from a friend or child. Nearly half (47%) said they had received training from their children and 15% from a friend in their neighborhood. These kinds of informal training had *no statistical impact* on respondents' behaviors. Given discourse that suggests online use will unfold seamlessly for new adopters as they get help from those around them, these findings are striking and important.

This is not the only research that shows a link between programmatic interventions and increases in broadband adoption and use. Research by Karen Mossberger, Caroline Tolbert, and Christopher Anderson, entitled "Measuring Change in Internet Use and Broadband Adoption," shows that "Smart Communities" in Chicago had higher rates of broadband adoption and use when studied over a five year period. Chicago's Smart Communities benefit from programmatic interventions of various sorts with a range of support. Nine Chicago neighborhoods were beneficiaries of a BTOP grant that supported the City of Chicago, the Local Initiatives Support Corporation, the Chicago Public Library, and more, to help residents of predominantly African American and Latino neighborhoods. The research found that, when compared to similar Chicago neighborhoods without Smart Communities interventions, there were 9 percentage points gains in measures of broadband use and adoption over 2008-2013 period.

The Future

What lessons does this offer as stakeholders look to the future?

- Expand community capacity building based on the PET model – starting with the partnership component as a way to get the most out of training funds.
- Expand the scope of stakeholders involved in addressing the problem.
    - Society's growing preference that people use the Internet for important functions matters a great deal. This means more players – government agencies, banks, or health care providers – have a growing stake in a connected and digitally ready population.

Five years ago, the goal was to close the digital divide and get broadband adoption numbers up. Since then, we've learned that the challenge is tougher than we might have thought, but research and practice have shown that it is solvable. The elements of the PET model serve as a roadmap not just to get more people online, but to promote society-wide digital readiness. With the emergence of the Internet of things – along with people's abiding concerns about trust & privacy online – there will remain a need for tools to help people acclimate to the digital world. The current state of knowledge in the field shows that there are viable models for continuing to develop these tools. As stakeholders and policymakers continue to look for ways to promote broadband adoption – whether at the community level or reform of the Lifeline program, they should look to lessons learned over the past several years.

Quality Competition in the Broadband Service Provision Industry

James E. Prieger
Pepperdine University
School of Public Policy
24255 Pacific Coast Highway
Malibu, CA 90263-7490
E-mail: james.prieger@pepperdine.edu

Gabor Molnar
University of Colorado at Boulder
Interdisciplinary Telecommunications Program
Campus Box 530, Boulder, CO, 80309-0530
E-mail: gabor.molnar@colorado.edu

Scott J. Savage
University of Colorado at Boulder
Department of Economics
Campus Box 256, Boulder, CO, 80309-0256
E-mail: scott.savage@colorado.edu

August 14, 2014

Key words
Broadband service provision, quality competition, structural estimation of static discrete games, entry, conditional logit.

JEL Classification
L1, L13, L96

Note
This version prepared for 2014 TPRC| 42nd Research Conference on Communication, Information and Internet Policy. This research reports preliminary findings and should not be cited or quoted without notifying the authors.

Abstract

We conduct an empirical analysis of quality competition between broadband Internet Service Providers (ISPs), using National Broadband Map data for 2011-2013 for almost a thousand local markets in California. We examine how incumbent ADSL firms respond to competition from CLECs and cable modem service providers. We use an important quality attribute, the downstream data rate, and estimate the strategic choice of quality for broadband ISPs.

Our paper follows a static game theoretic approach to the profit maximization decision of a broadband provider that leads to a simple two-stage method of estimation of the structural parameters of the ISPs' profit functions. The method accounts for both the strategic aspect of each firm's quality decision, as well as the endogeneity problems inherent in the estimation problem.

Our results include two main findings. First, ILECs improve the quality of their ADSL offerings when a cable player enters the market, and also when cable operators start to offer DOCSIS 3.0 speeds. Second, ILEC ADSL providers do not raise their service quality in response to ADSL competition from CLECs, but they do boost their speed when CLECs deploy fiber in the market. This research represents the first step in what we hope to be a major advance in the empirical analysis of broadband provision, where little structural econometric work has been done.

I. Introduction

Empirical studies to assess the relationship between competition in service provision and the quality of service offered to consumers are few. Furthermore, much of the scant empirical work that examines the broadband Internet Service Providers' (ISPs') choice of quality is not based on rigorous microfoundations for the firm's strategic decisions for entry and quality choice.

The objective of our study is to conduct an empirical analysis of quality competition among ISPs. We take advantage of recent advances in the industrial organization literature on feasible estimation of discrete games to model and estimate the determinants of an ISP's decision to enter a local market, and what speed of service to offer upon entry. Our research goal is to characterize the fundamentals of broadband services provision in the US with a structural model, allowing us to address various hypotheses concerning how firms respond to the entry and quality decisions of their rivals. The model will eventually allow us to address how competition would evolve under various policy counterfactuals that affect competition, although we do not pursue that in this preliminary version of the paper.

Acronyms

ADSL	Asynchronous Digital Subscriber Line
AME	Average Marginal Effect
CBG	Census Block Group
CLEC	Competitive Local Exchange Company
CPUC	California Public Utility Commission
DOCSIS	Data Over Cable Service Interface Specification
DSL	Digital Subscriber Line
FIRE	Finance, Insurance, and Real Estate
FTTN	Fiber to the Node
ILEC	Incumbent Local Exchange Company
IO	Industrial Organization
ISP	Internet Service Provider
Kbps	Kilobits per second
KF	Kilofeet
MADTR	Maximum Advertised Downstream Transmission Rate
Mbps	Megabits per Second
ME	Marginal Effect
MEMdn	Marginal Effect at the Median
MLE	Maximum Likelihood Estimation
NBM	National Broadband Map
SE	Standard Error
TV	Television

Our econometric model draws on the work of Bajari, et al. (2010), who propose a two-stage method to estimate models of strategic interactions for discrete strategy spaces. In the first stage, we estimate reduced form choice probabilities for the entry and quality decision of each potential entrant in a market. The estimated choice probabilities are then used in the second stage to estimate the structural parameters of the firms' profit functions. The method thus accounts for both the strategic aspect of each firm's decision, as well as the endogeneity problems inherent in estimation.[1]

In this first version of our paper, we focus on competition in nearly thousand local broadband markets in California, although eventually we will extend the analysis to the entire US. The markets are small geographic areas around central offices[2] of Incumbent Local Exchange Companies (ILECs). In these markets, we examine how ILEC ADSL[3] players respond to competition from Competitive Local Exchange Companies (CLECs) and cable players. We focus on an important quality attribute, the Internet data rate, and estimate the strategic choices of maximum advertised download data rates for ILEC broadband ISPs. We draw our firm count and speed data from five waves of the US National Broadband Map (NBM).

The economic literature on competition and quality shows that a higher degree of market competition may lead to higher or lower quality of service. While more competition increases the firm's incentives to supply high quality (holding output prices fixed), more competition also reduces the price–cost margin, which reduces the incentives to invest in quality. Considered another way, if greater competition leads to stronger share-stealing effects, there will be higher equilibrium quality. However,

---

[1] The endogeneity problem arises from the familiar simultaneity problem in incomplete information games: in equilibrium, the competitors' actions depend on their expectations about the firm's action, and vice versa.

[2] A central office (or "wire center") is the location where the telephone company's network switching equipment is installed. There is one central office for each local telephone exchange in the traditional Public Switched Telephone Network (PSTN).

[3] Asynchronous Digital Subscriber Line (ADSL) is a technology that enables broadband data transmission over the copper telephone lines in the local loop of the telephone exchange. ADSL service, as defined for purposes of the National Broadband Map and therefore this study, may involve use of copper lines all the way from the central office to the subscriber's premises, or may make use of fiber from the central office to a remote terminal. The common element is that ADSL involves using the copper telephone wires for the last part of the transmission path to the subscriber's premises. ADSL offers faster download than upload speed, as opposed to symmetric DSL (SDSL). SDSL is primarily a business product.

given that the elasticity of demand with respect to quality need not increase with competition, the premise may not hold. Thus, the net effect of competition on quality is a priori uncertain and empirical measurement is necessary. Outside of a few markets such as healthcare, airlines, and retail gasoline, this has rarely been done in the literature.

Our paper contributes to the existing literature by showing empirically that ILECs appear to respond to intermodal but not intramodal competition. ILECs improve the quality of their ADSL offerings when a cable operator enters the market, or when the incumbent cable operator deploys DOCSIS 3.0. We also found evidence that ILECs do not raise their ADSL service quality when the competing CLEC is offering ADSL only, regardless of speed, but that ILECs do boost their speed when CLECs deploy fiber in the market.

Future work will investigate how CLECs ADSL and cable modem players respond to intermodal and intramodal competition. Our future analyses will also include competition from fixed and mobile wireless broadband. Our results will shed light on whether policymakers should encourage more intramodal competition, for example, by easing entry through unbundling of local network elements or policy prohibiting exclusive video and internet franchises by localities. In addition, we can also assess the efficacy of the US strategy of encouraging mostly intermodal competition (DSL and fiber offered by AT&T and Verizon vs. cable modem service offered by Comcast and other CATV providers). The research represents the first step toward a major advance in the empirical analysis of broadband provision, where little structural econometric work has been done.[4]

The next section reviews the literature on competition and quality and provides an overview of the literature on entry games with a special focus on static games. Section III details the data, and

---

[4] The main structural contribution is by Nevo et. al. (2013), who examine the welfare effects from usage-based pricing and demand for residential broadband.

Section IV describes the empirical model. Results are presented in Section V. Section VI concludes with discussion of the results and outlines our plans for future work.

II. Background and Review

A. Competition and Quality

The relationship between competition and quality has attracted much recent attention from US policy makers. In particular, the education, electricity, finance, health, media and telecom sectors have experienced extensive legislative reform by state and federal governments intended to promote consumer choice, greater product variety, and increased quality through greater competition. In his examination of healthcare, Katz (2013) explains the key justification for this approach as being the "…intuition that, due to the potential to steal market share from rivals, a competitive care provider has stronger incentives to raise its quality to attract patients." Furthermore, this intuition is conditioned on the "belief that greater competition leads to stronger share-stealing effects and, thus, higher equilibrium quality." In this section we consider the theoretical and empirical evidence concerning this intuition.

1. Theoretical literature on competition in quality

The straightforward intuition that share stealing will lead to a positive association between competition and quality is supported by economic theory when prices are fixed, as is the case for some regulated markets (e.g., Douglas and Miller, 1974; Schmalensee, 1977). As long as the fixed price is set above marginal cost at some base level of quality, firms will increase quality in an attempt to gain market share.

When prices are not fixed, more competition will also lower the price–cost margin and this may reduce the incentives to invest in quality. Gaynor (2006) uses the Dorfman-Steiner (1954) condition, adapted to monopolistic competition, to show that the amount spent on quality by the firm depends on the ratio of the quality elasticity of demand to the price elasticity of demand. When an increase in market power reduces both elasticities, quality may increase or decrease depending on the relative strengths of the two effects. Gaynor notes that similar intuition is provided by several other studies but within a different modeling framework. For example, Kranton (2003) studied the effect of competition on quality when consumers are imperfectly informed about quality. Her model shows that if firms compete in price for market share, both price and quality can be lower, which is analogous to the price elasticity of demand exceeding the quality elasticity of demand.

Matsa (2011) describes the tradeoffs facing the firm in the short and long run. He notes that lower profit margins under more competition reduce the immediate cost of losing a "sale" so firms may shade quality. In the long run, however, competition may raise the likelihood that unhappy consumers switch to a competitor, so firms improve quality. In their growth model with incremental innovations, Aghion and Howitt (2009) show that competition fosters innovation in sectors where firms operate at the same technological level. Here, competition reduces pre-innovation rents and thereby increases the incremental profits from innovating and becoming a leader. In other sectors, competition reduces the post innovation rents of laggard firms and thus their incentive to catch up with the leader. Chen and Schwartz (2013) also use a model of innovation to outline conditions where the incentive to add a new, higher-quality product can be greatest under monopoly. The monopolist loses more profit on the old product but may earn more profit on the new one because it prices the old product in a way that internalizes the effect on the new one.[5] While these studies represent only a few examples, they are

---

[5] The key factor is the extent to which the monopolist can divert sales to the new product as opposed to leaking sales to outside goods if it raises the price of its old product (Chen and Schwartz, 2013).

illustrative of the overall finding from the theoretical literature. Competition can lead to lower or higher quality, depending on the underlying properties of demand, costs and information.[6]

2. Empirical literature on competition in quality

Given the theoretical ambiguity in outcomes, it is not surprising that recent empirical studies have produced mixed results on the relationship between competition and quality for different industries with different market conditions. The basic empirical approach has been to write down a firm's equilibrium quality function as the implicit solution to their profit maximization problem. A reduced-form quality equation is then specified that relates some measure of quality to cost and demand shifters and to a measure of the number of firms in the market. For example, Mazzeo (2003) shows that average flight delays are longer in more concentrated airline markets. Goolsbee and Petrin (2004) estimate that cable television (TV) channel capacity, number of over-the-air channels and number of premium movie channels increased in response to satellite entry, while Savage and Wirth (2005) document a similar effect with respect to potential entry from cable overbuilders.[7] Matsa (2011) finds that supermarkets facing more intense competition have more products available on their shelves, while Olivares and Cachon (2009) show that the inventories of General Motors dealerships increases with the number of competitors.

In contrast, Domberger and Sherr (1989) find no correlation between the threat of new entry and customers' satisfaction with their attorney used for home purchases. Prince and Simon (2013) show

---

[6] This ambiguity has a long history in industrial organization (IO) theory. Chamberlin (1933) and Abbott (1955) show that firms with market power may reduce product quality to save costs and maximize their profits. Swan (1970, 1971) demonstrated no relationship between monopoly power and product quality and defined conditions under which a competitive and monopoly market introduce a product with the same level of quality but the monopoly will charge a higher price. Schmalensee (1979) shows that this result holds up under some relaxation of the original assumptions but questions whether quality choice under oligopoly will be well approximated by either the competitive or monopoly models.

[7] An overbuilder in the cable industry is a second entrant into an existing cable franchise area to compete with the incumbent. Overbuilding using hybrid fiber-coax networks (i.e., a traditional cable system architecture) is relatively rare in the US in general and in California in particular.

that flight delays for incumbent airlines worsen in response to entry threats by Southwest Airlines. Chen and Gayle (2013) examine mergers and product quality (i.e., the ratio of non-stop flight distance to total flight distance used to get passengers from origin to destination) for the airline industry. They find that quality increased in markets where the merging airlines did not compete *ex ante*, and decreased in markets where they did. This is consistent with their theory that mergers improve coordination but diminish competitive pressure for firms to provide high quality products.

Similar studies have also been conducted with advertising as the proxy for quality. Dick (2001) examined the US retail banking industry using higher advertising intensity (i.e., marketing expenses divided by total asset value) as a measure of higher customer service quality. He found that dominant banks provide a higher level of service quality than fringe banks. Crawford (2007) analyzed the relationship between TV station ownership and the quality of their programming. He found no relationship between cross ownership with a local newspaper or radio station and the number of minutes of advertising included in TV programming, where more minutes indicating lower quality TV service. Hiller et al. (2014) analyzed consumer media bundles and showed a positive correlation between the number of independent TV stations and the amount of time and space devoted to advertising.

In telecommunications, Wallsten and Mallahan (2013) find that the number of wireline ISPs in a US census tract is positively correlated with the highest advertised downstream speeds. Nardotto et al. (2012) show a positive relationship between lower barriers to entry, measured by the presence of local loop unbundling, and average broadband download speeds in the United Kingdom.[8] Molnar and Savage (2013) show that wireline speeds are higher in census block groups with two or more wireline ISPs than

---

[8] Unbundling requires the incumbent telephone company to lease the connection from their central office to the household (“local loop”) to new entrants so they can compete in the final product market for broadband Internet.

with a single wireline ISP, but there is no relationship between wireline speeds and the number of wireless ISPs.

While reduced-form quality equations provide useful insights into the general relationship between competition and quality, they say nothing about the strategic interactions between firms with respect to their quality choices. Kugler and Weiss (2013) use a reaction-function approach to estimate the strategic quality choices for Austrian gas stations. Their empirical reaction function relates the opening hours of a station to those of its competitors. Their results suggest significant but imperfect coordination, in opening hours among stations of the same network, which implies that opening hours are strategic complements. They find a similar but weaker effect between independent stations or stations from competing networks. Brueckner and Luo (2013) use a similar model to investigate strategic interaction among US airlines in flight frequency. Using instrumental variables estimation, a positive reaction function is found in some specifications, suggesting complementarity in the choice of frequencies.

In summary, there is much work in the field of Industrial Organization (IO) on the effect of competition on prices, but not nearly enough has been done on quality. Both parties to a merger law suit would benefit from some empirical evidence in this direction. Lack of evidence on this is mostly due to the lack of data on quality. Schmalensee (1979) noted 35 years ago that “it is far from obvious that any single mathematical representation of quality can serve for a broad spectrum of products.” Even more so today, most industries sell highly differentiated products, making standardized quality measures difficult to collect, and the few previous studies looked at flight delays, product availability in supermarkets, and the number of TV channels. These are worthwhile quality dimensions in their respective industries, but pale compared to the importance of Internet speed in the broadband industry. This paper investigates an essential and standardized quality attribute – Internet speed in a highly relevant industry in the digital age – and estimates the strategic choices of download and upload speeds for Californian broadband providers. Estimation of the static model of strategic interactions with discrete-choice methods determines the probability of a particular level of quality for a representative broadband provider as a function of the expected quality choice of rivals and various market characteristics.[9]

B. Broadband Market Entry

When entering new markets, or re-evaluating their business plans regarding technology or quality in an existing market, ISPs face many decisions: which technologies to offer, what packages to create, how much to invest in service quality, what prices to charge, and how to promote service offerings. The ISP's customers can decide on the type of contract, what service level they purchase, and what additional products they take with a service bundle. ISPs must also consider the strategic reactions of the rival firms. Will they enter the market? How will a competing firm position its market play? The interrelated nature of these decisions suggests modeling them with empirical discrete games that can assume sequential or simultaneous move by the players.

We model the broadband Internet markets as a repeated simultaneous game. Simultaneous games are imperfect-information games; players do not have the knowledge about the actions of the others. In our model, ISPs choose a quality simultaneously, and they do not know the current-period actions of the other firms.[10] When we observe broadband markets and make an attempt to understand how the players behave, we also lack information on price, cost, or demand data. We can observe,

---

[9] Xiao and Orazam (2011) estimate a simple discrete-choice model of broadband entry and find that sunk costs are an important determinant of wireline entry in US zip codes. However, they are unable to distinguish between one, two or three providers due to data confidentiality, and they do not estimate the direct effects of entry on market outcomes such as quality.

[10] The reality is, or course, more complex; ISPs do not actually choose a service quality once per six months all at the same time. However, our modeling approach is commonly adopted in the literature when there is no clear first mover and is best seen as an approximate structure designed to reflect uncertainty regarding competitors' plans.

however, the entry and exit of players, the speeds that they provide, along with market demographics, and make inferences even in the case of incomplete information.

Inference about structural parameters of the profit function from observations on entry was made possible by Bresnahan and Reiss (1990) and the subsequent stream of literature triggered by their seminal paper. Bresnahan and Reiss inferred the effects of entry on competition from the relationship between the number of market entrants and the market size. By observing strategic entry decisions of small retail firms in isolated rural markets, they argued that firms must pay a fixed and sunk cost to enter the market. They also argued that the total industry profit depends on the number of firms on the market but not on the identity of the entrants. Bresnahan and Reiss proposed an estimator that maximizes the likelihood for the number of firms and introduced the idea of entry thresholds, i.e., the market size required to support a given number of firms. The two main disadvantages of their model are that firms' costs are homogenous, and that the firms do not offer differentiated products. In their later work (1994), they estimated firms' sunk costs from differences in the thresholds for entry and exit.

Berry (1992) relaxes this limitation by allowing heterogeneity between firms entering the markets. He develops a model of market entry considering a large number of heterogeneous potential entrants and applies the model to analyze competition in airline markets. Berry recommends using simulation methods to address the computational problem of calculating the linear combination of integrals that define the probability of events. Mazzeo (2002) extends the Bresnahan-Reiss model by allowing firms to offer heterogeneous (high-quality and low-quality) products. Using data from motel markets along US interstate highways, and endogenizing the quality choice of firms, he founds that hoteliers have strong incentives to differentiate. Ciliberto and Tamer (2009) broaden the literature by allowing for heterogeneity without making equilibrium selection assumptions. Applying a pseudo maximum likelihood estimation method to the US airline industry, and expanding on Tamer's earlier work (2003), they find evidence of heterogeneity across airlines in their profit functions.

Additional recent contributions include Seim (2006), Aguirregabiria and Mira (2007), and Bajari et al. (2010). Seim's static equilibrium model makes early use of spatial econometrics in market structure and product type choice studies. Her simulation results demonstrate the firms' incentives for spatial differentiation in the video rental industry and the importance of incorporating product-type choices into the market entry process. Aguirregabiria and Mira (2007) propose a two-step method to estimate static games of incomplete information and illustrate it using an example of a static game of market entry. Their method greatly reduces the computational complexity of earlier approaches, and the present work derives from theirs. In the spirit of Aguirregabiria and Mira (2007), Bajari, et al. (2010) implement a two-stage method to estimate models of strategic interactions for discrete strategy spaces. In the first stage, they estimate reduced form choice probabilities for the entry and quality decision of each potential entrant in the market. Then, in the second stage, they use these computed choice probabilities to estimate the structural parameters of the firm's profit function. As an application for the two-stage model, they study the determination of stock recommendation issued by equity analysis for high-tech stocks between the years of 1998-2003.

In telecommunications, employing the entry model in Mazzeo's (2002) study, Greenstein and Mazzeo find (2006) evidence that competitors are heterogeneous and that firms account for both potential market demand and the business strategies of their competitors when making their entry decisions. Following Bresnahan and Reiss (1991), Xiao and Orazem (2011) estimate a discrete-choice model of broadband entry, as discussed above.

In addition to these two works, in which estimation is based directly on theoretical entry models, most studies of market entry in broadband are nonstructural (reduced-form). Almost all of the existing works have been extracted from cross-sectional, static studies of existing players; the impact of potential new entrants are not studied. In a typical broadband market entry study, a cross-section of either the number of ISPs or an indicator for the presence of at least one competitor in the local area is regressed on demographic and other market characteristics (Prieger, 2003; Grubesic and Murray, 2004; Flamm, 2005; Prieger and Church, 2012; Prieger, 2013). These studies show that the decisions of telecom service providers to deploy network resources and offer service in a local market depend on both economic and regulatory considerations. Demand factors such as market size, average income, and other demographic characteristics all been shown to affect broadband penetration (Prieger, 2003; Grubesic and Murray, 2004; Flamm, 2005; Flamm and Chaudhuri, 2007; Prieger and Hu, 2008; Prieger, 2013). Some of these papers also show that population density or terrain also influence broadband penetration in the expected ways, and can be used as proxies for cost. For example, rural areas are more likely to be served only with lower-speed broadband or by few providers, or less likely to have broadband available at all than urban areas, due to low population density and rougher topography, (Stenberg et al., 2009; Li et al., 2011; Prieger, 2013). Intermodal and intramodal competition among broadband ISPs, both actual and potential, also affects the incentives to enter the local markets (Denni and Gruber, 2007; Prieger and Hu, 2008; Wallsten and Mallahan, 2013).

Like the work of Greenstein and Mazzeo (2006) and Xiao and Orazem (2011), we perform structural estimation to identify parameters of the potential entrant’s profit function. Unlike the earlier studies, we are particularly interested in those parameters relating to quality competition. We hope that our work will thus be a significant addition to the scant structural empirical work on broadband competition.

III. Data

In this paper, we focus on competition in broadband service provision in local markets in California. We chose California because it is large enough to contain many local markets, yet not so large as to make working with the voluminous NBM data unwieldy. In future versions of the paper, we hope to expand to other states.

A. Market definition

Any definition of the broadband Internet market is only an approximation of how ISPs may view their market play. Market definition is made difficult because the natural areas of deployment for different types of broadband providers, e.g., wire center serving areas and cable franchise areas, do not exactly match. Previous studies have used counties, census tracts, ZIP codes, and local telephone exchange boundaries to define the geographical market for broadband Internet (Gillett and Lehr, 1999; Prieger, 2003; Wallsten and Mallahan, 2013; Xiao and Orazem, 2011; Nardotto et. al., 2012; Prieger and Conolly, 2013; Prieger, 2013). Our market definition instead is similar in spirit to that of Prieger and Hu (2008b), who carefully examine the distance from the local phone company's central office to define the local markets for ADSL. Roughly speaking, our 965 broadband Internet markets are small geographic areas within the distance of 2.3 miles of the ILECs' central offices in California.[11] The main reason behind our market definition is that in areas close to an existing ILEC wire center, the incumbent ADSL player has the greatest ability to match the speeds of fiber and cable modem competitors, due to the degradation of DSL speed with line distance. The rest of this subsection gives further details on the market definition process and can be skipped by readers not interested in the technical details.

Our definition of the markets for the entry game is a three-step process. Market definition begins with drawing a circle of radius 12,000 feet (12 kilofeet (kf), about 2.3 miles) around each California ILEC wire center found in the NECA tariff #4. The threshold of 12 kf (along with a secondary threshold of 18 kf, discussed below) was chosen in accord with California Public Utility Commission methodology for validating information on the provision of DSL (CPUC, 2013). A radius of 12 kf from the equipment in the wire center also corresponds to the straight-line threshold for provision of DSL at 6.3

---

[11] For robustness, we also studied markets with a circle of radius of 18 thousand feet from an existing central office, and will also show results using this geographical delineation as well.

Mpbs.[12] Since last-mile network may be constrained to run along right-angled streets, a 12 kf radius by the Euclidean metric has a worst-case situation where the lines from the wire center have a taxicab distance of 18 kf long,[13] in which case DSL of at least 1.5 Mbps is possible. These speed limitations are relaxed in many markets by the installation of remote terminals to neighborhoods farther from the wire center, as in AT&T's fiber-to-the-node (FTTN) architecture for its U-verse service. However, we have no data on which markets include remote terminals.

In a second step, we limit the areas defined by the circles to actual wire center serving areas of the ILECs. Using GIS data from GDT on the service territory of the ILECs associated with each wire center, parts of the 12 kf radius circles not also in the actual service territory were excluded from each market. This step ensures that in dense urban areas, where wire centers are closer to each other, the market area associated with one wire center does not overlap with the territory served by an adjacent wire center.[14]

A third step is necessary to match the second-step market areas to the broadband provision data in the NBM, which is keyed to Census geography. The third-step market area, therefore, consists of the union of all Census block groups (CBGs) that lie wholly within an area from the second step. In a few rural locations, no CBG are contained within the areas from step two, and in such cases we instead use the set of CBGs that overlap the step-two area. All GIS processing for market definition was performed

---

[12] For the relationship between distance and ADSL speed, see http://whatis.techtarget.com/reference/Fast-Guide-to-DSL-Digital-Subscriber-Line.

[13] The taxicab distance, defined as the distance between two points measured along axes at right angles, is also called Minkowski's L_1 distance. The worst-case scenario is the maximum taxicab distance for a fixed Euclidean distance, and occurs when the communications lines run along the legs of a right triangle with hypotenuse equal to 12 kf in length. The line length to reach the 12 kf radius is about 16.97 kf in this case.

[14] In a few markets (29) the purported coordinates of the wire center from the GDT and NECA data sources did not agree to within 2 V&H units (about 0.6 miles) using the L2 norm. (The telephone industry uses a unique "V&H" coordinate system for central office locations.) For such wire centers we did not use the GDT wire center serving area to limit the market. Instead we used the entire area defined by the 12 kf radius around the wire center (as located using the NECA data) less any area already part of another market.

using ArcMap. The market definitions result in 965 markets.[15] For use in testing the robustness of the econometric conclusions, an equivalently constructed set of markets based on an 18 kf (about 3.4 miles) radius were also created.

In summary, our definition results in a set of local broadband markets that are distinct, small enough to represent an ILEC's decisions about infrastructure in a single wire center area for DSL, yet large enough so that local decisions about infrastructure and quality of service do not affect multiple markets.[16] Figure 1 shows the step-two and step-three market areas throughout the state. In Figures 2 and 3 we show a detailed view of some of these markets. The first figure shows some of the Los Angeles urban area, in which the markets are often constrained more by wire center boundaries than by the 12 kf radius. This is most apparent in the West L.A. markets in the upper left and the downtown L.A. markets in the upper right areas of the figure. The heavy dots on the map mark the ILEC central office locations, the blocky areas surrounding the points are the market areas from step three (each a collection of CBGs), and the larger, circular or smooth-bordered areas are the market areas from step two (the intersection of the 12 kf and wire center area boundaries). The other figure shows some extremely rural markets. A few markets at the top of Figure 3 are like those in Figure 2, where at least one CBG falls entirely within the step-two market area. The large markets in the middle of Figure 3 show examples of the few markets composed of CBGs that overlap with the step-two market area (because no CBG is wholly contained within it). These CBGs with low population density can be quite large, and this accounts for much of the rural areas in the right panel of Figure 1 being colored in. This is not likely to lead to overstating broadband presence, however. Whether an extremely rural central

---

[15] Four of the potential market areas were dropped after the second step because they were very small, and an additional market defined for a wire center on the Oregon border was dropped because it appeared to serve customers in Oregon instead of California.

[16] By which we mean infrastructure deployment in the central office and within the same wire serving area. Of course, backhaul infrastructure such as high capacity transmission lines between central offices or connections to the Internet backbone may affect multiple markets.

office is placed into a large or small market area, the maximum speed of any broadband provision of any type is highly likely to be present near the central office, which is typically in the center of town.

B. Broadband data

We draw data on the location and quality of broadband service in California from five semiannual waves of the US National Broadband Map, June 2011 to June 2013.[17] These were the latest data available at the start of this project. We chose not to include the first two rounds of the NBM data, from 2010, because those rounds used an earlier Census geography. We matched ISPs offering service anywhere in the market areas to the corresponding markets and recorded each firm's maximum advertised downstream rate, separately by technology and holding company.[18] While in theory the NBM also records actual transmission rates, those fields are missing for many firms, and we use the advertised rates instead. Technologies covered in the NBM include the ADSL, fiber, and cable modem services we investigate in this version of the paper, as well as wireless and less commonly used wireline technology.

In general, information on potential entry that did not occur cannot be found in the NBM. However, since the markets are defined around ILEC locations, the ILEC ADSL potential entrant is obvious. The ADSL quality choices of ILECs in California markets are shown in Figure 4. The data include 14 ILECs in 965 markets over four periods, choosing among five quality alternatives, for a total of 3,860 cases and 19,300 observations available for the estimations. The first-stage estimations include observations for 25 cable companies (4,365 cases and 26,190 observations) and 16 CLECs (37,743 cases

---

[17] Created from a collaboration between the National Telecommunications and Information Administration, the FCC, and all states, territories and Districts of the US, the NBM is an online tool that provides semi-annual information on the broadband service providers, their product type, technology, and their maximum advertised upload and download speeds in each US census block.

[18] Service providers are aggregated to the level of their holding company, even if they operate in the same market with multiple operating companies. We used a master list of holding companies constructed by one of the authors for previous broadband research that includes all firms appearing in the FCC Form 477 broadband filings in recent years. Our list of holding companies account for variation in company names, mergers, acquisitions, spin-offs, and cable system area swaps.

and 226,458 observations for ADSL; 52,234 cases and 313,404 observations for fiber). The set of potential entrants for CLECs for a market and technology type includes any CLEC offering service anywhere in California (except when the CLEC is already an ILEC in the market). The set of cable modem entrants includes any firm with a franchise area that at least partially overlaps a market.

For cable firms, the locations for entry into broadband service provision is limited by the extent of their franchise areas. In California, new cable franchises are awarded by the state, and the CPUC makes available GIS shapefiles of state-franchised areas.[19] We used these data to construct a variable measuring what fraction of the market area is covered by the franchise area. In one of our robustness tests, we weight the cable modem competition variables by this variable to account for market coverage that is less than complete.

C. Demographic data

Most of the demographic data for the markets come from Geolytics, based on the 2010 Census and 2008-2012 American Consumer Survey data from US Census Bureau for CBGs. However, to improve the precision of the population and household density variables, we instead counted population and households in Census blocks falling into our step-two market areas, and divided each by the step-two market area in square miles.[20] Similarly, the regressor for market area is for the step-two definition. We used the County Business Patterns 2011 data of the US Census Bureau to get information on finance,

---

[19] Not all franchise areas were awarded by the state in the past, however. Some legacy locally-awarded franchise areas with long terms are missing, and the fraction coverage variables described below in the text are missing for those. As local franchises expire, they are converted to state franchises.

[20] This avoids the difficulty that some of the step-three market areas are overly large in rural areas, even though the locus of economic activity is near the central office. If we included population and area of the entire step-three market areas, the resulting densities would be misleadingly low.

insurance, and real estate (FIRE) employment in our markets.[21] Table 1 contains summary statistics for the variables used in the study.

IV. The Econometric Model

A. Game-theoretic underpinnings

Our structural econometric model is based on a static game theoretic approach to the profit maximization decision of a broadband provider. We adopt the approach of Bajari et al. (2010) for estimation of static games of incomplete information with multiple equilibria, and we refer the interested reader to their article for presentation of the model at a high level of mathematical formality. The static game approach is a generalization of a discrete choice model that allows the quality choice of a firm to depend on the actions of the other firms. Firm *i* in market *m* at time *t* chooses an alternative $a_{imt} \in A = \{0,\ldots,J\}$ representing a quality level, where alternative 0 represents offering no broadband at all. Let the firm's profit u_{imt} be

$$u_{imt}(a_{imt}, a_{-imt}, s_{imt}; \theta) = \pi_{imt}(a_{imt}, a_{-imt}, s_{imt}; \theta) + \epsilon_{imt}(a_{imt}) + \eta_{imt}$$

where a_{-imt} represents the actions (chosen alternatives) of the other potential entrants in the market and period, s_{imt} is a vector of firm *i*'s state variables affecting profit, and θ is a finite vector of parameters. The state vector is assumed to be common knowledge to all firms, but $\epsilon_{imt}(a_{imt})$ is private information for firm *i*. For identification, we assume that after accounting for the actions of the other firms through argument a_{-imt}, the state variables of the other firms (s_{-imt}) do not affect directly firm *i*'s profits. This exclusion restriction will be used to identify the parameters of the deterministic

---

[21] The county level employment data were linked to our markets by calculating which county or counties each market is in. Since the FIRE employment variable describes the composition of employment in the are instead of counting employees, it is reasonable to apply these county-level data to our markets. When a market falls into more than one county, the data from the multiple counties is averaged, weighted by the market area falling into each.

part of the profit function in the two-step estimation. The final term, η, can be either private or common information, and includes all factors specific to the market, period, firm, or any combination of these that affect the profit of all alternatives equally. For example, η_{imt} can be a market-firm fixed effect η_{im} such as a firm's long-standing reputation in the area or a period fixed effect η_t stemming from the business cycle. The state variables include some factors, x_{imtj}, that vary across alternatives $j \in A$ and others, z_{imt}, that do not. The actions of the other firms enter observed profit through a set of competition variables $w_{imt}(a_{-imt})$. Observable profits are assumed to be linear in the state and competition variables:

$$\pi_{imt}(a_{imt}, a_{-imt}, s_{imt}; \theta) = \gamma_k' z_{imt} + \beta' x_{imtj} + \delta' w_{imt}(a_{-imt})$$

The firm does not observe the actions of other firms before choosing its action. Suppressing time and market subscripts, the firm's *expected* profit is

$$U_i(a_i, s_i; \theta) = E\pi_i(a_i, a_{-i}, s_i; \theta) + \epsilon_i(a_i) = \gamma_k' z_{imt} + \beta' x_{imtj} + \delta' E w_{imt}(a_{-imt})$$

where the expectation is taken over the space of other firms' private information. See Bajari et al. (2010) for a precise statement of this expectation. Informally, if it could observe the private information, firm *i* would know what each other firm would do. I.e., by assuming the other firms want to maximize profit, firm *i* can calculate the other firms' decision rules for quality choice that map their private information into their action space *A*. Taking expectation over the private information of the other firms yields an expected set of resulting competition variables, Ew_{imt}.

B. Estimation

We assume that $\epsilon_i(a_i)$ is drawn from the extreme value distribution as in the logit model. Then the firm chooses alternative $a_i = j$ such that $U_i(j, s_i; \theta) \geq U_i(k, s_i; \theta)$ for all $k \neq j$. Given the logit structure of the error terms, the probability that the firm chooses quality level *j* is thus

$$\Pr(a_{imt}=j|s_{imt})=\frac{\exp\left(\gamma_k' z_{imt}+\beta' x_{imtj}+\delta_j{}'(Ew_{imt})\right)}{1+\sum_k \exp\left(\gamma_k' z_{imt}+\beta' x_{imtj}+\delta_j{}'(Ew_{imt})\right)}$$

This formulation incorporates the usual identification assumption that the profit of base alternative 0, not entering, is normalized to zero. The expression above allows estimation of θ = (β,γ,δ) by maximum likelihood estimation (MLE) using a conditional logit model for choice of quality.[22] Note that the choice-invariant fixed effects η drop out of the conditional likelihood. For the same reason, the coefficients on z_{imt} must be alternative specific for the impact of the *z* to be estimable (as is familiar from the multinomial logit model).

The econometrician observes actual w_{imt} in the data, but not Ew_{imt}. We cannot substitute w_{imt} for Ew_{imt} in estimation, because the former is endogenous due to the simultaneity of the game. In the two-step method of Bajari et al. (2010), reduced form choice probabilities for competitors are estimated in the first step. By observing quality choices in a large number of markets, the econometrician forms a consistent estimate of the equilibrium choice probabilities. In our application, we use conditional logit in the first step, where the quality choices of each firm are regressed on z_{imt} and x_{imtj}, but not the competition variables. We then use the estimated choice probabilities to form the expected values of competition variables Ew_{imt} for the second-step estimation.

Since the state variables *x* and *z* are used to identify both the effects of (β,γ) and δ on the observed choices and profit is linear, an exclusion restriction avoids collinearity problems and helps identification. As mentioned above, the state variables specific to competitors are not included in the second-step estimation for the ILEC's decision. However, those excluded state variables are used to estimate the choice probabilities of competitors *k≠ i* in the first step, which then appear in the estimate of the expected action and consequence $Ew_{imt}(a_{-imt})$. In our application, the excluded instruments are

---

[22] Estimation was performed using the `asclogit` command in Stata 13.1.

infrastructure variables of the other firms, which we describe in the next subsection. Although these affect the actions of the other firms, conditional on those actions the infrastructure costs of the other firms should not affect directly the quality choice of firm *i*.

C. Specifics

For this version of the paper, we focus on ADSL provision by ILECs, the dominant telecommunications firms in the local markets. In California, this includes the “U-verse” DSL service offered by AT&T but not the “FIOS” fiber-to-the-home service from Verizon, which instead counts as fiber-based broadband. The competitors include ADSL and fiber broadband from CLECs and cable modem broadband offerings.[23] For now, we set aside competition from less closely related offerings such as fixed or mobile wireless broadband, but recognize that these competitors may be more relevant in the future.

For estimation, the many speed categories in the NBM are collapsed into four alternatives for ADSL: greater or equal to 768kbps but less than 3 Mbps, 3Mbps to 6 Mbps, 6 Mbps to 10 Mbps, and 10 to 25 Mbps. No ILECs report offering ADSL with maximum speed below 768 kbps or greater than or equal to 25 Mbps during our time period. Choice $j = 0$ of not offering ADSL in the market gives the base alternative. These speed categories are also presented in Table 2 for reference. Variables z include demographic variables reflecting market characteristics and infrastructure variables. Variable *NearestAnySpeed* is the distance in miles (or log miles, in one of the estimations) to the Census block nearest to the center of market *m* where firm *i* offered broadband using the same technology (ADSL, for

---

[23] The distinction between ILECs and CLECs is clear within a market, because the NECA tariff identifies the locations of ILECs, and all other ADSL or fiber providers in that market must be CLECs. However, AT&T (and other large firms) may be treated as ILECs in some markets and CLECs in other markets if they do out-of-region entry.

the ILECs here) last period, per the NBM.[24] Thus, when the ILEC already offered ADSL in the market at *t*-1, *NearestAnySpeed* will be small. When the ILEC did not offer ADSL in the market the previous period but did in a nearby area, *NearestAnySpeed* will be smaller than if the firm offered ADSL in some distant location. Due to the presence of sunk costs in broadband infrastructure deployment, we thus expect that higher values of *NearestAnySpeed* will lower the probability of higher ADSL quality.

There are two *x* variables in the model: *NearestSameSpeed* and *SameSpeednotFound*. The former is constructed similarly to *NearestAnySpeed* but only ADSL in the same quality category counts in the calculations. Since this variable is missing when the firm did not offer a particular quality level anywhere in California the previous period, it is set to zero for such cases and an indicator variable *SameSpeednotFound* is set to one. *SameSpeednotFound* thus captures the impact of the variable *NearestSameSpeed* when the latter would logically be infinite. By logic similar to the above, we expect *NearestSameSpeed* and *SameSpeednotFound* for quality *j* both to impact negatively the probability of the firm offering quality *j*.[25]

The competition variables we choose are indicators for the presence of at least one competitor in a quality category. These are the *w* variables, which are functions of quality choice decision a_{-i} as introduced above. The indicators are cumulative, defined as $w_{imt}^{bj} = 1$ if broadband of type *b* and speed *j* or higher is offered in the market this period, with $w_{imt}^{bj} = 0$ otherwise. In the second-step estimations, Ew_{imt} are the expected values of these variables, where the w_{imt}^{bj} are arranged into a column vector including all *b* and *j*, and thus can take values between 0 and 1. For CLECs offering ADSL, there is an additional alternative *j* = 5 of greater or equal to 25 Mbps but less than 50 Mbps. For cable

---

[24] All distance variables were calculated based on the latitude and longitude of the central offices and Census block centroids. The great circle distance metric was computed and the nearest broadband locations for each firm and market were found using a FORTRAN program.

[25] Each *x* variable also has cross-impacts. For example, *NearestSameSpeed* for the highest quality level has a marginal effect on the probability of the firm offering ADSL in the lower quality categories. We calculate but do not report these cross-effects in the tables for the sake of brevity.

modem, the categories are *j* = 1 (less than 10 Mbps), 2 (10 Mbps to 25 Mbps), 3 (25 Mbps to 50 Mbps) 4 (50 Mbps to 100 Mbps), and 5 (100+ Mbps). For fiber, the quality categories are 0 (no entry or any fiber below 1 Gbps, grouped because there is little fiber below that speed) and 1 (gigabit fiber). These speed categories are also presented in Table 2 for reference.

Since the demographic variables are specific to the market and apply to all firms and periods, and because it is unrealistic to assume that two observations from different periods for the same firm and market are independent, we use standard errors that are robust to clustering within markets. Finally, even though we have panel data, in this version of the paper we do not exploit the panel structure of the data in estimation. All estimations use pooled data from the latest four periods. Data from the earliest (fifth) period is used only to calculate the distance regressors for the first period included in the regression. We pool the data for several reasons. First, note that any market or market-firm fixed effect (η_m or η_{im} from above) that affects identically the profits of all quality levels is already accounted for in the conditional logit formulation. More practically, adding alternative-specific market fixed effects would add about four thousand coefficients to the model, making estimation difficult and possibly leading to incidental parameter bias. Finally, most of the variation in the data occurs in the cross section, not the time series within each market, and so fixed effect modeling would reduce greatly the effective size of the sample.

V. Results

The conditional logit estimations return a large number of estimated coefficients, since each regressor not varying over alternatives has a different coefficient for each of the four alternatives apart from the baseline choice. In our main estimation, we have 84 coefficients. While exponentiated coefficients from a conditional logit estimation have meaning as odds ratios relative to the base alternative, it is often more natural for econometricians and policy analysts to think in terms of marginal effects. The marginal effect of a regressor is the impact of a one unit increase in the regressor on the probability that the firm chooses a particular quality alternative. When the regressor is log(*x*), 0.01 times the marginal effect measures the impact of a 1% increase in *x*. Given our interest in the top end of the quality ladder, we show marginal impacts on the top three quality categories.[26] In the tables, we present the marginal effects at the median (*MEMdn*) and average marginal effects (*AME*) for the regressors of interest instead of the coefficients.[27] With MEMdn, the marginal effect is calculated once, setting all covariates at their median values. With AME, the marginal effect is calculated for each observation in the sample using actual values of regressors, with the results then averaged over the sample. The coefficients for our main estimation are included in the appendix for the interested reader.

A. First-step results

In the first step of estimation, the quality choices of competitors are regressed on the market demographics and the infrastructure variables used to proxy the costs of the firm. Since the first step is akin to a reduced form forecasting exercise, we err on the side of including a large set of predictors without regard for causal meaning of the coefficients, parsimony, or the significance of the estimates. The demographics include those also included in the second step: market area in miles, population density and growth rate, median household income (averaged across Census block groups in the market) average age and age squared, average highest educational grade level achieved, the fraction of housing units that are rented or vacant, the fraction of the labor force working at home, the proportion of area employment that is in the financial, insurance, or real-estate (FIRE) sector, and the fraction of the market area that is under water. Where these variables are right skewed they are in logs, as noted in the tables. These variables were chosen based on a review of previous literature on the determinants

---

[26] We focus on the top end, in addition, because few ADSL offerings by ILECs are in the 0 or 1 categories anyway.

[27] Another reason not to present the coefficients is that the marginal effects are functions of all the coefficients, and thus it is possible for the ME of a regressor to be statistically significant even when its coefficient is not. Thus checking for significance stars on coefficients can give a misleading sense of which regressors are truly important.

broadband entry decisions. An additional set of demographics are included only in the first-step estimations: the density of households in the market, the fractions of nonwhite people, the percentage female, the standard deviation of education attainment, and the proportion of workers with long commutes. These variables are not included in the second step because of concerns about near multicollinearity with other demographic variables or insignificance and for the sake of parsimony in presenting results.

While we do not present the results from the first step in tables here, we note two things. The infrastructure variables are clearly highly relevant. The coefficients for *NearestSameSpeed* and *SameSpeednotFound* are statistically significant at the 1% level for all competing broadband types. The coefficients for *NearestAnySpeed* and *AnySpeednotFound* (where the latter variable is defined similarly to *SameSpeednotFound* but across all speed categories > 0) are also generally (but not uniformly) highly significant. The high significance and impact of these infrastructure variables implies that they are likely to be effective instruments to identify separately the impact of the competition variables in the second-step estimation. We also note that some of the demographic variables have insignificant coefficients, even those that we would expect to have strong impacts on quality choice. While that does not mean that they have no significant marginal effect on choice probabilities (see footnote 27), it does mean that the infrastructure variables alone are capturing much of the variation in quality choice. The same first-step estimates of Ew_{imt} are used for all the second-step specifications.

B. Second-step results

1. Estimation 1: Demographics only

We begin with a simple specification for ILEC ADSL quality choice in which only demographic variables are included. Table 3 contains the marginal effects, MEMdn and AME. The marginal effects are expressed in percentage points. Here we focus mainly on the marginal effects for the highest speed

ADSL, contained in the rightmost set of columns. This speed, from 10 to 25 Mbps, ("high-speed ADSL" in the following discussion) is offered in California by ILECs held by 10 holding companies, the largest of which is AT&T offering its U-verse service.[28] Another four firms offer ADSL only with lower speeds.[29] The results show that several of the demographic variables significantly[30] increase the probability of offering high-speed ADSL: income, population density, age (at the 10% level only), rental housing %, and FIRE employment. One variable, the vacancy rate, significantly lowers the firm's probability of offering high-speed ADSL. For an example of interpreting the numbers, consider the income variable. The MEMdn for log income, 16.68, implies that an increase of market-area household income of 10% increases the probability of the ILEC offering high-speed ADSL by 1.67 percentage points. The MEM for log income, 16.54, is similar in this case, although we observe that often the AME's are somewhat smaller than the MEMdn's.

2. Estimation 2: Add competition variables

Estimation 2 in Table 4 repeats the previous specification but with the competition variables included. In this estimation the infrastructure variables are still not included in the second step. Thus, the impacts of the competition variables may be biased due to endogeneity. For example, cost factors for firm *i* that are omitted in this regression, such as the presence of previously installed or nearby infrastructure, may be correlated with the quality choices of rivals through unobserved local factors. The estimates show apparently large impacts of competition on the ADSL quality choice. We consider

---

[28] The holding companies of these service providers are AT&T Inc., Calaveras Telephone Company, Frontier Communications Corporation, LICT Corporation, Ponderosa Communications, Inc., Sebastian Enterprises , Sierra Tel Communications Group, SureWest/Consolidated, Telephone and Data Systems, Inc., and Volcano Communications Company.

[29] The holding companies of these service providers are : Bryan Family Inc., Siskiyou Telephone Co., VARCOMM, Inc., and Verizon Communications. Verizon offers lower speed ADSL in some markets, but for higher qualities offers subscribers fiber (FIOS) instead.

[30] Here we mean "significant in either MEMdn or AME," and so below as well.

the impact of each type of competitor in turn. The marginal effect given for a speed category *j* is calculated to pertain to changing the competitors' maximum speed category from *j* – 1 to *j*.

The cable modem quality choice seems to affect the ILEC ADSL quality decisions a lot. When the cable modem service is relatively slow, the negative marginal effects for high-speed ADSL indicate that the ILECs are less likely to offer fast service themselves.[31] Once the cable companies move up into the DOCSIS 3.0 speed tiers, 50 Mbps and above, however, the ILEC is more likely to offer high-speed ADSL. The apparent effect of CLEC ADSL competition is similar: when the competitors' offerings are worse quality, the ILECs quality is less likely to be of the highest. The impact of CLEC gigabit fiber is small and insignificant. Given the omission of the infrastructure variables, we do not yet assign causal interpretation to these results. Comparing the results of the first two estimations, we see that the addition of the competition variables did not greatly change the marginal effects of the demographics. For this reason and to save space in the tables, we will not show the impacts of the demographic variables in the other tables below.

3. Estimation 3: The main specification, adding infrastructure variables

The addition of the infrastructure variables in the second-step estimation brings us to our preferred estimation (see Table 5). *SameSpeednotFound* and *NearestSameSpeed* have the expected negative impacts on same-choice alternatives. *NearestAnySpeed* has a further negative impact on the high-speed ADSL choice. After controlling for the same-speed infrastructure, the marginal effect for *NearestAnySpeed* can be interpreted as the impact of the distance to lower-speed infrastructure.

As expected from the discussion of the potential endogeneity problems in Estimation 2, the impacts of the competition variables, while qualitatively similar to before, have very different

---

[31] In this preliminary work, not all standard errors (s.e.'s) are available yet. Difficulties with numerical derivatives, and lack of time to program analytic derivatives yet, leads to the omitted SE's in this and following tables. Furthermore, as is common in the IO literature, the second-step s.e.'s do not account for estimation error in the first step. Thus our reported s.e.'s are smaller than those from the valid asymptotic variance-covariance matrix.

magnitudes in Estimation 3. The differing results show how the infrastructure variables help control for omitted variable bias. Since this is our main specification, we go through the results in greater detail here. When cable competitors switch from having no service to offering the slowest service (< 10 Mbps), the probability of high-speed ADSL rises by 12 percentage points. Looking at the columns in Table 5 for alternatives *j* = 2 and *j* = 3, we see that about two-thirds of these 12 percentage points come from upgrading from ADSL of speed between 3 and 6 Mbps, while about a quarter come from upgrading from ADSL of speed between 6 Mbps and 10 Mbps). Thus, whether an ILEC faces any cable competition at all appears to spur investment in ADSL speed. This impact (and those that follow) is not merely from the coincidence of DSL and cable modem service in more attractive markets, because the demographic regressors in the model control for the key market factors of income, population density, and so on. Furthermore, these apparent impacts are not merely reflections of favorable cost conditions for broadband provision, since last period's infrastructure variables account for that. Thus, they likely reflect the strategic considerations of ILECs in California and we interpret them as such.

However, as the cable competitors rise up the quality ladder, the impacts are not monotonic. When cable modem quality rises from between 10 and 25 Mbps to between 25 and 50 Mbps, the ILEC is 76 percentage points less likely to offer high-speed ADSL. Looking at the other columns of the table, we find that probability lost from alternatives 3 and 4 went to alternative 2. There are relatively few observations (88 out of 3,860 cases) with ILEC DSL entrants facing cable competition in speed category 3, and the large negative impact may merely be a small sample phenomenon, statistical significance notwithstanding. However, it may also be that the ILEC is responding with slower broadband to what the cable company did *not* do: upgrade to DOCSIS 3.0.[32] When the maximum speed of the cable

---

[32] The DOCSIS 2.0 standard can provide capacity up to 38 Mbps downstream and 27 Mbps upstream. DOCSIS 3.0 can dramatically increase downstream and upstream capacity by a factor of the number of channels used in the network (up to four channels), but this upgrade may take up to two years to complete. This means that DOCSIS 2.0

modem service rises to the two highest speed categories, the marginal effects on high-speed ADSL are positive. The largest marginal effects are for when the cable competitors upgrade from 25-50 Mbps (mostly DOCSIS 2.0) to 50-100 Mbps (DOCSIS 3.0). The MEMdn for high-speed ADSL is 88.3 percentage points, and the AME is 57.7. The probability gained comes mainly from alternative 2. In summary, ILECs generally respond to cable competition by upgrading their ADSL quality when they face any competition at all and when the quality of the competition becomes high.

In contrast with cable modem competition, there is no strongly significant evidence that ILECs pay much attention to the quality of their CLEC ADSL competition. The results are also in contrast to the previous estimation, in which CLEC ADSL had some highly significant marginal effects on high-speed ILEC ADSL. This difference shows the importance of the infrastructure variables in controlling for omitted variable bias. The largest impact on high-speed ILEC ADSL, 10 percentage points for the MEMdn and 8.8 for the AME, comes from when the CLECs move into the same speed category (between 10 and 25 Mbps). However, the MEMdn is not significant at the 5% level. The generally weak response to CLEC ADSL quality may indicate that ILECs do not perceive CLECs in California to be much of a competitive threat to their largely residential-oriented ADSL service. The largest CLEC ADSL provider, by far, is MegaPath (held by Platinum Equity, Inc.), which targets the business market.

Finally, the presence of gigabit CLEC fiber spurs a 6.2 percentage point increase (per MEMdn; 7.6 for AME) in high-speed ADSL. While Google fiber does not appear in our data, these results are in line with anecdotal accounts of incumbent broadband providers increasing their quality of service in response to Google fiber elsewhere in the country.[33]

---

could be used to move from the 10 to 25 Mbps range to the 25 to 50 Mbps range, whereas moving to the 50 to 100 Mbps range requires DOCSIS 3.0.

[33] For example, in the Austin area where Google Fiber is available, AT&T's U-verse DSL service has offered downstream speeds up to 300 Mbps. See http://www.fiercetelecom.com/story/att-begins-upgrading-austin-customers-1-gbps-service/2014-08-11 (accessed on 8/13/2014).

4. Additional estimations and robustness checks

Here we briefly consider three additional estimations performed as robustness checks. A subset of the results is in Table 6, where only the marginal effects for high-speed ADSL are shown. In Estimation 4, we use the log of the distances for the infrastructure variables instead of the level. The results are qualitatively similar to Estimation 3, except that the marginal effects of the slowest cable modem service and CLEC gigabit fiber lose significance. In Estimation 5, we replace the cable modem competition variables with their market-coverage-weighted counterparts. The results are very close to those of Estimation 3, even for the cable competition variables. Finally, we include all available demographics in Estimation 6. Again, the results are similar, except that the impact of CLEC gigabit fiber loses significance.

VI. Discussion and Conclusions

A. General

Our results show that ILECs respond to the quality choices of rival broadband providers. Their responses are heterogeneous to type of provider and to the level of quality. Specifically, ILECs appear to care more about rivals using cable modem or fiber technologies than rivals using a similar ADSL technology. The likely expectation for the ILEC is that if a consumer is going to switch services they are more likely to switch to a rival with a technology that can provide a very fast Internet service. Moreover, the level of speed matters in strategic responses. Particularly, when cable modem rivals move from no service to a “low” speed service tier or from a “medium” service tier to a “high” speed service tier, the ILEC also increases speed. This suggests strategic complementarity in the provision of quality and is consistent with the findings from Kugler and Weiss (2013) and Brueckner and Luo (2013) in the reaction-function literature as well as with Goolsbee and Petrin (2004), Savage and Wirth (2005) and Matsa (2011) in the reduced-form competition and quality literature.

Interestingly, however, when cable modem rivals move from “low” to “medium” speed service tiers, the ILEC is less likely to provide high quality ADSL, which suggests strategic substitutability in the provision of quality. This response may be due to changes in the price and quality elasticity of demands as suggested by theory. However, it is possible that this reflects the ILEC’s reaction to an underlying capacity constraint facing its rival, which does not have DOCSIS 3.0 installed. The result may even merely be an artifact of the data that will not persist once we expand our analysis. More empirical and theoretical analysis is required to fully understand this intriguing result. Overall, our empirical finding of a non-monotonic relationship between the quality choices of rival broadband providers also resembles the findings of Chen and Gayle (2013) from the airline industry.

B. Further work

The empirical results presented here are preliminary, and in this final section we discuss how we will refine them and which other questions we will address with the data. The next step in the project is to add additional competitors to the estimation of ILECs’ quality choice for ADSL. Besides the CLECs and cable modem providers, we have data on fixed and mobile wireless providers and other types of wireline providers (symmetric DSL, “other copper”, and so forth). In this initial work we focused only on the main competitors, but these other types of broadband may also affect the ILECs’ decisions. We also plan to explore the strategic quality choice of other broadband players besides ILECs offering ADSL. That is, even those we model the CLEC and cable modem providers in the first step of estimation, we have not yet included them in the second step. It may also be interesting to use the count of competitors in each quality category instead of merely an indicator for at least one competitors’ presence there. This is not likely to be important for cable companies, since overbuilding is rare in California. However, it may be more important for CLEC competition. We also have not yet re-estimated the model using the alternative 18 kf market definition, which we can use as a further robustness check.

Once we have a more mature set of estimation results in hand, we plan to exploit the model to address important questions of public policy. For example, our results so far suggest that intermodal competition is much more important than intramodal competition among ILECs and CLECs. If this result persists as we refine our results, it may indicate that unbundling policies aimed at the local loop would be largely ineffective at increasing the quality of service offerings to residents. Our model will allow us to address this question quantitatively. As a second example, we will also be able to address the impact of a merger between cable companies or fiber broadband providers on the quality offered by those firms and their competitors in the market. Such analyses will be useful for purposes of merger review.

References

Abbott, L. (1955). *Quality and Competition; an Essay in Economic Theory*. Columbia University Press.

Aghion, P., & Howitt, P. (2009). *The Economics of Growth*. MIT Press.

Aguirregabiria, V., & Mira, P. (2007). Sequential estimation of dynamic discrete games. *Econometrica, 75*(1), 1–53.

Bajari, P., Han, H., Krainer, J., & Nekipelov, D. (2010). Estimating Static Models of Strategic Interactions. *Journal of Business & Economic Statistics*, *28*(4), 469-482.

Berry, S. (1992), Estimation of a Model of Entry in the Airline Industry. *Econometrica*, *60*(4), 889–917.

Bresnahan, T. F. & Reiss, P. C. (1990). Entry in Monopoly Markets. *Review of Economic Studies, 57*, 531-553.

Brueckner, J., & Luo, D. (2013). Measuring Strategic Firm Interaction in Product-Quality Choices: The Case of Airline Flight Frequency. Work-in-Progress, Department of Economics, University of California, Irvine.

California Public Utility Commission (CPUC). (2013). *Data Processing Methods*. October 1st. Retrieved from http://www.cpuc.ca.gov/NR/rdonlyres/73AE2E37-3E8C-45E8-8B13-97C3B8434D86/0/CA_Methodology_2013_10_01.pdf

Chamberlin, E. (1933). *The Theory of Monopolistic Competition*. Harvard University Press.

Chen, Y., & Gayle, P. (2013). Mergers and Product Quality: Evidence from the Airline Industry. MPRA Paper No. 51238, http://ideas.repec.org/p/pra/mprapa/51238.html.

Chen, Y., & Schwartz, M. (2013). Product Innovation Incentives: Monopoly vs. Competition. *Journal of Economics and Management Strategy*, *22*(3), 513-528.

Ciliberto & Tamer, E. (2009). Market Structure and Multiple Equilibria. *Econometrica*, *77*(6), 1791-1828.

Crawford, G. (2007). Television Station Ownership Structure and the Quantity and Quality of TV Programming. FCC Media Ownership Study #3, Washington, D.C., FCC. Retrieved from https://apps.fcc.gov/edocs_public/attachmatch/DA-07-3470A4.pdf

Denni, M., & Gruber, H. (2007). The diffusion of broadband telecommunications in the U.S.—the role of different forms of competition. *Communications & Strategies*, *68*, 139-157.

Dick, A. (2007). Market Size, Service Quality, and Competition in Banking. *Journal of Money, Credit and Banking*, *39*(1), 49-81.

Domberger, S., & Sherr, A. (1989). The Impact of Competition on Pricing and Quality of Legal Services. *International Review of Law and Economics*, *9*, 41-56.

Dorfman, R., & Steiner, P. (1954). Optimal Advertising and Optimal Quality. *American Economic Review*, *44*, 826-836.

Douglas, G., & Miller, J. (1974). Quality Competition, Industry Equilibrium, and Efficiency in the Price-Constrained Airline Market. *American Economic Review*, *64*, 657-669.

Flamm, K. (2005). The Role of Economics, Demographics, and State Policy in Broadband Availability. Paper presented at the PURC/London Business School Conference on “The Future of Broadband: Wired and Wireless, 2005”, Gainesville, Florida.

Flamm, K., & Chaudhuri, A. (2007). An analysis of the determinants of broadband access. *Telecommunications Policy*, *31*, 312–326.

Gaynor, M. (2006). What do We Know About Competition and Quality in Health Care Markets? NBER Working Paper 12301, Cambridge, MA.

GeoLytics, Inc. (2012). *Census 2010 and ACS 2008-2012* [Custom Data Files]. Available from http://www.geolytics.com

Goolsbee, A., & Petrin, A. (2004). The Consumer Gains from Direct Broadcast Satellites and the Competition with Cable TV. *Econometrica*, *72*, 351-81.

Greenstein, S., & Mazzeo, M. (2006). The Role of Differentiation Strategy in Local Telecommunication Entry and Market Evolution: 1999-2002. *The Journal of Industrial Economics*, *54*(3), 323-350.

Grubesic, T. H., & Murray, A. T. (2004). Waiting for broadband: Local competition and the spatial distribution of advanced telecommunication services in the United States. *Growth and Change*, *35*(2), 139-165.

Hiller, S., Savage, S., & Waldman, D. (2014). Market Structure and Media Diversity. forthcoming in *Economic Inquiry*.

Katz, M. (2013). Provider Competition and Healthcare Quality: More Bang for the Buck? *International Journal of Industrial Organization*, *31*, 612-625L

Kranton, R. (2003). Competition and the Incentive to Produce High Quality. *Economica*, *70*, 385-404.

Kugler, A., & Weiss, C. (2013). Time as a Strategic Variable: An Empirical Analysis of Opening Hours in the Austrian Gasoline Market. Work-in-Progress, Vienna University of Business and Economics, Vienna.

Lee, S. [Sangwon] , Marcu, M., & Lee, S. [Seonmi]. (2011). An empirical analysis of fixed and mobile broadband diffusion. *Information Economics and Policy*, *23*, 227-233.

Matsa, D. (2011). Competition and Product Quality in the Supermarket Industry. *Quarterly Journal of Economics*, *126*, 1539-1591.

Mazzeo, M. (2002). Product Choice and Oligopoly Market Structure. *Rand Journal of Economics 33*(2), 221-242.

Mazzeo, M. (2003). Competition and Service Quality in the U.S. Airline Industry. *Review of Industrial Organization*, *22*(4), 275-296.

Molnar, G., *&* Savage, S. (2013). Market Structure and Product Quality: Evidence from United States Broadband Internet Markets. Work-in-Progress, University of Colorado, Boulder.

Nardotto, M., Valletti, T., & Verboven, F. (2012). Unbundling the Incumbent: Evidence From UK Broadband. CEPR Discussion Papers 9194.

Olivares, M., & Cachon, G. (2009). Competing Retailers and Inventory: An Empirical Investigation of General Motors’ Dealerships in Isolated U.S. Markets. *Management Science*, *55*(9), 1586–1604.

Nevo, A., Turner, J., & Williams, J. (2013). Usage-Based Pricing and Demand for Residential Broadband. Retrieved from: http://dx.doi.org/10.2139/ssrn.2330426

Prieger, J. E. (2003). The supply side of the Digital Divide: Is there equal availability in the broadband Internet access market? *Economic Inquiry*, *41*(2), 346-363.

Prieger, J. E. (2007). The Impact of Cost Changes on Industry Entry and Exit, *Journal of Economics*, *91*(3), 211–243.

Prieger, J. E. (2012). The Broadband Digital Divide and the Economic Benefits of Mobile Broadband for Rural Areas. *Telecommunications Policy*, *37*(6-7), 483-502.

Prieger, J. E. & Church, T. (2012). Deployment of Mobile Broadband Service in the United States. In *Mobile Services Industries, Technologies, and Applications in the Global Economy*, I. Lee (Ed.), Hershey, PA: IGI Global.

Prieger, J. E. & Hu, W.-M. (2008a). Competition in broadband provision and the digital divide. In Y.K. Dwivedi, et al. (Eds.), *Handbook of Research on Global Diffusion of Broadband Data Transmission*, (pp. 241-259). Hershey, PA: IGI Global.

Prieger, J. E., & Hu, W.-M. (2008b). The broadband digital divide and the nexus of race, competition, and quality. *Information Economics and Policy*. *20*(2), 150-167.

Prieger, J. E. & Lee, S. (2008). Regulation and the Deployment of Broadband. In Y.K. Dwivedi, et al. (Eds.), *Handbook of Research on Global Diffusion of Broadband Data Transmission*, (pp. 278-303). Hershey, PA: IGI Global.

Prince, J. E., & Simon, D. (2013). Do Incumbents Improve Service Quality in Response to Entry? Evidence from Airlines' On-Time Performance. Indiana University School of Public and Environmental Affairs Research Paper No. 2010-09-01. Retrieved from http://ssrn.com/abstract=1670203 or http://dx.doi.org/10.2139/ssrn.1670203

Savage, S., & Wirth, M. (2005). Price, Programming, and Potential Competition in U.S. Cable Television Markets. *Journal of Regulatory Economics*, *27*, 25-46.

Schmalensee, R. (1977). Comparative Static Properties of Regulated Airline Oligopolies. *Bell Journal of Economics*, *8*, 565-576.

Schmalensee, R. (1979). Market Structure, Durability, and Quality: A Selective Survey. *Economic Inquiry*, *17*, 177-196.

Seim, K. (2005). An Empirical Model of Firm Entry with Endogenous Product-Type Choices. *Rand Journal of Economics 37*(3), 619-640.

Stenberg, P., Morehart, M., Vogel, S., Cromartie, J., Breneman, V., & Brown, D. (2009). Broadband Internet's value for rural America. U.S. Department of Agriculture, Economic Research Service, Economic Research Report No. 78.

Swan, P. (1970). Durability of Consumer Goods. *American Economic Review*, *60*, 884-894.

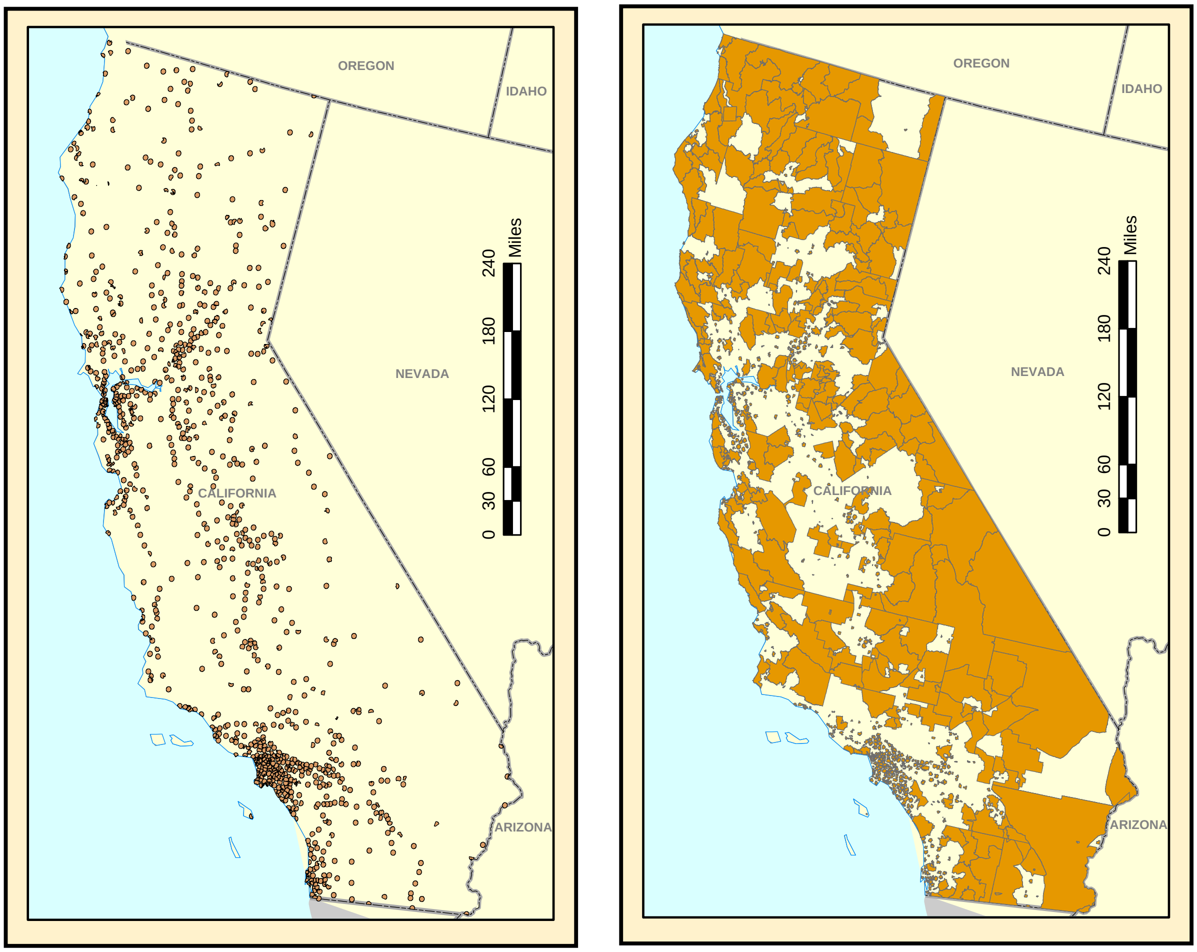
Swan, P. (1971). The Durability of Goods and the Regulation of Monopoly. *Bell Journal of Economics and Management Science*, *2*, 347-357.

Tamer, E. (2003). Incomplete simultaneous discrete response model with multiple equilibria. *Review of Economic Studies 70*(1), 147–167.

Wallsten, S., & Mallahan, C. (2013). Residential Broadband Competition in the United States. in Shane Greenstein, Avi Goldfarb, and Catherine Tucker (Eds.), *The Economics of Digitalization*, Elgar Publishing.

Xiao, M., & Orazem, P. (2011). Does the Fourth Entrant Make Any Difference?: Entry and Competition in the Early U.S. Broadband Market. *International Journal of Industrial Organization*, *29*(5), 547-561.

Figure 1: The Market Areas in California



Step-Two Market Areas

Step-Three Market Areas

Figure 2: Example of Urban Market Areas: Los Angeles Area

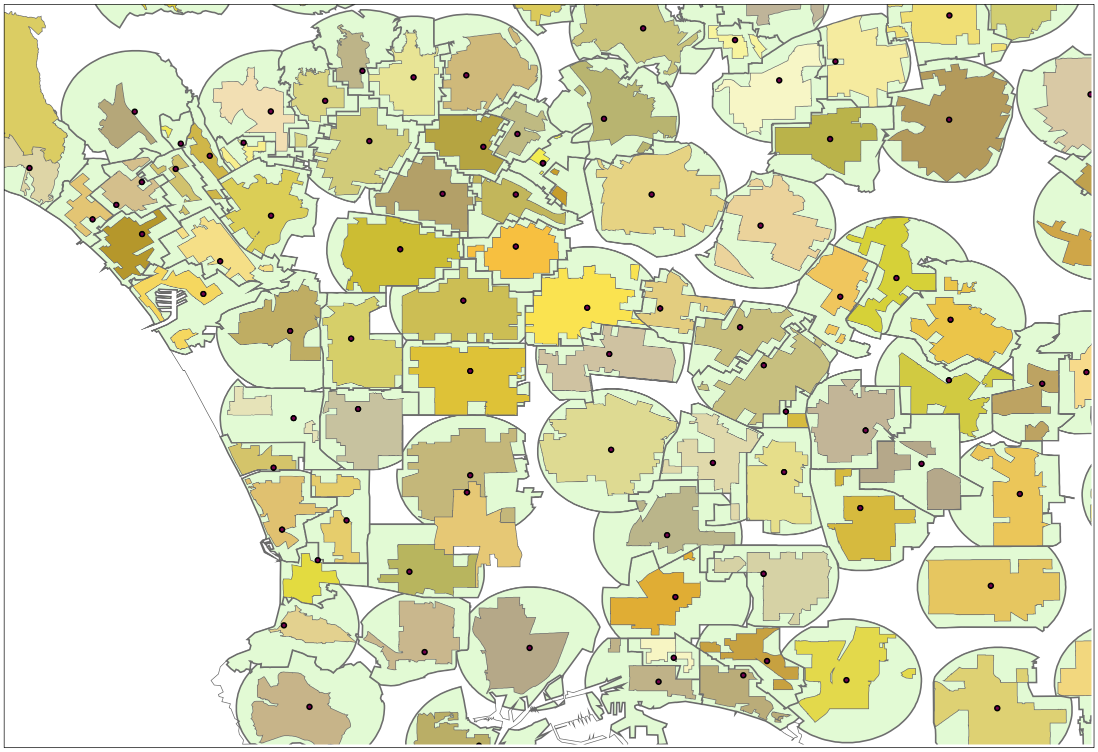


Figure 3: Example of Rural Market Areas: Rural Fresno County

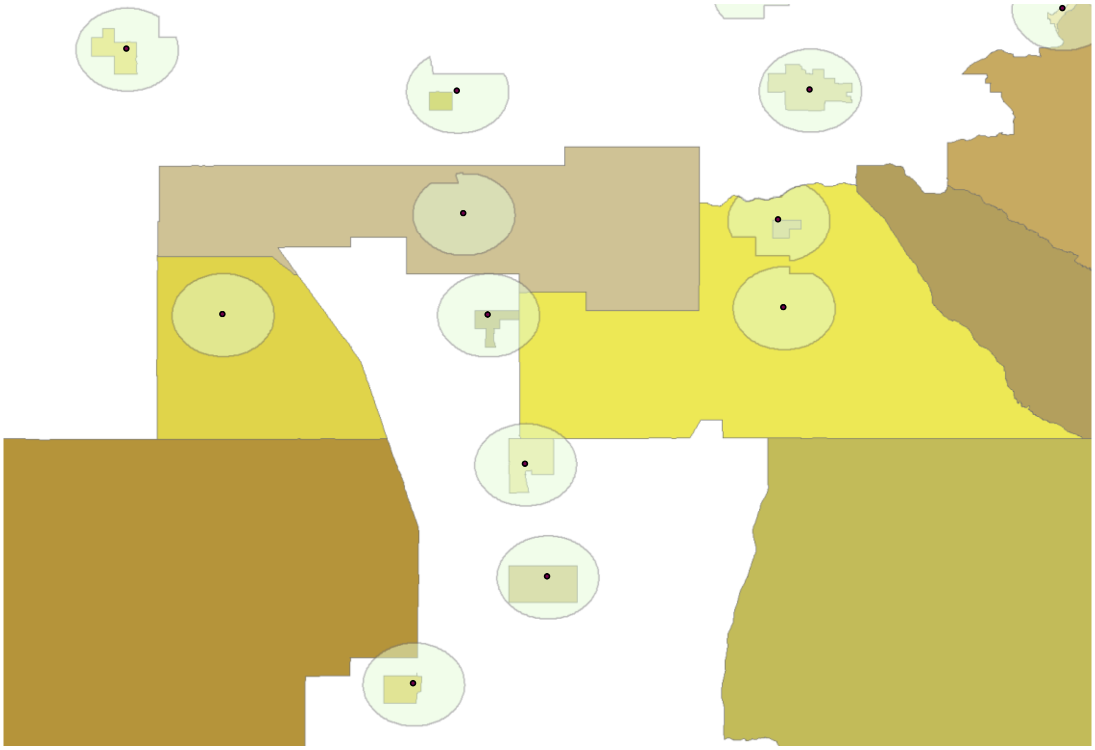
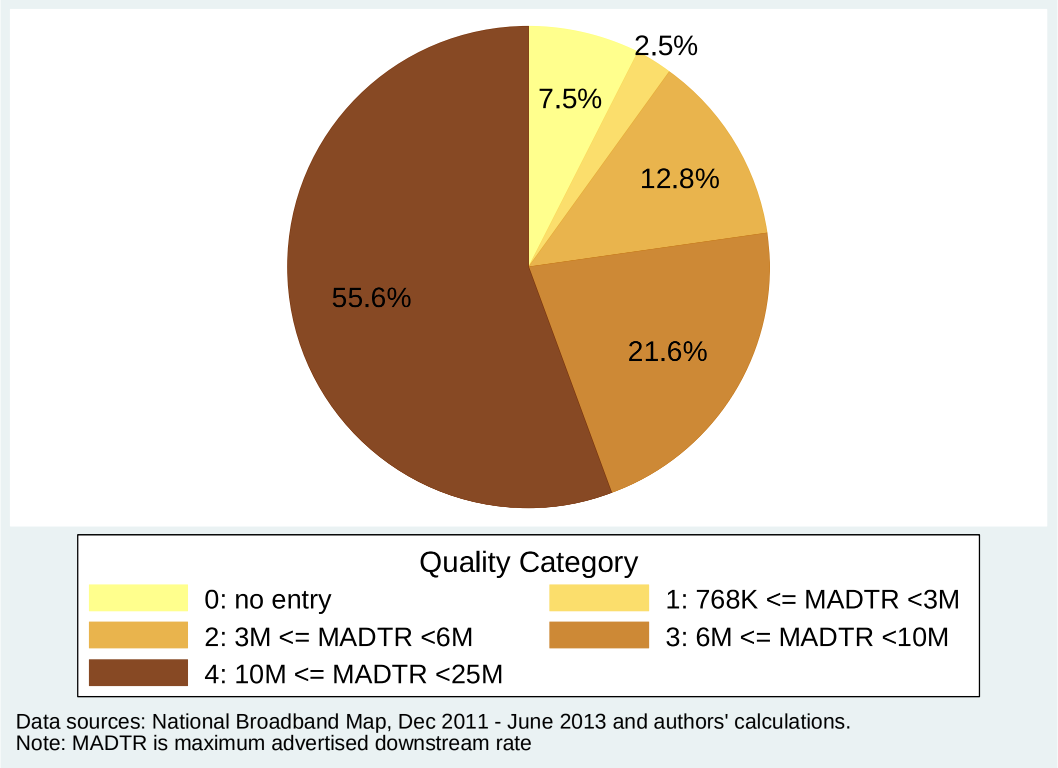


Figure 4: ILEC ADSL Quality Choice in California



Data sources: National Broadband Map, Dec 2011 - June 2013 and authors' calculations.
Note: MADTR is maximum advertised downstream rate

Table 1: Summary Statistics

	Mean	s.d.	Min	Max
Y (chosen alternatives)	3.154	1.197	0	4
Competition variables†				
Cable modem service				
CM <10 M	0.853	0.355	0	1
10M < CM < 25M	0.826	0.379	0	1
25M < CM < 50M	0.761	0.427	0	1
50M < CM <100 M	0.738	0.440	0	1
100M < CM	0.471	0.499	0	1
CLEC ADSL				
768K < CLEC ADSL <3M	0.634	0.482	0	1
3M < CLEC ADSL <6M	0.509	0.500	0	1
6M < CLEC ADSL <10M	0.458	0.498	0	1
10M < CLEC ADSL <25M	0.384	0.486	0	1
25M < CLEC ADSL <50M	0.109	0.312	0	1
CLEC fiber > 1G	0.223	0.417	0	1
Main Demographics				
Area (log mi)	2.473	0.460	-1.150	2.986
Pop. density (log)	6.479	3.082	-2.387	11.327
Pop. growth	0.110	0.801	-0.916	22.752
Age	37.854	6.136	20.093	62.661
Education (grade)	13.694	1.690	7.895	17.445
Rental housing (%)	0.413	0.177	0.010	1.000
Work at home	0.204	0.070	0.000	0.638
Water area (%)	0.013	0.035	0.000	0.493
Income (log)	11.090	0.399	9.322	12.544
Vacancy rate (log)	-2.339	0.786	-4.615	-0.109
FIRE employment (log %)	-2.831	0.272	-3.597	-2.054
Additional Demographics				
HH density (log)	5.424	3.044	-3.294	10.205
Nonwhite %	0.270	0.174	0.000	0.884
Female %	0.493	0.051	0.038	0.631
Education, s.d.	514.379	342.021	63.649	2262.061
Long commute %	0.199	0.106	0.000	0.868

†The competition variables are defined to represent all possible speeds that competitors could offer in the market. When the maximum speed in the NBM of competitors is in a particular category, then the indicator variables equal 1 for that and lower speed categories.

Table 2: Quality Alternatives – The Downstream Speed Categories

Alternative	**Speed Category**		
	ILEC ADSL	**CLEC ADSL**	**Cable Modem**
0	No entry	No entry	No entry
1	768K ≤ MADTR <3M	768K ≤ MADTR <3M	MADTR <10 M
2	3M ≤ MADTR <6M	3M ≤ MADTR <6M	10M ≤ MADTR < 25M
3	6M ≤ MADTR <10M	6M ≤ MADTR <10M	25M ≤ MADTR < 50M
4	10M ≤ MADTR <25M	10M ≤ MADTR <25M	50M ≤ MADTR < 100 M
5	NA	25M ≤ MADTR <50M	100M ≤ MADTR

Table notes: *MADTR* is the maximum advertised downstream transmission rate.

Table 3: ILEC ADSL Estimation 1 – Demographics only

	Prob(3M < ADSL speed <6M)				Prob(6M < ADSL speed <10M)				Prob(10M < ADSL speed <25M)			
	ME at median		Average ME		ME at median		Average ME		ME at median		Average ME	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE
Demographics												
Area (log mi)	-1.022	1.953	-1.294	2.037	-9.903***	3.238	-8.533***	3.044	9.883**	3.897	7.194*	3.674
Pop. density (log)	-0.447	0.435	0.145	0.443	0.937	0.815	1.551**	0.643	0.284	0.821	1.355**	0.635
Pop. growth	0.600	1.100	0.470	1.190	-7.035	4.691	-6.111	4.022	6.104	4.086	4.408	3.001
Age	-1.645	1.046	-1.475	1.101	-1.498	2.023	-0.921	1.753	3.246	2.045	3.015*	1.711
Age squared	0.021*	0.012	0.019	0.013	0.008	0.025	0.003	0.021	-0.029	0.025	-0.028	0.021
Education (grade)	-0.068	0.751	0.110	0.768	1.475	1.358	1.464	1.160	-1.142	1.349	-0.549	1.136
Rental housing (%)	-14.497**	7.313	-13.237*	7.016	-37.210***	12.456	-28.700***	10.136	53.210***	12.739	45.438***	10.415
Work at home	-8.695	11.971	-10.910	12.095	-31.780	22.402	-28.841	18.529	37.473	23.134	26.151	19.458
Water area (%)	32.732*	17.782	39.309**	18.081	-87.169*	48.996	-67.503	41.034	61.927	44.643	55.993	35.660
Income (log)	-6.483*	3.761	-4.939	3.656	-8.468	6.624	-4.726	5.486	16.684**	6.867	16.539***	5.747
Vacancy rate (log)	2.466	1.841	2.258	1.790	10.395***	3.244	8.296***	2.622	-12.939***	3.247	-10.712***	2.682
FIRE employment (log %)	-5.347*	2.828	-4.563	2.774	-22.731***	6.094	-17.710***	4.899	28.384***	6.190	24.037***	4.945

*10% sig level ** 5% sig level ***1% sig level.

Table notes: Estimation method is Conditional Logit. Since there are no competition variables, the two step estimation described in the text (refer to section IV.B in the text) is not required in this specification. SE's are robust to clustering on markets. "ME at median" is the marginal effect of a one unit increase in the regressor in the row label on the choice probability given in the column superheading, calculated at the median values of all regressors. "Average ME" is the marginal effect of a one unit increase in the regressor in the row label on the choice probability given in the column superheading, calculated at actual regressor values and averaged over the sample. Choice alternatives not shown in the table but included in the estimation are 1) no ADSL broadband, and 2) Prob(768K < ADSL speed <3M).

Table 4: ILEC ADSL Estimation 2 – Competition variables and demographics

	Alternative 2				Alternative 3				Alternative 4			
	Prob(3M < ADSL speed <6M)				Prob(6M < ADSL speed <10M)				Prob(10M < ADSL speed <25M)			
	MEMdn		AME		MEMdn		AME		MEMdn		AME	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE
Competition variables†												
CM <10 M	-9.846	7.332	-5.055		39.410***	14.640	32.850		-26.950**	12.110	-19.072	
10M < CM < 25M	43.690***	12.730	33.305		-31.520**	15.870	-22.652		-11.910	11.700	-12.425	
25M < CM < 50M	40.230***	11.610	43.626		-28.470***	10.240	-27.666		-11.770*	6.197	-15.258	
50M < CM <100 M	-76.51***	5.089	-71.584		31.460***	5.296	29.042		45.820***	5.658	46.794	
100M < CM	-11.15***	3.926	-8.644		-9.640*	4.925	-8.585		20.520***	4.956	15.782	
768K < CLEC ADSL <3M	1.388	5.856	3.524		-2.155	6.883	0.319		2.129	7.114	5.269	
3M < CLEC ADSL <6M	-5.243	6.984	-5.304		2.416	14.730	1.387		2.220	14.070	0.391	
6M < CLEC ADSL <10M	-8.570	5.639	-8.353		40.230***	15.550	26.689		-33.750**	14.270	-30.193	
10M < CLEC ADSL <25M	0.869	2.938	0.225		-23.050***	8.682	-17.865		22.030***	7.460	17.120	
25M < CLEC ADSL <50M	-1.176	3.037	-0.088		-39.920***	5.804	-28.432		42.920***	5.925	34.986	
CLEC fiber > 1G	-13.11***	3.540	-9.706		11.280	8.536	12.109		2.555	8.439	8.028	
Demographics												
Area (log mi)	-2.676	2.545	-1.680	1.821	-7.311**	3.703	-4.717*	2.697	10.048**	4.471	7.659**	3.446
Pop. density (log)	0.116	0.664	0.598	0.467	1.061	0.950	1.344**	0.638	-0.971	0.908	0.205	0.627
Pop. growth	2.463	2.027	1.299	1.515	-13.576**	5.813	-9.899**	4.128	10.956**	4.957	7.180**	3.140
Age	-0.010	1.581	-0.020	1.114	-2.889	2.288	-1.975	1.709	2.875	2.272	2.051	1.673
Age squared	0.005	0.018	0.004	0.013	0.025	0.027	0.017	0.020	-0.030	0.027	-0.020	0.020
Education (grade)	-0.286	1.048	0.067	0.753	0.617	1.689	0.740	1.225	-0.218	1.632	0.321	1.185
Rental housing (%)	3.607	9.950	2.297	7.010	-53.267***	15.314	-37.238***	10.400	49.788***	14.795	35.859***	10.390
Work at home	-7.971	15.331	-8.766	11.236	-37.606	24.265	-29.685*	16.733	44.447*	23.856	26.861	17.726
Water area (%)	38.599*	23.005	35.697**	16.719	-83.318	52.120	-48.197	35.932	48.603	45.621	50.388	32.816
Income (log)	3.346	4.424	2.934	3.183	-5.649	8.181	-3.133	5.828	2.465	7.845	2.732	5.728

Table continued next page

	Alternative 2				Alternative 3				Alternative 4			
	Prob(3M < ADSL speed <6M)				Prob(6M < ADSL speed <10M)				Prob(10M < ADSL speed <25M)			
	MEMdn		AME		MEMdn		AME		MEMdn		AME	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE
Vacancy rate (log)	-0.668	2.302	-0.650	1.612	13.733***	3.784	9.300***	2.498	-13.188***	3.554	-9.925***	2.491
FIRE employment (log %)	-0.588	4.511	-0.284	3.168	-38.049***	8.272	-26.087***	5.267	38.768***	7.835	28.438***	5.356

*10% sig level ** 5% sig level ***1% sig level.

†For the competition variables, the marginal effect is for moving from the speed category below to the category in the row label. E.g., for the row labeled “10M < CM < 25M” the ME is for moving from facing CM competitors with a max speed of less than 10M (“CM <10 M”) to facing CM competitors with a max speed of between 10M and 25M.

Table notes: Estimates are from the second step, and all the regressors labeled “competition variables” are the expected values calculated from the first step (refer to section IV.B in the text). See also notes to previous table.

Table 5: ILEC ADSL Estimation 3 (Main estimation) – Demographics, Competition variables, and Nearest Infrastructure variables

	Alternative 2			Alternative 3			Alternative 4		
	Prob(3M < ADSL speed <6M)			Prob(6M < ADSL speed <10M)			Prob(10M < ADSL speed <25M)		
	MEMdn		AME	MEMdn		AME	MEMdn		AME
	Est.	SE	Est.	Est.	SE	Est.	Est.	SE	Est.
Competition variables									
CM <10 M	-8.215***	2.050	-14.204	-2.950	5.733	9.159	12.190**	6.210	13.605
10M < CM < 25M	11.000*	6.371	17.335	5.262	8.081	-7.888	-16.960	11.820	-15.807
25M < CM < 50M	88.550***	6.431	77.247	-11.400*	6.359	-18.942	-76.460***	10.870	-46.781
50M < CM <100 M	-96.230***	1.008	-84.814	7.890***	1.793	22.649	88.310***	2.359	56.720
100M < CM	-1.865**	0.906	-3.333	-2.413	1.645	0.029	4.285**	2.160	3.874
768K < CLEC ADSL <3M	-2.262	1.529	-0.276	-2.829	2.837	0.784	5.422*	3.277	5.784
3M < CLEC ADSL <6M	-0.466	1.777	-1.912	-0.822	3.334	-1.535	1.198	3.803	0.234
6M < CLEC ADSL <10M	-1.228	1.681	-8.962	10.180*	5.387	9.726	-9.436	6.110	-6.464
10M < CLEC ADSL <25M	-0.629	0.789	1.546	-8.855*	4.568	-3.453	10.030*	5.207	8.767
25M < CLEC ADSL <50M	3.318**	1.571	12.446	-5.660***	1.485	-22.302	1.310	2.409	-3.930
CLEC fiber > 1G	-3.101***	0.822	-9.931	-3.069*	1.774	3.048	6.212***	2.033	7.570
Demographics									
Included but not shown									
Cost/Infrastructure									
SameSpeednotFound (own effect)	-3.142***	0.657	NA	-7.237***	1.293	NA	-88.510***	1.586	NA
NearestSameSpeed (log mi, own effect)	-0.100***	0.018	NA	-0.221***	0.043	NA	-0.307***	0.048	NA
NearestAnySpeed (log mi)	0.313**	0.152	0.324	0.527	0.321	0.036	-0.853*	0.436	-0.785

*10% sig level ** 5% sig level ***1% sig level.

Table notes: Estimates are from the second step. See also notes to previous table.

Table 6: Additional ILEC ADSL Estimations (Robustness Checks) – Results for the Choice of High Speed ILEC ADSL, Prob(10M < ADSL speed <25M)

	Estimation 4 **Log distance variables**			**Estimation 5** **CM coverage-adjusted**			**Estimation 6** **Expanded set of demographics**		
	MEMdn		AME	MEMdn		AME	MEMdn		AME
	Est.	SE	Est.	Est.	SE	Est.	Est.	SE	Est.
Competition variables									
CM <10 M	4.910	3.128	11.231	11.000**	5.141	13.161	11.740*	7.002	12.096
10M < CM < 25M	-4.850	5.424	-9.504	-19.310	12.740	-17.818	-20.680	13.420	-16.731
25M < CM < 50M	-91.610***	5.722	-46.370	-75.860***	12.170	-46.378	-72.210***	12.130	-45.683
50M < CM <100 M	91.710***	2.702	47.058	88.880***	2.330	56.697	86.870***	2.676	56.384
100M < CM	4.659***	1.788	7.546	3.925*	2.183	3.722	4.810**	2.446	4.166
768K < CLEC ADSL <3M	1.081	2.527	3.047	6.265*	3.276	6.507	1.618	3.382	2.540
3M < CLEC ADSL <6M	2.997	2.431	-0.033	1.524	3.820	0.337	1.257	4.400	-0.084
6M < CLEC ADSL <10M	-5.894	3.731	-3.093	-9.273	6.161	-6.138	-7.285	6.234	-3.641
10M < CLEC ADSL <25M	2.784	3.559	0.996	9.293*	5.228	7.787	6.009	5.192	4.600
25M < CLEC ADSL <50M	-0.760	3.418	-4.793	1.613	2.502	-2.887	2.285	3.955	-3.664
CLEC fiber > 1G	0.475	2.050	3.022	6.791***	2.028	8.025	2.673	2.916	2.594
Demographics									
Included but not shown									
Cost/Infrastructure									
SameSpeednotFound[†]	-90.240***	1.541		-88.310***	1.605		-87.410***	1.811	
NearestSameSpeed (log mi)[†]	-4.660***	0.860		-0.313***	0.047		-0.330***	0.054	
NearestAnySpeed (log mi)	-2.366***	0.553	-4.071	-0.886**	0.448	-0.798	-0.924*	0.487	-0.777

*10% sig level ** 5% sig level ***1% sig level. [†] Own effect (effect on alternative 4, in this case).

Table Notes: Each set of three columns pertains to a different estimation specification. 1st: distances in the infrastructure variables are in logs instead of levels. 2nd: cable modem competition variables are weighted by fraction of market covered by the cable service area. 3rd: all available demographics are included.

Appendix: Raw Estimation Output for Estimation 3 (Main Specification)

This appendix contains the estimated coefficients for Estimation 3, our preferred specification. These are the coefficients underlying the marginal effects reported in Table 5.

```
Alternative-specific conditional logit         Number of obs      =     19280
Case variable: caseID                          Number of cases    =      3856

Alternative variable: sbin_option              Alts per case: min =         5
                                                              max =         5

                                                  Wald chi2(98)   =   1097.19
Log pseudolikelihood = -2032.1477                 Prob > chi2     =    0.0000

                             (Std. Err. adjusted for 965 clusters in market)
------------------------------------------------------------------------------
                     |               Robust
              choice |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------------------+--------------------------------------------------------
sbin_option          |
      SameBBnotFound |  -6.705426   .3666517   -18.29   0.000     -7.42405   -5.986802
nearestSameBBifFound |  -.0328376   .0046148    -7.12   0.000    -.0418824   -.0237928
---------------------+--------------------------------------------------------
0                    |  (base alternative)
---------------------+--------------------------------------------------------
1                    |
   nearestSameTechBB |   -.151646   .0288382    -5.26   0.000    -.2081678   -.0951243
Expected_CM_Quality1 |  -1.606585   2.428441    -0.66   0.508    -6.366242    3.153072
Expected_CM_Quality2 |  -2.245731   3.062607    -0.73   0.463     -8.24833    3.756869
Expected_CM_Quality3 |   9.618941   4.032032     2.39   0.017     1.716304    17.52158
Expected_CM_Quality4 |  -7.919902   3.662783    -2.16   0.031    -15.09882   -.7409792
Expected_CM_Quality5 |   2.729569   1.313439     2.08   0.038     .1552752    5.303863
 Exp_ADSL_CLEC_Qlty1 |  -2.027241   2.627857    -0.77   0.440    -7.177747    3.123265
 Exp_ADSL_CLEC_Qlty2 |   2.740354   3.469344     0.79   0.430    -4.059436    9.540143
 Exp_ADSL_CLEC_Qlty3 |  -3.455721   2.373179    -1.46   0.145    -8.107067    1.195625
 Exp_ADSL_CLEC_Qlty4 |   .9968364   1.936998     0.51   0.607    -2.799611    4.793284
 Exp_ADSL_CLEC_Qlty5 |  -1.028206    1.74516    -0.59   0.556    -4.448657    2.392245
Exp_Fiber_CLEC_Qlty5 |   .7254898   1.618034     0.45   0.654    -2.445799    3.896779
              areaLn |   .4020212   .8366828     0.48   0.631    -1.237847    2.041889
            popDenLn |  -.2090502   .1254246    -1.67   0.096    -.4548779    .0367775
           PopGrowth |   .1483223   .2739474     0.54   0.588    -.3886047    .6852494
        AGE_MEAN_mkt |   .1743126   .3205165     0.54   0.587    -.4538882    .8025135
               ageSq |  -.0013766   .0038701    -0.36   0.722    -.0089618    .0062086
        EDUC_MEAN_mkt |   .0205232   .2514362     0.08   0.935    -.4722827    .5133291
           _RENT_mkt |   2.682896   1.404381     1.91   0.056    -.0696403    5.435432
       _WORKHOME_mkt |  -3.276944   2.775446    -1.18   0.238    -8.716719    2.162831
     _WATERSHARE_mkt |   -9.33643   6.036809    -1.55   0.122    -21.16836    2.495499
            incomeLn |   .5662495   .9756816     0.58   0.562    -1.346051     2.47855
       VacancyRateLn |   -.800186    .506348    -1.58   0.114     -1.79261    .1922379
           FIREempLn |   .4413859   1.268756     0.35   0.728    -2.045329    2.928101
               _cons |  -16.23696   12.83698    -1.26   0.206    -41.39698    8.923071
---------------------+--------------------------------------------------------
```

Table continued next page

Table continued from previous page

```
--------------------------------------------------------------------------------------
                     |               Robust
              choice |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------------------+----------------------------------------------------------------
2                    |
   nearestSameTechBB |  -.1777229    .034506    -5.15   0.000    -.2453534   -.1100924
Expected_CM_Quality1 |  -.6136592    1.65967    -0.37   0.712    -3.866552    2.639233
Expected_CM_Quality2 |  -.0345336   2.058637    -0.02   0.987    -4.069388    4.000321
Expected_CM_Quality3 |   12.14041   3.565601     3.40   0.001     5.151961    19.12886
Expected_CM_Quality4 |  -10.58655   3.076683    -3.44   0.001    -16.61674   -4.556366
Expected_CM_Quality5 |   .9277611   1.016057     0.91   0.361    -1.063673    2.919195
 Exp_ADSL_CLEC_Qlty1 |   2.081478   2.114908     0.98   0.325    -2.063665    6.226621
 Exp_ADSL_CLEC_Qlty2 |  -.9551984   2.949447    -0.32   0.746    -6.736008    4.825611
 Exp_ADSL_CLEC_Qlty3 |  -4.057481   1.793257    -2.26   0.024      -7.5722   -.5427621
 Exp_ADSL_CLEC_Qlty4 |   2.644875   1.280144     2.07   0.039     .1358395    5.153911
 Exp_ADSL_CLEC_Qlty5 |  -2.278049   1.234551    -1.85   0.065    -4.697726     .141627
Exp_Fiber_CLEC_Qlty5 |  -.5608268   1.394808    -0.40   0.688    -3.294601    2.172947
              areaLn |  -.3420503   .3312069    -1.03   0.302    -.9912039    .3071034
            popDenLn |  -.0072046   .1163476    -0.06   0.951    -.2352417    .2208326
           PopGrowth |  -.0147865   .2916122    -0.05   0.960    -.5863359    .5567628
        AGE_MEAN_mkt |  -.0682873   .2935228    -0.23   0.816    -.6435815    .5070069
               ageSq |   .0019221   .0035754     0.54   0.591    -.0050854    .0089297
       EDUC_MEAN_mkt |   .0196186   .2104462     0.09   0.926    -.3928485    .4320857
           _RENT_mkt |    2.93263   1.384258     2.12   0.034     .2195339    5.645726
       _WORKHOME_mkt |  -7.095614   2.684007    -2.64   0.008    -12.35617   -1.835057
     _WATERSHARE_mkt |   .0487735   3.898462     0.01   0.990    -7.592071    7.689618
            incomeLn |   .3991842   .8799402     0.45   0.650    -1.325467    2.123835
       VacancyRateLn |  -.8423771   .4424499    -1.90   0.057    -1.709563    .0248087
           FIREempLn |  -.1848963   1.321602    -0.14   0.889    -2.775189    2.405396
               _cons |  -9.969011   12.57489    -0.79   0.428    -34.61535    14.67733
---------------------+----------------------------------------------------------------
3                    |
   nearestSameTechBB |  -.2045352   .0440739    -4.64   0.000    -.2909184    -.118152
Expected_CM_Quality1 |   2.321899   1.937166     1.20   0.231    -1.474877    6.118674
Expected_CM_Quality2 |  -3.027535   2.354243    -1.29   0.198    -7.641767    1.586697
Expected_CM_Quality3 |   4.398966   3.874984     1.14   0.256    -3.195863     11.9938
Expected_CM_Quality4 |  -2.075728   3.355484    -0.62   0.536    -8.652355    4.500899
Expected_CM_Quality5 |    1.28571   .9971871     1.29   0.197    -.6687413     3.24016
 Exp_ADSL_CLEC_Qlty1 |   2.199401   2.133433     1.03   0.303     -1.98205    6.380853
 Exp_ADSL_CLEC_Qlty2 |  -.9590137   3.004932    -0.32   0.750    -6.848571    4.930544
 Exp_ADSL_CLEC_Qlty3 |  -2.363884   1.820055    -1.30   0.194    -5.931126    1.203358
 Exp_ADSL_CLEC_Qlty4 |   2.252075   1.213814     1.86   0.064    -.1269563    4.631106
 Exp_ADSL_CLEC_Qlty5 |  -6.185145   1.499646    -4.12   0.000    -9.124397   -3.245893
Exp_Fiber_CLEC_Qlty5 |   .8785445   1.321831     0.66   0.506    -1.712197    3.469286
              areaLn |  -.2591963   .3543046    -0.73   0.464    -.9536205    .4352279
            popDenLn |   -.064573   .1167247    -0.55   0.580    -.2933492    .1642031
           PopGrowth |  -.5781168   .4325214    -1.34   0.181    -1.425843    .2696096
        AGE_MEAN_mkt |  -.0641535   .3035055    -0.21   0.833    -.6590132    .5307063
               ageSq |   .0014001   .0037028     0.38   0.705    -.0058573    .0086574
       EDUC_MEAN_mkt |  -.0032367    .211315    -0.02   0.988    -.4174065    .4109331
           _RENT_mkt |   2.017749     1.3496     1.50   0.135    -.6274183    4.662917
       _WORKHOME_mkt |  -7.445426   2.614233    -2.85   0.004    -12.56923   -2.321623
     _WATERSHARE_mkt |  -7.152064   4.262157    -1.68   0.093    -15.50574     1.20161
            incomeLn |    .000525   .9197481     0.00   1.000    -1.802148    1.803198
       VacancyRateLn |  -.4264742    .441726    -0.97   0.334    -1.292241    .4392929
           FIREempLn |  -1.007138   1.282111    -0.79   0.432    -3.520029    1.505753
               _cons |  -5.700624   12.97699    -0.44   0.660    -31.13506    19.73381
---------------------+----------------------------------------------------------------
```

Table continued next page

Table continued from previous page

```
------------------------------------------------------------------------------------
                     |               Robust
              choice |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------------------+--------------------------------------------------------------
4                    |
      nearestSameTechBB |  -.2867449   .0608138    -4.72   0.000    -.4059378   -.1675519
Expected_CM_Quality1 |   2.851677   1.937474     1.47   0.141    -.9457034    6.649057
Expected_CM_Quality2 |  -3.843347   2.273202    -1.69   0.091    -8.298741    .6120472
Expected_CM_Quality3 |   2.877306   3.894042     0.74   0.460    -4.754875    10.50949
Expected_CM_Quality4 |  -.0436249   3.416056    -0.01   0.990    -6.738971    6.651721
Expected_CM_Quality5 |   1.695662   1.020376     1.66   0.097    -.3042389    3.695563
 Exp_ADSL_CLEC_Qlty1 |   2.663995   2.117224     1.26   0.208    -1.485688    6.813678
 Exp_ADSL_CLEC_Qlty2 |   -.790212   2.956526    -0.27   0.789    -6.584896    5.004472
 Exp_ADSL_CLEC_Qlty3 |  -3.599671   1.830709    -1.97   0.049    -7.187794   -.0115469
 Exp_ADSL_CLEC_Qlty4 |   3.253188   1.304257     2.49   0.013     .6968909    5.809485
 Exp_ADSL_CLEC_Qlty5 |  -3.738563   1.179449    -3.17   0.002    -6.050241   -1.426886
Exp_Fiber_CLEC_Qlty5 |   1.472453   1.333173     1.10   0.269    -1.140519    4.085425
              areaLn |  -.5987834   .3615515    -1.66   0.098    -1.307411    .1098446
            popDenLn |   .0207623   .1183542     0.18   0.861    -.2112078    .2527323
           PopGrowth |  -.1246738   .2665312    -0.47   0.640    -.6470652    .3977177
        AGE_MEAN_mkt |  -.2474044   .2990171    -0.83   0.408    -.8334671    .3386584
               ageSq |   .0034348   .0036747     0.93   0.350    -.0037675     .010637
       EDUC_MEAN_mkt |   .1461308   .2129376     0.69   0.493    -.2712192    .5634807
           _RENT_mkt |   1.394047     1.3606     1.02   0.306     -1.27268    4.060774
       _WORKHOME_mkt |  -7.182032   2.861303    -2.51   0.012    -12.79008   -1.573982
      _WATERSHARE_mkt |   -2.34731   3.944568    -0.60   0.552    -10.07852    5.383902
            incomeLn |  -.3398597   .9353834    -0.36   0.716    -2.173178    1.493458
       VacancyRateLn |  -.9037069    .432946    -2.09   0.037    -1.752265   -.0551484
           FIREempLn |  -.7777628   1.265103    -0.61   0.539    -3.257319    1.701793
               _cons |   1.772227   13.04927     0.14   0.892    -23.80388    27.34833
------------------------------------------------------------------------------------
```

Chapter 14

The High Cost of a Cheap Lesson in Wireless Access

> [Economic experiments] ... include experimentation with new forms of economic organization as well as the better-known historical experiments that have been responsible for new products and new manufacturing technologies.
>
> Nathan Rosenberg, 1992[i]

In the late 1990s many industry developers were unsure whether any design for wireless Internet access would ever become embodied in a mass-market product. As it turned out, similar to other developments in the commercial Internet, the crucial decisions behind wireless local area networking—today called Wi-Fi—did not come to fruition by a simple or straightforward path. Its birth arose from a series of experiments, and these experiments involved decisions by several key industry executives, principally Steve Jobs and Michael Dell. Those decisions built upon several other experiments by a committee of engineers, and a set of policy makers at the Federal Communications Commission, which this chapter explains. Although it may not be apparent at first glance, the economic experiments that led to Wi-Fi contained many features found in other experiments that led to the commercial Internet.

The institutional details are generally underappreciated by all but insiders. The committee of engineers was sponsored by the IEEE (Institute of Electrical and Electronics Engineers), which helped form many committees that endorsed standards for interoperable products, and these designs helped coordinate designs from multiple suppliers. The standard for what later became Wi-Fi came from Subcommittee 802.11, the eleventh subcommittee to explore issues within the domain of Committee 802. Committee 802 was formed in 1980 and explored standards for local area networking, while subcommittee 11 was formed in 1990 and explored standards for wireless data communications. The subcommittee designed a standard for both antennae and laptop receivers, and that became crucial for the birth of mass-market wireless Internet access over short distances, such as a hundred feet. The crucial design emerged in 1999, when the committee published Standard 802.11b, which altered some features not found in Standard 802.11a (changing the frequency of electromagnetic spectrum it used, among other things).[ii] The draft of 802.11b eventually caught on.[iii]

How did Steve Jobs become involved? Jobs had just returned as Apple's CEO in 1997, and he initiated a meeting with executives at Lucent, who had supplied many key engineers to Subcommittee 802.11, including its leader, Vic Hayes. Lucent's wireless LAN management welcomed the phone call and initially viewed it as payback for all of its investment of time and personnel in the subcommittee.

Lucent got more than they bargained for. Lucent's management anticipated bargaining hard to become the dominant equipment supplier of the hardware to make wireless LANs operate within laptop personal computers. Apple was still a fraction of Lucent's size, so Lucent expected a certain amount of deference from Apple, though, that is not how it played out. Cees Links, from Lucent, attended the meeting and in a memoir about the growth of Wi-Fi (written much later) he describes how it began awkwardly.[iv] One side showed up in suits and ties, while the West Coast engineers showed up in more relaxed wear. Jobs did not show up at first, and nobody would start without him. After some awkward small talk, Jobs finally walked in late. From thereon he did the majority of the talking. Links began a planned slide presentation, and described it this way:

> Then Steve asked, "Are there any questions?" I tried to show a few slides: key wins, market positioning, product offering, value creation, etc. Presenting slides with Steve Jobs is actually quite easy: you put up the slide, and he will do the talking, not necessarily related to the slide: then he asks for the next slides.

Links goes on to describe a short dialogue between Jobs and the senior management team from Lucent. This is where the management made their pitch and where they had planned to bargain hard. In response Jobs described what he wanted. Links's description of the end of the meeting is the most revealing:

> Turning the conversation back to wireless LANs, [Jobs declares,] "We need the radio card for $50 and I want to sell at $99." Then Steve apologized; he had to leave. Standing up, he said "Hi!" and went. The room fell silent.

The silence was understandable. Up until that point none of the wireless local area networking producers had ever achieved that price point, or anything near it. This price level was regarded as quite an ambitious target by Lucent, and any other equipment supplier.[v] Steve Jobs, on the other hand, acted as if he was delivering the simple hard truths about making mass-market products, and, as he did often with his unique managerial style, expressed impatience with anyone who did not see the vision he regarded as obvious.

Later events favored Jobs's point of view. NCR/Lucent eventually achieved that price point, albeit only after negotiations continued with Apple throughout production, as Apple changed the product requirements.[vi]

The Apple Airport—the first mass-market Wi-Fi product—debuted in July 1999 at a MacWorld convention in New York. As far as the insiders in subcommittee 802.11 were concerned, there was nothing technically new about it; it merely embedded the 802.11b design in a functioning product that Apple sold. However, it did something no prior wireless product had done: it was aimed at the mass market. It came in a branded product from Apple, Apple distributed the entire system, and the price and functionality—such as the data throughput speed—were good enough for a typical PC user.

That is not the whole story. Lucent's cards for Apple laptops and the Apple Airport system did not serve those who had PCs from suppliers other than Apple. That still left a large part of the PC industry uncovered. Most PCs used a Windows operating system from Microsoft. As it happened, that market became first addressed at Dell Computer.[vii]

Michael Dell, founder and CEO of Dell Computer, by then one of the largest PC providers in the world, heard about the announcement of the Apple Airport and called Lucent. According to the account from Cees Links, Dell was “furious” with them because Dell was not first to experiment with a product release.[viii] Lucent executives had to remind Michael Dell that he had an opportunity to be in on discussions as early as 1992. However, Dell had decided in 1993 to stop the discussions because he concluded (incorrectly, as it turned out) that there was no market for the technology.[ix]

The two parties subsequently came to a deal. Making a version for Dell became a priority thereafter, and making it compatible with Windows XP was the main challenge for the team at Lucent. Eventually Lucent would succeed at that as well. To do that Lucent and Microsoft cooperated in changing the design of Windows XP, and a new version was released in 2001. It supported 802.11b in all Windows-based systems. Just as with Apple, Lucent made a hardware card for an external slot in a PC.

Those first two projects established the mass market for laptop use, pioneering the technical issues affiliated with the challenges of the Apple and Windows operating systems. Both were investments in product designs embedding 802.11b, aimed at fostering sales as part of either Apple’s or Dell’s portfolio of products. In both examples, one pioneering firm, Lucent, would gain considerable sales from its position, and (in retrospect) would retain the position as a leading provider of equipment for several years.

The importance of those two experiments for the market’s development is more readily apparent in retrospect than it was at the time. After those two projects, the consumer-oriented mass market for wireless Internet access took off, with a large number of other firms also entering into production. Those two projects served as the bridge between years of experimentation with prototype designs and the design and distribution of products for mass markets, showing other equipment firms that real money could be made if oriented toward the demand the pioneering firms perceived.

Looking more deeply behind events, a complete explanation requires understanding both technology and market institutions. This chapter must go inside a committee that few nonengineers ever have heard of, IEEE Committee 802.11. This committee developed a new standard and made the design available without restriction. As this chapter describes, the subcommittee designed a standard around a set of flexible government rules for the electromagnetic spectrum, the electromagnetic wavelengths invisible to the eye. Those rules were themselves an experiment. Such flexible rules had never been deployed by any government, or by the Federal Communications Commission, and this chapter must explain the sense in which the rules were novel. The key insight seems too simple to be so profound: flexible rules would enable commercial firms to put many options in front of users, and users would choose which applications gave them the most value. User choice in this market determined something unprecedented: few of the original use cases for the spectrum remained popular. Instead, the vast majority of use migrated to an application they liked more—namely, wireless Internet access.

That last observation takes steps toward the deeper economic lesson of this episode. Flexible rules allowed spectrum to move from low-value to higher-value activities. That sentence may seem obtuse at this point in the chapter, but it represents a profound shift in government policy for the spectrum. Policy makers did not intend to foster innovation from the edges, at least not explicitly, but that is what they ended up doing nonetheless.

Adopting Rules for the Spectrum

The story of Wi-Fi begins in the early 1980s in the Federal Communications Commission, which is based in Washington, DC. As is typically the case in Washington, DC, a simple proposal does not get far until sensible voices of all ideological stripes see the wisdom in it, albeit each may see something different in it.

Like every other government, until the 1990s the United States government had a very restrictive system for allocating the spectrum. Governed by the 1934 Communication Act, the law gave the Federal Communications Commission (FCC) authority to license or bar companies from using the spectrum. Known as allocation through “command-and-control,” the FCC allocated each unique wavelength to a particular firm for a specific purpose, such as radio, television, and mobile telephony. This system was first adopted in order to minimize one user interfering with another. At one time it was thought that interference was a primary concern in all uses of spectrum. Hence, a central government administration could allocate rights to use spectrum, as well as determine other technical details, such as power over frequencies, which prevented one user’s activity from stepping on top of another.

To put it in very human terms, until the 1990s the owner and the purpose for the spectrum were determined far in advance by expert committees comprised of engineers, whose deliberations were approved of by the FCC. Spectrum was given to specific firms and for very specific purposes. These choices had the force of statute behind them and could not be undone except by the FCC. The choices committees made were rarely reversed, and only in exceptional circumstances.[x]

Circumscribing use eliminated many economic experiments before any ever got started. Why bother with an experiment if it would require moving spectrum from one use to another that command-and-control would not approve? For many years that begged a question: If market participants had had the ability to decide how to employ the spectrum for a range of economic experiments, would they come to a different conclusion about how to deploy it? If it were possible, would spectrum move from a use with low value to one with high value?

Once again, it is possible to put a very human face on that question. In the early 1980s, one employee at the FCC, Michael Marcus, asked this question about the spectrum for short-range uses.[xi] The question was rather pointed in the context of short-range uses, because one short-range application was less likely to interfere with another. Distance (or low power) prevented equipment in one location from interfering with another only a few dozen feet away. Moreover, in some of the primary short-range applications—for example, garage door openers, baby monitors, and wireless handsets—households used the spectrum infrequently, perhaps only a few times a day. Why would the FCC have to worry about interference if a user could transmit a signal no more than one hundred feet? Perhaps neighbors could work out the issues themselves, or perhaps simple technical solutions could be found (such as automated selection among multiple channels). The FCC could just leave market participants to find those solutions. Why would the FCC have to designate the licensee, an owner and designer, when plenty of firms could supply such equipment? More to the point, why would the FCC want the administrative burden of licensing hundreds of firms who made use of the spectrum in thousands of places?

Initially Marcus received a warm reception from those who had sympathy for free-market ideology. They saw an opportunity to make use of markets. The recent election of the Reagan administration had installed many commissioners with such sympathies. A task force was appointed and began to consider how such a system would work.

But the initiative did not make it from blackboard to implementation at any point in the 1980s. Adherents to the established system raised many questions, and momentum stalled. After many years another set of commissioners determined the priorities in the agency. Eventually backlash inside the FCC bureaucracy became powerful again, especially among those who did not see any merit to departing from command-and-control mechanisms. Marcus then became a target of deliberate efforts aimed to make him leave the FCC; he received terrible employee reviews and was hounded out of his job.[xii]

After the 1992 election the Clinton administration installed commissioners who had a taste for reform. Giving the spectrum to users appealed to those who wanted to experiment with new forms of government, such as diffusing discretion to users and small manufacturers. Reed Hundt, the new chair of the FCC, felt he had a mandate for action, and he took it in many different areas, including spectrum policy. He foresaw information and communication technologies as a bridge toward a revolution in new services and productivity growth.[xiii]

The FCC took up Marcus's proposal again, and this time pushed it through. By late April 1996 the FCC took the last legal step. The FCC initiated a "Notice for Proposed Rule Making" to make available a small amount of unlicensed spectrum for what became known as Unlicensed National Information Infrastructure (U-NII) devices. It was understood from the FCC's order that the commission anticipated "short range, high-speed wireless digital communications" and devices that supported "the creation of new wireless local area networks ('LANs') and … facilitate wireless access to the National Information Infrastructure ('NII')."[xiv] Beyond that, however, little else was specified about the design or application of the spectrum. After deliberating over that summer, the commission made the spectrum available. The order that emerged on January 9, 1997, stated, “we are adopting the minimum technical rules necessary to prevent interference to other services and to ensure that the spectrum is used efficiently.”[xv]

Taking a minimal approach was, in fact, the key administrative innovation, though that was not apparent to all observers at the time. Traditional defenders of command-and-control regarded the allocation, known as the Part 15 rules, as “garbage spectrum,” a throwaway to uses with low value, and a symbolic salvo in an ideological battle. The standard use cases used for reference—as mentioned, garage door openers, wireless handsets, and baby monitors—were also thought to be low value in comparison to radio, television, and mobile telephony. Perhaps a business case could be made for use in a warehouse, but this was not regarded as particularly valuable.[xvi] More to the point, forecasts about mass market wireless access to Internet data services did not play a central role in the design of these rules, or tip sides in the ideological fights in favor or against aspects of these rules. These rules issued as the commercial Internet began, and connection between these commercial events was distant. Use of euphemisms like the “National Information Infrastructure,” as quoted above, was symptomatic of that distance.

Something crucial was embedded in the Part 15 rules, however. Consistent with Mike Marcus’s original proposals, the spectrum did not have tight restrictions on its purpose—that is, the FCC did not control how the spectrum was used, or the purpose to which it was put. Any application was acceptable as long as it did not interfere with other activities outside the band. Even different applications could use the same spectrum. Related, the spectrum did not get designated to a single owner, and the FCC did not choose the equipment makers. Equipment makers were free to design their products in response to what they learned about its value from experimenting with its use. The spectrum was released with a few minor usage cases in mind, but the rules were made flexible enough to accommodate additional invention of new uses.

Standard Committees and Their Designs

The FCC has the force of law behind its actions. In contrast, any standard emerging from discussions at an IEEE committee was not legally binding on industry participants, and, accordingly, it was called a “voluntary standard.” The committees did their work in the hope that such a design could act as a focal point for an interoperable design. In the best case, firms would embed the design in their products, such as routers and receivers, and these would become interoperable as result. Firms could differentiate along other dimensions, such as other aspects of product design, brand, or distribution.

Committee 802 was formed in the early 1980s, before the privatization of the Internet was ever proposed. By the late 1980s the committee was well known among computing and electronics engineers because it had helped design and diffuse the Ethernet standard for local area networks.[xvii] By the late 1980s the 802 committee had become a victim of its success, and it had grown larger, establishing subcommittees for many areas, ostensibly to extend the range of uses for Ethernet.

Subcommittee 802.11 was established in 1990. Like all subcommittees of this broad family of committees, it concerned itself with a specific topic, in this case, designs for interoperability standards to enable wireless data traffic over short ranges—ostensibly doing with wireless technology what a local area network did with wires. A close look at the engineering suggests, however, that this label was mere window dressing for otherwise complex deliberations. For example, while “wireless local area networking” accurately described the aspirations for users, for suppliers the technical issues in this area hardly had any connection to the technical issues in existing local area networking. As it would turn out, due to the very different error-correction issues, the software for a wireless local area network would end up bearing only a slight resemblance to that for a wired local area network, and contained many important technical differences.[xviii]

At first the committee did not get very far, lacking any clear direction. But then a new chair was appointed, Vic Hayes. Hayes was a technologist with a visionary outlook, a cheerful demeanor, and, more importantly, the patience to see his vision to realization. Hayes first developed prototypes of wireless technologies for National Cash Register, or NCR. At the time it was a subdivision of AT&T, which would later become Lucent (and later it was a division of Agere Systems). In that capacity Hayes first developed prototypes for wireless terminals for stockbrokers. Other applications for the technology were forecast, such as easy rearrangement of retail space.[xix] From this experience he had a steady and flexible vision of the value of a standard that many component vendors could use to make interoperable equipment.

Other potential applications for this standard came up in the earliest meetings. One of the earliest prototypes had been a wireless local area network for a university campus.[xx] Another was short-range wireless Ethernet in warehouses with complex logistical operations. Several firms had built expensive prototypes of these devices but had not found many buyers, or otherwise experienced very limited commercial response. Indeed, throughout the first half of the 1990s, as the 802.11 committee met and continued to do its work, pioneering firms continued their experiments with different uses and generated little interest among potential buyers, who regarded these prototypes as expensive and not very useful.[xxi]

As with most such committees, Hayes tried to involve members who brought appropriate technical expertise and who represented the views of most of the major suppliers of equipment in which this standard would be embedded. At first, therefore, the group was comprised of enthusiastic designers focused on the needs of big users with potential warehousing applications (for example, FedEx, United Parcel Service, Wal-Mart, Sears, and Boeing), where the application's value did not seem specious. Although it is obvious in retrospect, notably absent from the earliest meetings were representatives of many of the suppliers of valuable equipment a decade later, such as firms from electronics and computing.

Those firms would not be absent for too long. Several related efforts, such as HomeRF and Bluetooth, were founded in 1998. The former was organized by firms such as Motorola and Siemens, and at its peak involved over a hundred companies before it disbanded.[xxii] The latter was established by Ericsson, Sony-Ericsson, IBM, Intel, Toshiba, and Nokia, and currently still exists; today it involves thousands of firms. It focused on very short-range uses—under a few feet, and, as such, tended to have a set of applications distinct from Wi-Fi. Subsequent events would change the predominant use case, as economic experiments showed participants that high market value lay in a different configuration of technology, operations, and pricing than had originally been envisioned.[xxiii]

Embedding the Design in Products

Wi-Fi did not emerge from a set of prespecified designs and classified stages. Economic experiments played a role in shaping that path, as pioneering firms took actions in response to the actions of the standard committee. As defined at the start of the chapter, economic experiments involved more than just new designs. They also involved new forms of economic organization and new products and new manufacturing technologies.

As it turned out, these experiments with Wi-Fi by Apple and Dell, described in the introduction, generated a response from many buyers, who also began to experiment. Users began deploying this equipment in a variety of settings, campuses, buildings, public parks, and coffee shops. Unsurprisingly, other vendors tried to meet this demand as well. Around the same time as the publication of 802.11b, firms that had helped pioneer the standard—including 3Com, Aironet (now a division of Cisco), Harris Semiconductor (now Intersil), Lucent (now Agere), Nokia, and Symbol Technologies—formed the Wireless Ethernet Compatibility Alliance (WECA). WECA branded the new technology "Wi-Fi," which was a marketing ploy for the mass market, since WECA's members believed that "802.11b" was a much less appealing label.[xxiv] They aimed to nurture what enthusiasts were doing and broaden it into sales to a broader base of users.

WECA also arranged to perform testing for conformance to the standard, such as certifying interoperability of antennae and receivers made by different firms. This is valuable when the set of vendors becomes large and heterogeneous, as it helps maintain maximum service for users with little effort on their part. In brief, while the IEEE committee designed the standard, a different body (of similar firms) performed conformance testing.

Events then took on a momentum of their own. Technical successes became widely publicized. Numerous businesses became users of Wi-Fi and began directed experiments supporting what became known as *hot spots*, which was an innovative business idea. A hot spot is a data transmission mediated by a third party for local use in a public space or on a retail premise. A hot spot in a public space could be free, it could be installed by a homeowner, or it could be maintained by a building association for all building residences. It could be supported by the café or by a restaurant or by a library trying to serve its local user base. Or, it could be subscription based, with short-term or long-term contracts between users and providers. The latter became common at Starbucks, for example, which subcontracted with T-Mobile to provide the service throughout its cafés.

Hot spots were similar to, but outside of, the original set of use-cases for the standard.

Since nothing precluded this unanticipated use from growing, grow it did. It grew in business buildings, in homes, in public parks, and in a wide variety of settings, eventually causing the firms behind HomeRF to give up. The growing use of Wi-Fi raised numerous unexpected issues about interference, privacy, and appropriation of the signals of neighbors. Nevertheless, these issues did not slow Wi-Fi's growing popularity.[xxv] Websites sprouted up to give users, especially travelers, directions to the nearest hot spot. As demand grew, suppliers gladly met it. As in a classic network bandwagon, the growing number of users attracted more suppliers and vice versa.

Collective invention played its familiar role. Economic experiments among a community of participants led to many new insights that accumulated over time. While several pioneering firms took important steps in initiating market development, no single firm was responsible for all the economic experiments that eventually altered the state of knowledge about how to best operate equipment using IEEE Standard 802.11b. Rather, many firms responded to user demand, demonstrations of new applications with tangible market experience, and the lessons accumulated.

In this way, Wi-Fi became an industry-wide platform. All participants took actions using standards that invited activity from complementary component providers. In this instance of Wi-Fi equipment, the presence of standard, related institutions for conformance, and the universal participation of virtually all the industry, encouraged experiments in antennae and receiver design, as well as in deployment of final equipment in new operational modes (such as a hot spot). Because Wi-Fi deployed at an industry-wide level, experimenters could presume (safely) that other complements would make use of the same design, which led each experiment to specialize on narrow issues and specific issues of interest to the experimenter.

Interplay

Later events in the development of Wi-Fi illustrate how a firm can invest in building on an economic experiment. Reacting to the experiment that generated Wi-Fi, Intel created Centrino, a large program that would install wireless capability in its notebook computers. It was officially launched in March 2003.

As with many aspects of growth in wireless access, the Centrino program is easy to misunderstand. Centrino was much less obvious to Intel in advance than it was in retrospect. It too was an experiment, just on a very large scale.

At the turn of the millennium Intel's strategic plans were responding to multiple market forces. While demand for desktop and notebook computers had grown along with the Internet in the 1990s, Intel's marketing department forecast an imminent slowdown in the share of desktop sales, as well as increasing engineering challenges supplying faster chips. More worrisome, Intel had branded itself as a firm that always marketed better and faster microchips, while it was no longer clear that demand for bigger and faster would arise across all segments of computing. Notebook users valued mobility, for example, and that placed value on distinct attributes, such as longer battery life, less energy-intensive chips, smaller storage, more compact designs, less weight, and less heat. Even by the late 1990s many mobile users had shown a willingness to give up improvements in bigger and faster microprocessors in order to get improvements on these other attributes.

In 2001, in response to a number of initiatives and studies, Intel's management decided it was time to change priorities. Labeling this a "left turn," the company chose to embed a Wi-Fi connection in all notebooks that used Intel microprocessors.[xxvi] This was easier said than done. The choice *not only* involved redesigning the Intel microprocessor, Intel's core product, stressing lower power and lower processing speeds, but it also involved redesigning the motherboard for desktop PCs and notebooks, adding antennae and supporting chips. Intel made a number of reference designs and made them widely available at low cost.

Intel's management hoped that its endorsement would increase demand for wireless capabilities within notebooks by, among other things, reducing weight and size while offering users simplicity and technical assurances in a standardized function. The firm also anticipated that its branding would help sell notebooks using Intel chips and motherboard designs instead of using microchips from Advanced Micro Devices (AMD). Furthermore, antenna and router equipment makers anticipated that a standardized format for wireless notebooks might help raise demand for their goods.

The redesign brought one concrete benefit to users—namely, it eliminated the need for an external card for the notebook, which was usually supplied by a firm other than Intel and installed by users or original equipment manufacturers (OEMs) in an expansion slot. Intel hoped for additional benefits for users, such as more reliability in wireless functionality, fewer set-up difficulties, longer-lived batteries due to less need for heat reduction, and thinner notebook designs due to smaller cooling units. Seeking to inform users about all those changes, Intel adopted "Centrino" as a label, and it initiated a program to certify compliance.

Intel's management further worried that wireless notebooks would not be used widely enough to merit their investment in Centrino, so Intel's management considered exploring activities far outside of its core product microprocessors. These exploratory actions were not far outside its philosophical approach to managing the demand for microprocessors. Intel long ago made a distinction between managing its first job, making microprocessors, and managing anything that helped it sell more microprocessors, which was often given the label "Job 2."[xxvii]

For example, the company launched a program to change the demand for the wireless feature of notebooks. As part of Job 2, Intel eventually certified fifteen thousand additional hot spots in hotels, airports, and other public places by the time Centrino launched.[xxviii]

As another example of Job 2, Intel made motherboard designs available to others. The firm had crept into the motherboard business slowly over the prior decade as it initiated a variety of improvements to the designs of computers using its microprocessors. Years earlier, the firm had designed prototypes of these motherboards and by the time it announced the Centrino program, it was making some motherboards, branding them, and encouraging many of its business partners to make similar designs. The wireless capabilities of a notebook had not been the focus on these earlier programs, so the announcement of the Centrino program represented a shift in strategic aims and direction for the Intel programs affiliated with motherboards.[xxix]

This latter program illustrates one of the interesting conflicts that emerged in Wi-Fi's development. Intel's motherboard designs could increase the efficiencies of computers, but that benefit was not welcomed by every OEM that assembled PCs or other industry players. Firms such as Texas Instruments and Intersil had lobbied earlier for different designs for the 802.11g upgrade, investing heavily in the efforts at Committee 802.11. Neither of them had intended to help Intel's business, and neither of them wanted to see Intel increase its influence over the designs that were deployed to most users.

Moreover, Intel's designs eliminated some differences between OEMs and other component providers. Many of these firms resented both losing control over their designs and losing the ability to strategically differentiate their own designs. At the same time, other OEMs liked the Intel design, since it allowed the firms to concentrate on other facets of their business. That competitive rivalry eventually generated cooperation from every small OEM, especially after Intel initiated marketing programs for Centrino. These programs were especially necessary to induce cooperation from many big OEMs.

Intel ran into several unanticipated crises, such as insufficient parts for the preferred design and a trademark dispute over the use of its preferred symbol for the program. However, the biggest and most important resistance came from the largest distributor of PCs, Dell Computer, one of the earliest firms to get the market started. Dell insisted on selling its own branded Wi-Fi products, buying internal cards from others that handled Wi-Fi, bypassing Intel altogether. Dell branded its solution and had grown a good side business from its pioneering efforts.

Despite Dell's resistance, the cooperation from antenna makers and (importantly) users helped Intel reach its goals. By embedding the standards in its products, Intel made Wi-Fi, or rather Centrino, easy to use, which proved popular with many users. (Indeed, eventually this success would induce reluctant cooperation from Dell.)

The Centrino example illustrates the array of deliberate firm activities taken during a short period that built on top of learning from an earlier undirected economic experiment. The activities in IEEE Committee 802.11 ended up affecting the activities of many other firms, such as equipment manufacturers, laptop makers, chip makers, and coffee shops, which then shaped new activities in Committee 802.11 as well.

This example also illustrates that economic experiments can—and do—happen in spite of overt conflict between firms. Those firms may be either direct competitors or participants in a value chain with diverging interests. Conflict arises, as it did here, when all can forecast that the success of one firm's experiment adversely affects the business fortunes of another.

Experiments and Creating Value

What do we learn from the evolution of Wi-Fi? The growth of wireless access illustrates many of the elements underlying economic experiments, and, interestingly, these elements were present with the experiments that took place in dial-up access. It also illustrates the errors of many economic myths, and we start with those.

First consider the myth that new market opportunities evolved organically, independently generated by market incentives. Market incentives eventually played a role, but the historical facts suggest a nuanced reading of their role. In this instance a committee of engineers, many employed at industry suppliers, and not all of them participating with the same motive, met for years and designed a standard for a market in which products designed around proprietary standards had not appealed to many users. That involved collaborative ideation and iteration. Only after the emergence of a standard did profit-making activity emerge. Incentives worked indirectly, creating the shadow of future market events on collective action.

Second, incentives also did shape market actions, but not all innovative outcomes. There was a potential disconnect between those who incurred the costs of experimentation and those who learned from it. The disconnection occurred as an unintended by-product of deliberate experiments.

Where did incentives matter directly? By helping market participants learn about the nature of demand in quickly evolving environments, companies tried to position their offerings and pricing structures. Such lessons increased value for the experimenter by generating more revenue through improvement of an existing service, enhancing profits from lowering operation costs or avoiding higher investment expenses, or enhancing pricing power through targeting services to customers better than rivals. In general, many of these benefits could not be measured. If they could be measured—even partially—the private value could be measured in terms of the additional revenue contributed to a firm's business and/or the additional cost savings it generated.[xxx]

Revenue might have increased through altering pricing practices. For example, the acceptable pricing norm for hourly caps changed over time, as ISPs learned about the reaction of different customer segments to distinct menus of choices. Similarly, Wi-Fi prices in many hot spots reflected carrier perception about what the market demand could support.

Pricing experiments often coincided with experiments regarding the range of services offered. During the mid- to late 1990s, for example, virtually all ISPs experimented with changes to the standard bundle offered, such as default e-mail memory, instant messaging support, and hosting services in which the ISP maintained web pages for clients. Most Wi-Fi hot spots, in contrast, did not alter the standard bundle much, restricting access to one simple function. While ISPs experimented with performing services complementary to access, such as hosting services, networking services, and web design consultations, most Wi-Fi hot spots retained their status as stand-alone services.

Learning oriented toward cost reduction resembled learning oriented toward enhanced revenue. For instance, as dial-up ISPs learned from one another about the efficient deployment of 56K modems, those who deployed it found they could charge a modest price premium for faster service (approximately five dollars), but that that premium disappeared in less than a year, after the modems became more common.[xxxi] Similarly, many Wi-Fi hot-spot providers initially charged for access but later found competition reducing their ability to price the service. Instead, Wi-Fi merely became an element of service for a location, often at no charge at all. While that might have led to better customer retention for a café, and eventually manifest as greater sales or higher firm prices, it would have been difficult to attribute a specific change in price or volume to only that investment in Wi-Fi.

More broadly, events in wireless access illustrate that incentives generate action, but others benefit. Pioneers reacted to incentives and expended costly resources on economic experiments. Those costs involved personnel taking time and effort. They also could involve real resources to build prototypes. In some circumstances, however, learning was cheap to an entrepreneur when others took pioneering action. In most circumstances an inexpensive lesson to a later beneficiary came at a high expense to a pioneer, such as a failed business. Succinctly, accounting for industry-wide costs, cheap lessons came at a high cost.

Industry-Wide Benefits

Events in Wi-Fi can be understood with reference to a traditional economic concept called a "learning externality." A learning externality arises when one party learns from the experiment of another but does not take part in the experiment and does not compensate those who incurred the costs from the experiment. Learning externalities were pervasive during the growth of the commercial Internet, and this example illustrates a widely pervasive phenomenon.

Looking back on the experiment in Wi-Fi and other access markets, at least two externalities shaped the experience with learning. There was an information externality *between* firms, as when one firm's directed experiment taught another firm a lesson, or a set of actions interacted in an experiment and taught every industry participant a lesson. There were also information externalities *over time*, as when the lessons of prior experiments generated lessons on which further experiments were built. The example of Wi-Fi shows that these two externalities were pervasive, as well as difficult to distinguish from one another.

Many of these learning externalities were *positive*, that is, one market participant benefited from the actions of another.[xxxii] These positive externalities took one of two forms. In one case, what worked for one firm became known and imitated by others. For example, success from an experiment at one hot spot in one location in 2001 implied it might be profitable in another location with similar features. Alternatively, what did not work for one firm becomes known and, therefore, avoided. For example, the difficulties with the first design for 802.11 become known from experiences in 1997, leading equipment firms to save money by delaying building plans until a more suitable design emerged.

Intertemporal externalities also were common. For example, lessons about how to avoid commercial failure could be as valuable to observers as those who employed them. These inexpensive valuable lessons often emerged this way, and it is easy to see why. The firm whose failure illustrated the lesson for others rarely, if ever, did so for the purpose of teaching others, and almost never under contract with the others who (later) gained the benefit of the lessons learned from the failure. For example, AOL learned from the experiments of many prior ISPs who altered their pricing or product lines, but did not arrange the rest of their product offerings in an appealing fashion. Similarly, many hot-spot developers of Wi-Fi learned from the pricing experience of others who did not get their pricing right.

Which externalities operated quickly and which operated slowly? The answer emerges by distinguishing among four distinct types of lessons that led to positive externalities. The first were *market lessons*. These pertained to norms and patterns of market-based actions, such as how to write a contract that users found acceptable, and how to price services, and so on. Second, *technical lessons* pertained to the design of a piece of equipment—for example, knowing how to configure Wi-Fi so that it worked in the type of space/location at all times that fit the supplier’s needs. Third, *heuristic lessons* combined technical knowledge with either market or operational knowledge about how employees behaved in firms and how customers reacted to firm behavior. For example, knowing how to deploy Wi-Fi for a maximal set of users was such a heuristic lesson. Fourth, *complex lessons* were marketing and operational lessons that involved many functions inside an organization—for example, knowing how to integrate the use of Wi-Fi into a wide variety of other offerings. These four types of lessons spread at different rates.

Consider technical lessons. In 1999 the technical lessons about Wi-Fi were often rather trivial for an ISP to learn. Generally, these technical skills were common among those who operated bulletin boards, computers, ISPs, or related equipment. Most local and national ISPs already had procedures in place to, for example, implement billing, publicize their services to local users, or address user service calls. Doing so for Wi-Fi in a coffee shop or restaurant was easy. Although the market actions changed, these were relatively easy to execute within existing organizational procedures.

Technical lessons tended to spread easily for another reason: they tended to become codified quickly.[xxxiii] For example, most equipment suppliers in competitive markets would not consider selling equipment if information about it were not codified because most buyers demanded it as a condition of purchase. Related, vendors of equipment also would have developed a set of marketing parameters for their buyers, which guided them toward best-practice deployment.

Others lessons pertained to heuristic knowledge about how to operate that equipment efficiently, and these too spread comparatively quickly. For example, lessons about how to manage a Wi-Fi router at peak usage levels was not known initially after a new piece of equipment became available for use, but such lessons would be learned through trial and error.

As it turned out, those lessons spread to different coffee shops through a variety of mechanisms—administrators in key locations coordinated it (for example, at Starbucks), franchises communicated with one another (for example, at McDonalds), bulletin boards emerged to support different types of user groups, and the Wi-Fi association invested in support activities as well.

Several factors affected the speed at which heuristic lessons spread, and, as a result, these could spread quickly or slowly. On the one hand, some heuristic lessons spread slowly because, as sources of potential competitive advantage, the firms that first discovered them guarded them. For example, firms guarded their strategies for how to deploy equipment efficiently, and they may also have guarded information that indicated details about their future designs. On the other hand, some firms, such as equipment providers, had strong incentives to spread lessons, since their spread contributed to further sales. Such tension was inherent in the diffusion of Wi-Fi, for example. Intel's program to further fund development of certification of hot spots is another illustration.

The similarity between vendor organizations also shaped spreading of lessons. Most dial-up ISPs used similar software tools for monitoring users, particularly after these showed up in the discussion boards at an open source project, such as Apache, the most popular web server. The community effectively coordinated many innovative efforts for dial-up ISPs in the mid- to late 1990s, by sharing multiple upgrades and fixes to the source code among ISPs. By supporting similar technology, operations, and heuristics, the designs embedded in standards in many organizations also contributed to sharing of lessons. For instance, organizations, such as the Internet Engineering Task Force, (IETF) and the World Wide Web Consortium (W3C) facilitated the movement of lessons. Seen from this perspective, the 802.11 committee for Wi-Fi and WECA helped the lessons spread quickly.

The other side of the same coin is variance in idiosyncratic factors, which could slow the codification of such heuristic lessons. One community of users may differ from another. For example, though peak ISP usage occurs around the same time of day in different locations, surfing behavior varied according to gender, family status, age, education, and income of the members of the household. The sum of these varied across cities, and even from one vendor to another within the same city. Such variety interfered with finding commonalities in, for example, marketing strategies for a new feature across locations or vendors.

Complex lessons spread slowly. In part this was due to the lengthy investigations by firms seeking to lower cost or generate extra revenue. They often were interdependent, where one operational goal reinforced the other, or associated with unique firm features, such as scale. Almost by definition, these lessons resisted immediate codification and moved slowly from firm to firm. For example, management at one hotel chain would not lightly discuss with other hotel chains which type of customer showed a willingness to pay for Wi-Fi. This is not surprising, since firms often hesitated to share information about what sort of costly activities built customer retention most effectively—for example, did users have greater willingness to pay incrementally for access or as a standard part of their contract?

That does not means complex business lessons never spread. Rather, they spread with more effort and at greater cost. In general, they spread more slowly. Even while technical information and market lessons moved quickly between locations and firms, the ability of a firm to prevent direct rivals from imitating its business actions immediately slowed others. Some complex lessons also did not tend to spread to others, at least for a short time.

Economic Experiments in a Complex World

Stepping back from events, it is possible to see that policy makers did not foresee—indeed, could not have foreseen—the consequences of their flexible spectrum policies. Flexibility unleashed two complementary economic processes: movement of value and learning from experimentation.

The spectrum moved from low-value to higher-value activities as it became embedded in different products. Users bought Wi-Fi products instead of garage door openers and baby monitors. In comparison with the old command-and-control rules for allocating spectrum, which often fixed the application, the movement between uses occurred quickly. Users made their choices, and suppliers followed demand. More to the point, it occurred much more quickly than it would have occurred if government managers had retained the right to approve of a change in application.

Flexibility also enabled the firms to learn from experimentation. There were few restrictions on how accumulated lessons got used and by whom. Technical, operational, and heuristic lessons spread quickly, while some complex lessons did not spread at all. That allowed the spectrum to move between different users for Wi-Fi, such as between users who deployed Wi-Fi in business spaces and homes and those who deployed it in public spaces, such as café's and airports. Suppliers learned from one another's experiments, and moved to supply Wi-Fi wherever they could capture some of the created value. In comparison with the old command-and-control rules for allocating spectrum, which often fixed the identity of the seller who embedded the spectrum in products, this movement between uses occurred quickly.

Said simply, the flexible rules allowed lessons from experience in the market to spread quickly and widely. Accumulated lessons were built on the experience of other mistakes and triumphs. Almost by definition, the knowledge pool contained more lessons than any single firm could have learned on its own.

Accumulated knowledge also exhibits a mismatch between cost and benefit. Those who paid for lessons were not necessarily those who used them most profitably. As it typically turned out, many firms taught others lessons, and that had little to do with the original motives. Pioneers conducted the earliest experiments, and the timing of investment was determined by concerns that later participants did not—indeed, could not—influence. Only later, the most important lessons for value became known, creating the potential for regret over an earlier experiment that could have benefited others.

That last observation raises another question: What motivated a manager to pay for an experiment in the first place? Surely nobody was trying to do a favor for an unknown later participant in the market, so why conduct a lesson at all? While some lessons were conducted for the gains they generated for a firm, sometimes it seemed as if economic incentives did not drive a trial. Other inducements mattered just as much—the itch of curiosity, or the sporting thrill from doing something before others. Similarly, why did anyone let a lesson spread? Sometimes competitive pressures induced firms to learn from one another. Yet it also seemed as if lessons spread for reasons far less weighty than the market consequences, as when a manager boasted to a friend about inventing a clever enhancement, or about earning customer kudos for solving a common problem in a novel way.

Regardless of the reasons, the experience in wireless spectrum illustrates the benefits to enabling experiments by market participants. Expensive lessons appeared cheap to later borrowers, although no accountant would (or could) have recorded their value in a ledger, and both users and suppliers benefited.

---

i Rosenberg (1992), 181.

ii The subcommittee first proposed a standard in 1997 that received many beta users, but also failed to resolve many interoperability issues (among many issues). Learning from this experience, and viewing its efforts as racing against those of a private consortium—called HomeRF—the subcommittee rewrote the standard over the next two years. What came to be known as standard 802.11a was ratified in early 2000, just after 802.11b was ratified.

iii Standard 802.11a was licensed for usage in Europe and Asia as well as North America, while for some time 802.11a was only licensed in North America. Liu (2001) or Kharif (2003).

iv Lemstra, Hayes, and Groenewegen (2010), chapter 4, 129–31.{{is the chapter number necessary?}}

v A $100 retail price would have been anything from one-fifth to one-tenth the price of equipment in the first half of the decade. A cost of $50 would require economies of scale in production and extensive use of standard components, as the production cost of cards was higher than $100 at the time of the meeting between Apple and Lucent. As described in Lemstra, Hayes, and Groenewegen (2010), chapter 4, 131, it required Lucent to put into the initial price some of the learning curve benefits it anticipated, which was a departure from existing practice.

vi See Lemstra, Hayes, and Groenewegen (2010), chapter 4, 130, which describes changes in product requirements linked to “all-or-nothing type of negotiations.”

vii For an account see Lemstra, Hayes, and Groenewegen (2010), chapter 4, 131–32.

viii Lemstra, Hayes, and Groenewegen (2010), chapter 4, 131. Links says Michael Dell “was furious about the fact that he had been beaten by Apple.”

ix Lemstra, Hayes, and Groenewegen (2010), chapter 4, 131.

x Perhaps the best-known reversal occurred during the design of the broadcasting standard for color television. The Korean War delayed the deployment of the first approved design, and the after the war it was reconsidered, leading to deployment of a technically superior design from another designer.

xi Marcus (2009).

xii Marcus (2009).

xiii Hundt (2000).

xiv See the review of FCC policies found on http://www.cybertelecom.org/broadband/wifi.htm, accessed May 2007. Subsequent clarifications and rules aligned the spectrum in the United States with similar policies elsewhere.

xv See the review found on http://www.cybertelecom.org/broadband/wifi.htm, accessed May 2007.

xvi See the discussion in Marcus (2009).

xvii See Burg (2001) for analysis of the growth of a local area network market and activities in Committee 802.

xviii See chapters 2, 3, and 4 of Lemstra, Hayes, and Groenewegen (2010).

xix See Kharif (2003).

xx See the description of Hills (2005), who began developing the equivalent of a Wi-Fi network for the Carnegie Mellon campus in Pittsburgh, starting in 1993.

xxi Also see chapters 2, 3, and 4 of Lemstra, Hayes, and Groenewegen (2010).

xxii HomeRF did not generate the enthusiastic sales that those who designed it predicted—even though the

---

designers considered it technically superior to the alternatives. For speculation about why HomeRF failed, see, e.g., http://www.cazitech.com/HomeRF_Archives.htm, accessed in November 2010.

xxiii See Rosenberg (1996).

xxiv The choice of the label "Wi-Fi" resembled "Hi-Fi" or high fidelity, a term commonly used to describe high quality and expensive musical components. The label was meant to signal high quality transmission. Yet 802.11b actually has little to do with music or fidelity, and "Wi-Fi" is a made-up phrase.

xxv In high-density settings it was possible for there to be interference among the channels, or interference with other users of the unlicensed spectrum reserved by the FCC, such as cordless telephones. The diffusion of so many devices also raised questions about norms for paying for access in apartment buildings, from neighbors, and others. See Sandvig (2004).

xxvi For a full account see Burgelman (2007).

xxvii Gawer and Cusumano (2002).

xxviii Burgelman (2007).

xxix For history and analysis of why management chose to invest heavily in some complementary technologies and not others, see, e.g., Gawer and Cusumano (2002), and Gawer and Henderson (2007).

xxx Filtering between noise and cause is a key challenge in such experiments. See Thomke (2003a) for approaches for designing experiments so they can be measured in settings where firms control many of the key aspects of the experiment.

xxxi See Stranger and Greenstein (2007).

xxxii In standard economic parlance, if it were possible to anticipate the benefit and to write a contract over its measured levels, the beneficiary would have *paid* for the benefit.

xxxiii In this context, "codified" refers to an idea put in a structured format that another technically trained individual can understand without having the author present—e.g., words, mathematical formulas, plans, pictures, or professional drawings. See, e.g., the discussion in Nelson (2007).

OPPORTUNITY COSTS AND THE VALUE OF THE MARGINAL PRODUCT ACROSS SPECTRUM ALLOCATIONS[1]

Thomas W. Hazlett[2] & Michael Honig[3]

April 28, 2015

---

[1] Paper for the Colorado/Northwestern Cable Academic Workshop, May 5-6, Northwestern University School of Law. This is preliminary work-in-progress for Workshop distribution only. Please do not cite without permission of the authors.

[2] H.H. Macaulay Endowed Professor of Economics, Clemson University: hazlett@clemson.edu.

[3] Professor of Electrical Engineering & Computer Science, Northwestern University: mh@eecs.northwestern.edu.

ABSTRACT

Rapid increases in data flows through both cellular networks and local hotspots have raised new contentions about radio spectrum policy. Observing trends in which wi-fi, Bluetooth and Zigbee have become widely popular, and mobile operators offload data traffic to local hotspots, some argue that unlicensed allocations hosting such wireless flows are increasingly valuable relative to licensed bands. A shift in regulatory practice, favoring unlicensed, is advocated.

We show that this argument fails both empirically and theoretically. In empirical terms, studies estimating unlicensed bandwidth values grapple with the lack of market prices. Errors emerge with the use of various proxies for value, such as bits. Innovation gains associated with unlicensed bands, proxied by FCC equipment authorizations, are unsupported by economic evidence. More fundamentally, a unilateral approach reflects a lack of symmetry. Marginal allocations of liberal licenses have been shown to produce very high consumer welfare gains (Hazlett & Munoz 2009), dominating unlicensed magnitudes, while some unlicensed bands have performed poorly relative to licensed alternatives. Indeed, the oft-cited “mobile data tsunami,” the development of 2G, 3G, and 4G cellular networks, and the robust input market for licenses yielding exclusive spectrum rights, suggest that restricting such allocations will incur substantial opportunity costs. Abstracting from such considerations inappropriately truncates empirical assessment.

As a matter of economic logic, the value of the marginal product (VMP) of a given spectrum allocation is miscalculated. In essence, the asserted popularity of given applications are put forward proxies as for the value of spectral inputs. This approach ignores key elements of VMP estimation, including input substitution, and the resulting policy claims commonly ignore the opportunity costs of spectrum as well as the transaction costs incurred by the regulatory regime. The result is what we deem the *Broadcast TV spectrum Valuation Fallacy* – the idea that, just because TV programs are popular, TV channels are efficiently deployed.

We propose a more fulsome evaluation of spectrum policy, one that includes not only the appropriate measures of VMP and opportunity cost for rival allocations, but incorporates administrative costs. Spectrum rules that delay useful activities and impose rigidities to stymie future adjustments to emerging opportunities are costly whenever approaches more conducive to dynamic change are available.

Table of Contents

I. INTRODUCTION

With wireless usage booming, political pressure is felt by policy makers to allocate more radio spectrum for useful services. Regulators have long maintained that the most useful frequency spaces have already been allocated, that we have "run out of available spectrum." Yet, new services or technologies can be accommodated by reconfiguring access rights. Entrants can utilize frequencies by politely sharing bands with existing communications. Alternatively, incumbent users can switch to other frequencies, technologies, service models, or cease operations, making greater bandwidth available for entrants. Such progress may enable greater wireless output.

In choosing where and how to initiate new rules, the Commission follows administrative procedures baked into spectrum law since the Radio Act of 1927. Regulators evaluate the various options, receiving input from various citizens, stakeholders and policy makers, and designate how bands will be deployed – or redeployed – according to "public interest, convenience or necessity." Traditionally, this has invoked regulatory *diktat*, also known as "command and control," where the administrative agency determines precisely what wireless services are provided and how.

In recent decades, however, regulators have specified less, relying on market forces more. This is achieved, in the first instance, by issuing licenses that define spectrum spaces, but then allow licensees the freedom to determine how such spaces are used. These liberal licenses, often analogized to *de facto* property rights in spectrum, are widely used to host mobile voice and data networks via wireless wide area networks, (WWANs). In the second instance, a 1980s deregulation of certain unlicensed bands granted greater flexibility to device makers and radio users, leaving service models unspecified and allowing scope for technical innovation. Coordination is achieved via power limits and (comparatively limited) technical specifications imposed by regulators, as well as industry standards nested within the FCC rules. These bands have become popular for extending fixed networks via wireless local area networks (WLANs).

The popularity of applications using unlicensed airwaves, including cordless phones, wi-fi, and Bluetooth, has led some analysts to declare that the administrative regime should pivot. Harvard law professor Yochai Benkler, for instance, posits that licensed spectrum is becoming obsolete. An "epochal switch" (Benkler 2012, 76) is occurring in which unlicensed spectrum is "the basic model for wireless communications, while various exclusive models — both property-like and command-and-control — are becoming a valuable complement for special cases that require high mobility and accept little latency" (Benkler 2012, 75). The implied policy course: allocate more spectrum via license-exempt rules, less via licenses.

In a similar vein, Stanford University economists Paul Milgrom, Jonathan Levin and Assat Eilet find that "regulation of unlicensed spectrum can be viewed as a successful example of a *managed commons* approach" (Milgrom et al. 2011, 14; emphasis original). Their study, commissioned by Google, asserts that "unlicensed spectrum is an enabling resource that, like other enabling resources and technologies, encourages innovation by many parties." In contrast, "Licensing or ownership that limits access to such resources discourages innovation by giving too much power to the licensee or owner." (Milgrom et al. 2011, 28) The authors conclude that regulators "expand the quality and quantity of unlicensed spectrum alongside that of licensed spectrum" (Ibid., 29).

Both Benkler and MLE place great importance on trends in mobile networks, where operators offload traffic to local hotspots when subscribers are stationary (not in a moving automobile, e.g.) and geographically close to a fixed broadband network available for interconnection. While WWANs have been rapidly growing, MLE note that as much as one-half of all mobile network data traffic is received or transmitted via a WLAN connection. Some mobile networks (in particular, Republic Wireless, an MVNO[4]) are "wifi first," switching to WWAN connectivity only when out of reach of fixed networks. The increasing use of unlicensed frequencies to host traffic originating on mobile networks (using liberal licenses) or broadband networks (using privately owned "spectrum in a tube") is claimed to suggest that the social value of unlicensed spectrum is increasing relative to alternatives, and that shifting government allocations in this direction would improve efficiency in wireless markets.

The analysis is deeply flawed, however. In this paper, we demonstrate that the central confusion in such analyses is to conflate a *wireless use model*, a given way of providing services, with *a property regime*, a legal system that determines how ownership rights are defined and distributed. In this mode, the value of spectrum used in one way is incorrectly compared with the relevant options; relevant opportunity costs, correct allocation margins, and the associated regulatory costs – including those ensuing from the creation of anti-commons tragedies – are simply ignored. In this manner, additional spectrum allocations are justified merely on the observed existence of productive use. That the opportunities foreclosed by this use may have greater value, on some margin, or that existing services could more efficiently supplied with other spectrum rights, must be considered. So, too, must reductions in consumer welfare imposed by the case-by-case allocation system that is recommended to impose the recommended unlicensed bandwidth, given that lower-cost regulatory alternatives are available.

The evolution of wireless markets is, indeed, proceeding at rapid pace. Much can be learned from the market deployments now available for observation. For decades, regulators have over-allocated spectrum to this service on the grounds that the benefits of television programming were deserving of categorical support via large spectrum set asides. The errors in this approach are now widely acknowledged, even by the agency that created them (FCC 2002a, 2010).

In general, spectrum allocations cannot be evaluated solely according to whether a

---

[4] Soon to be joined by Google. The search giant announced a deal in early 2015 to lease capacity from Sprint and T-Mobile networks for WWAN service, while relying heavily on wi-fi access to fixed networks for a mobile consumer service to debut in future months.

given set of applications accessing the allocations generates social value. To calculate the value of the marginal product, the relevant metric, much more is required, including an accounting as to whether the relevant allocations:

(1) support the most socially valuable services
    -- given possible input substitutions and spectrum opportunity costs
(2) when focusing on the appropriate quantity margins – more or less MHz –
(3) and recognizing future opportunities that will require costly reconfigurations
(4) and considering the overhead of the administrative structure required.

This simple methodology can be called a *Spectrum Allocation Valuation Grid.*

Contemporary policy debate, following historic "public interest" allocations, tend to treat (1) in a casual, ad hoc manner, and then to skip (2) through (4). Rather than rely on a thorough cost-benefit analysis, regulators then place great emphasis on categorical arguments claiming that a particular use of spectrum is highly productive and should be, therefore, allocated more bandwidth. That conclusion is easy to argue provided that political support for the position is forthcoming. But the conclusions rendered may be only weakly supported – or convincingly rejected -- by marketplace evidence.

Perhaps the results of this truncated analytical approach are best seen in the allocation of VHF/UHF bandwidth for terrestrial broadcast television. This illustration is especially useful, given the well-developed consensus that, while television broadcast content is highly valuable, the use of 49 digital TV channels (294 MHz) to distribute it via terrestrial broadcast is not. This demonstrates the important distinction between the *value of a wireless application* and the *value of spectrum set aside to supply that applicatio*n – as seen in this context, if not in others.

But the bulk of our article deals with those others – specifically, the problematic use of a truncated policy template applied to the question of licensed v. unlicensed allocations.[5] The decision to reallocate television "white spaces," unallocated bandwidth in the FCC's TV Band, is an example of the reasoning we critique. There exist multiple ways that white spaces, long used as buffers to limit interference between TV broadcast stations, may potentially perform more socially valuable functions. One path is to permit unlicensed devices to access particular channels, setting power limits and other rules to as to protect existing TV stations against "harmful interference." An alternative is to have regulators reassign broadcast channels, packing TV stations more densely in spectrum space,[6] and then auction licenses allocated the marginal spectrum "cleared" by the

---

[5] Licensed spectrum is allocated, broadly speaking, in two ways: traditional licenses, which narrowly define the permitted uses of spectrum, and liberal licenses, which do not. See, e.g., Hazlett & Spitzer (2006). The former are notably inefficient, constraining competition and blocking the use of new technologies (by forcing innovative deployments to obtain new regulatory permissions). Unless otherwise noted, this article will refer to liberal licenses when discussing the choice between licensed and unlicensed spectrum allocations.

[6] A given set of broadcasters can reduce their bandwidth occupancy by various means, including co-location of transmit facilities and shared broadcasts (with one digital TV broadcast stream of 19.4 mbps

process. This is what the FCC did, 2002-2008, in reallocating the "digital dividend" (bandwidth made available by shuttering analog stations) to liberal licenses, which it then assigned by competitive bidding. Yet a third way is to auction "overlay licenses" for TV Band spectrum, encumbered to respect the broadcast rights of incumbent TV licensees. These would entitle new spectrum rights holders to use white spaces as economic circumstances dictate, while granting them secondary rights to use incumbents' bandwidth. Overlay owners and incumbents are then enabled to trade, with newcomers paying incumbents to shift allocations to new services. Such transitions have been achieved in secondary license markets, and with overlays issued in personal communications services (Cramton & et al., 1998), advanced wireless services , and 700 MHz licenses (Hazlett 2008a, b).

The approach in the FCC's white spaces proceeding, however, did not seriously consider such alternatives, or any others. Its decision-making process is instructive.

First, it made a categorical claim favoring a given regulatory model based on the popularity of certain wireless applications. "The Commission's rules for unlicensed transmitters have been a tremendous success" (FCC 2002, par. 6). The agency pointed to "cordless telephones, home security systems, electronic toys, anti-pilfering and inventory control systems and computer local area networks" developed for use in the 2.4 GHz ISM (Industrial, Scientific and Medical) band, where unlicensed devices are authorized.

Yet, other bands operating under different rules have also proven "a tremendous success." And some unlicensed allocations have not proven so conducive to wireless activity, including U-PCS (in the 1.9 GHz band), the 3650 MHz allocation, and now the TV Band white spaces, as discussed below. To claim that one set of wireless services implies that the regulatory system supporting them is superior to alternatives would seem to require that the alternatives be considered and the costs and benefits of the rival approaches examined. These would necessarily include the costs of the regulatory process itself, as well as adjustment costs in future years when new services, business models or technologies render existing modes of operation – and, perhaps with them, particular spectrum allocations -- relatively inefficient.

Yet, instead of such a regulatory analysis, the approach – common far beyond the white spaces proceeding – chose a much narrower path. For the most part, it simply considered whether unlicensed devices could be accommodated in the white spaces, and ignored all but one other opportunity: the status quo. In this vein, the FCC found that "multiple vacant channels are generally available" and "Allowing unlicensed devices to operate on TV channels that are not being used in a particular area would be a more efficient use of the spectrum" (FCC 2002, par. 14). Only if the extant allocation is the one relevant alternative is this true, and there is no logical reason to truncate the world's many choices in such fashion.

---

hosting two or more rival stations programming), by reducing the power emitted by stations, or by using alternative delivery platforms (such as cable, satellite, or broadband).

The FCC evinced some understanding of this in its rule making. In passing, it suggested that comparisons were in order. In cursory fashion, it presented a case for unlicensed allocations, given the asserted (and categorical) disutility of licensed spectrum use in a compact analysis given here in whole:

> Unlicensed use of this spectrum as opposed to licensed use appears to be appropriate because the operating power levels of unlicensed devices are generally lower than the power levels used in commercial mobile radio services, making it easier for unlicensed devices to identify and operate on unused frequencies without causing interference to authorized services. Further, the frequencies and amount of unused TV spectrum vary from location to location and could change over time as TV stations or other authorized services are added or change frequency, potentially complicating the licensing of commercial services in unused TV spectrum. We note also that the unlicensed uses we identify in this NOI are not intended to limit future licensed use or to guarantee spectrum access rights for this band (2002, par. 14; footnote omitted).

This dismissal of licensed spectrum use is ad hoc; in legal language, arbitrary and capricious. (This methodology was not expanded or remedied in subsequent FCC Notices.) First, the Commission assumes that licensed used must use higher power levels than unlicensed, when in fact the use of licensed frequencies often features emission levels markedly lower than unlicensed. Small cells using licensed spectrum (pico or femtocells) routinely operate at just a fraction of the power levels used in wifi radios (Cisco 2009). Second, the idea that different white space allocations, market to market, is a problem for licensed use simple ignores the fact that it is also a problem for unlicensed use. It is an empirical question as to which approach will best solve the standardization problem when the FCC is unable to allocate spectrum uniformly. Third, the FCC claims that it can issue unlicensed use rights without compromising any future licensed use – when the claim is falsified by the Commission’s extant determination. In choosing to reject new licensed spectrum rights in favor of unlicensed white spaces, it has already limited current and future licensed use.

More basically, however, the Commission fails to consider how market-based bargaining might create efficiencies. In particular, overlay owners could induce TV stations to alter frequency use, opening spectrum for other productive employments. There is no reason to assume, as does the FCC implicitly, that the only way to share the TV Band is to protect existing broadcasting. This configuration, after all, constrains other uses – licenses or unlicensed – with respect to available bandwidth and power emissions, while imposing overhead on devices (as with listen-before-talk or database-checking routines). If new TV band users find that TV station owners will accept payments to curtail their emissions that are less than the benefits thereby gained, markets may achieve efficiencies by constraining not the entrants’ but the incumbents’ operations.

This bargaining is deterred with non-exclusive spectrum rights supplied by the government for unlicensed allocations; rights there are highly fragmented and free rider problems (in arranging bargains with incumbents) dominate. Analytically, this forces a consideration of exclusive rights into the spectrum allocation policy determination, in order to consider identify the costs and benefits of rival choices. The policy options in the TV White Spaces proceeding, however, were artificially truncated: potentially beneficial choices were simply ignored. This is not an aberration. As we will show, the standard template used to analyze the licensed v. unlicensed allocation question is seriously flawed. When the appropriate factors are accounted for, we show in this paper that:

> A) a liberal exclusive rights regime in spectrum is socially beneficial whenever the benefits derived from defining and enforcing such rights exceed costs (this statement is a truism, but it is useful – and moves policy – to set it as the default);
>
> B) for WWANs, benefits are easily in excess of costs, and mobile network creation and usage suggest that exclusive rights support consumer welfare gains far exceeding those flowing from alternative rights structures;
>
> C) for WLANs or PANs, extremely localized services, private rights are less productive (as spectrum use rights default to real property owners) but some exclusions – power limits and technical restraints – are highly useful, requiring either regulators or spectrum owners to set rules;
>
> D) exclusive, flexible-use rights support deployments associated today with both licensed and unlicensed applications;
>
> E) these rights can flow from generic templates, reducing administrative overhead by obviating the need for specific regulatory determinations, thus mitigating "tragedy of the anti-commons."

II. THE VALUE OF THE WIRELESS APPLICATION IS NOT THE VALUE OF THE SPECTRUM

The U.S. 600 MHz Band is being re-organized. The Federal Communications Commission has determined that continued use of the frequencies for terrestrial television broadcasting is no longer in the public interest. In 2010, it announced its intention to reallocate up to 120 MHz of the 294 MHz TV Band, creating new licenses to support, among other things, mobile voice and data services (FCC 2010). In engineering this shift, the Commission considers the social benefits of moving from an allocation supporting one industry, broadcast TV, to one supporting another, mobile network services. Hence, the Commission proposed its reallocation by citing the strong demand for additional spectral inputs in wireless phone networks, the so-called "spectrum crunch" or "mobile data tsunami."

This spectrum allocation system is not free to operate. Regulatory control of access to airwaves, while potentially preventing certain "tragedies of the commons,"[7] can

---

[7] The term refers to the inefficient dissipation of a resource due to insufficient exclusion. It was coined in Hardin (1968). But "tragedy of open access" would constitute the better description, as resources owned

impose delays or rigidities that reduce competition and innovation in wireless markets. In particular, the top-down structure of decision-making distorts market outcomes. Moreover, given the transaction costs, including free rider problems, necessary to correct resulting misallocations, bureaucratic processes tend to lock-in old technologies, service models, and market structures.

This is not an outsider's critique; it is the argument the FCC used in finding that the TV Band was inefficiently depriving growing markets of spectrum. To correct the problem it would be necessary to employ "market-based mechanisms" departing from standard administrative rule makings. The two-sided auction was introduced to allow actors with superior information as to spectrum values to guide the reallocation process (FCC 2010; Hazlett 2014).[8] Its key innovation allows the government to *buy back licenses it has issued to broadcasters*, implicitly rejecting the use of administrative powers to unilaterally change spectrum use.

The conclusion rendered by regulators is notable not only in exposing frailties in the FCC's administrative process, but in evaluating the analysis that created and yet protects the TV Band spectrum. Television broadcasting has been allocated overly generous bandwidth – and the allocation long protected from encroachment[9] -- on the logic that

** broadcast TV programs are highly valuable*
** therefore, the spectrum allocated to broadcast TV is highly valuable.*

In its National Broadband Plan, the FCC defied this empty economic logic. Broadcast video's popularity could not alone establish terrestrial TV broadcasting, on the currently sized band, an efficient spectrum use. Predictably, broadcast licensees opposed the reallocation of TV Band spectrum, advancing traditional arguments that conflate application value with the value of an enabling spectrum allocation. Industry studies are produced to build on the fact that more than $20 billion in TV receivers (now, flat screen panels) are sold in the United States annually,[10] while TV broadcasters generate more than $40 billion in advertising sales.[11] The economic impact is claimed as:

---

by a group (in "common") may be utilized without over-dissipation. Michael Heller (1998) edited the phrase to describe a situation where ownership rights were, in theory, sufficient to exclude over-dissipation, but too fragmented to be of practical value. The concepts are closely related, if not identical. See Fennell (2004).

[8] Studies by economists long ago concluded that television broadcast frequencies were over-allocated (e.g., Kwerel & Williams 1992; Spiller & Cardilli 1999; White 2000; Hazlett 2001a, b; Faulhaber & Farber 2003). Tech visionaries have been arguing the case for about as long (Gilder 1990; Negroponte 1995, 24).

[9] Not all encroachment on the TV Band has been resisted. Originally, some 82 channels were set aside for broadcasting on 6 MHz TV channels (492 MHz) in the 1939-1953 period. Fourteen channels (70-83), or 84 MHz was stripped from broadcasting in the 1970s, some of which was used for cellular. In the 700 MHz auctions, held 2002-2008, 18 channels (52-69), or 108 MHz, was reallocated to adjust for the analog TV switch-off following the digital TV transition. In 2015, 49 channels (294 MHz) remain allocated to broadcast TV.

[10] CITE XXX

[11] CITE XXX

- “$1.24 trillion of Gross Domestic Product originat[es] in the commercial local radio and television industry annually [and]

- 2.65 million jobs [are] attributable to the local radio and television industry on an annual basis.”[12]

Moreover, TV stations are said to provide important public service, including free over-the-air programs that extend information to voters about elections and candidates; educational shows that provide learning opportunities for children; and emergency alerts during natural (or other) disasters.[13] These spectrum allocation justifications are transparently protectionist. For many years the great majority of public affairs programming, including the airing of presidential debates, has occurred on cable networks such as CNN and C-SPAN (Hazlett 2000). Similarly, emergency warnings, from tornado watches to Amber Alerts, are distributed more efficiently and targeted more effectively via mobile texting and other push technologies. Yet, such arguments have been sufficient to keep the TV Band locked into place for decades, and will continue to protect significant set-asides (at least about sixty percent of the current band) even after the FCC’s Incentive Auction.

The point here is to observe the structure of this uncompelling policy argument, what what we shall call the *TV Spectrum Valuation Fallacy*. TV broadcasters do indeed have FCC licenses that permit certain wireless activities. And TV broadcasters also produce and distribute video programs of great popularity. But the spectrum rights issued, while connected to the outputs, cannot be valued solely on that basis.

Regulators reveal that they understand something about that separation in deciding to push back the margins of the TV Band allocation from 294 MHz to, as roughly planned in the Incentive Auction, 174 MHz. In resisting this reform in 2009 Comments to the National Broadband Plan, the broadcast industry articulated the position that such actions were unwarranted because:

> Various spectrum re-allocation proposals would undermine local broadcasters’ ability to invest in local news operations and other existing services, and they would prevent them from launching new services that would expand the benefits they provide to the public and help defray their sunk costs. The combined effect of these harms ultimately would threaten the fragile viability of the country’s broadcast service in a difficult economic environment where challenging long-term trends are likely to intensify (NAB 2009, 12).

---

[12] *Local Broadcasting: An Engine for Economic Growth*, Woods & Poole Economics (2014), 1. See also, National Association of Broadcasters, *Study Finds Local Broadcasting Generates $1.24 Trillion in Economic Activity Annually; Report finds 2.65 million jobs attributable to commercial broadcast industry; 500 broadcasters to deliver state-specifc [sic] economic reports to Members of Congress*, Press Release (Feb. 24, 2014).

[13] CITE XXX

And the NAB further noted, “Broadcasters are rolling out mobile DTV services” (Id.). This was not just another pretty commercial product, but a matter of life and death: “Mobile DTV will dramatically expand the distribution of emergency information” (Id). In fact, the broadcasters noted – correctly – that the mobile data tsunami, the premise of the FCC’s reasoning in making more spectrum available for mobile networks, was itself driven by emerging demands for mobile video. “’TV’ is driving great demand for spectrum access, but using new models” (Id.). And they forecast that broadcast licenses, occupying the existing TV Band spectrum allocation, would most efficiently meet this demand. Mobile TV will supply “desirable, popular programming on a mobile basis, and it will do so on a spectrally efficient, point-to-multipoint basis” (Ibid., 13).

As the facts stand, video has been largely responsible for the extremely rapid growth in mobile network traffic loads (Cisco 2015). But the spectrum allocated to terrestrial broadcast TV stations has played virtually no role in this. A 2011 tech press headline suggests a more plausible causation: “The Mobile Tsunami Is Near: Blame Netflix & Apple.”[14] Regulators embraced this view, concluding that additional bandwidth made available to mobile networks would provide the requisite patch. Moreover, the bandwidth could be shifted from the TV Band where spectrum is over-supplied given the existence of alternative video delivery platforms -- including cable, satellite, fixed broadband and mobile broadband – to offer low-cost substitute service, and where digital TV stations can be packed together much more tightly (with many stations sharing 6 MHz channels) without substantial loss of video service. As estimated (Hazlett 2014), replacing the entire terrestrial TV broadcasting system with a nationwide satellite service distributing “free, over-the-air” (OTA) broadcasting services would cost no more than $3 billion.

Insofar as public safety is concerned, the emergency warnings issued by TV station broadcasts, while valuable, have been almost wholly eclipsed by a push technology: mobile texting. This sends messages to mobile device users in given geographic areas, 24/7/365, alerting even those not watching broadcast TV. At any given time of the day or night, this is a minority of the total population, often a tiny fraction. In contrast, mobile phone subscribers tend to keep their devices with them, and network-connected, through the day (and the bedside at night). In this sense, extreme weather warnings or Amber Alerts are more transmitted via applications not using TV licenses.

The FCC’s 2010 Incentive Auction initiative not only rejected the broadcasters’ policy position but the thinking it was based on. The Commission did not argue that broadcast TV services were not valued, but incorporated additional valuation criteria. If reflected upon spectral input substitution options, considered bandwidth opportunity costs, and focused on appropriate increments over categorical alternatives. A determination was made – to be adjusted using competitive bidding data – that will move marginal TV Band frequencies into Commercial Mobile Radio Service (CMRS)

---

[14] Stacey Higginbotham, *The Mobile Tsunami Is Near: Blame Netflix & Apple*, GIGAOM (Jan. 11, 2011).

allocations. New licenses will authorize liberal spectrum access, such that are free to obtain such rights and deploy services such as warranted by economic conditions.

It is key that the information supplied in the Incentive Auction helps to drive the FCC's reallocation. Instead of simply determining that the TV Band, being too large, simply transfer some arbitrary increment of bandwidth to mobile services, regulators will allow markets to help determine the outcome. Only if bidders in the forward auction (where CMRS licenses are offered for sale) pay a sum equal to or exceeding the stations participating in the reverse auction (where TV licenses are offered for sale to the government) will the target level of MHz established by the FCC be transferred. If forward bids fail to reach the dollar levels demanded by the stations needed to sell in order to achieve a given reallocation, the FCC will drop its MHz target and start the Incentive Auction over again. This repeats until forward auction bids exceed reverse auction offer prices. While key elements of the auction design have been criticized (Hazlett 2014), the point here is that the Commission chose to incorporate key market-based information in the reallocation process on the (correct) presumption that markets generally incorporate better data about relative resource valuations than can be discerned by government regulators acting administratively (i.e., without competitive bidding).

We attempt to briefly summarize the lessons learned from the FCC's implicit rejection of the *Broadcast TV Spectrum Valuation Fallacy* in the Incentive Auction policy formulation in Table 1.

TABLE 1. SPECTRUM ALLOCATION VALUATION GRID: THE CASE OF THE FCC'S TERRESTRIAL BROADCAST TV 2010 REFORM

TV Band Regulatory Determination	*FCC Conclusion*	*FCC Solution*	*Comment*
Hosts valuable services?	Yes.	To yet look for gains from spectrum reallocation	This is progress.
Opportunity cost of spectrum?	Yes, particularly given demands for additional capacity in cellular networks.	Reallocate MHz from TV Band to mobile licenses.	This breaks asserted identity between app value and spectrum value.
Substitute delivery systems?	TV stations can exit, using cable, satellite, broadband; tighter packing possible	FCC to buy TV licenses in reverse auction, repack TV Band, reallocate frequencies to mobile licenses for auction.	Ditto.
Appropriate margin?	294 MHz excessive	174 MHz better	Regulatory fiat informed by Incentive Auction bids
Low cost future spectrum reconfigurations?	No consideration.	Liberal licenses will allow; protected TV Band will not	
Consideration of regulatory process?	Yes. Long reallocation delays documented.	Instituting complex two-sided Incentive Auction.	Reallocation process to stretch years beyond the 5-year proposal.

III. ERRORS IN CATEGORICAL CLAIMS FOR A SHIFT TO UNLICENSED ALLOCATIONS

In their 1992 FCC paper, economist Evan Kwerel and engineer John Williams considered the following question: *What if one television station in Los Angeles, California were permitted to use the spectrum allocated to its license to provide cellular phone service instead of broadcasting?* Their exercise:

- Estimated the loss in value to viewers from the subtraction of one TV channel.
- Estimated the additional value created for mobile subscribers given an additional 18 MHz of bandwidth in Los Angeles (the elimination of one TV station broadcast could allow the use of not only the 6 MHz allocated its license but the bandwidth allocated adjacent licenses); gains from entry in the mobile market yielding lower retail service prices (assuming Cournot competition and plausible cost conditions).
- Contrasted the offsetting magnitudes, concluding that more than $1 billion in consumer gains would be realized, in present value terms, by the contemplated switch (Kwerel & Williams 1992; 1993).

Their results were compelling because their general methodology was rooted economic reality. They had produced persuasive evidence that the value of the marginal product for specific TV band spectrum was higher in one allocation versus another. While the analysis did not include an examination of other possible employments beyond cellular, the policy recommendation – that licenses be issued that allowed "flexible use" – would not have blocked those other possible uses. Having shown that just one possible non-TV use would increase social value, proof of concept supporting a market-based reallocation was presented.

Kwerel-Williams (1992) also incorporated the possibilities for input substitution. By using Cournot assumptions for how an additional allocation of spectrum could facilitate entry into a well-defined (cellular) market, it implicitly steered their analysis to consider output changes when all efficient possibilities for substitution had been exhausted.

One important margin not addressed directly by the 1992 paper were administrative costs. The idea that allowing licensees to reallocate the spectrum set aside for their licensees might result in additional costly processes could offset perceived gains from the immediate changes in wireless applications. Yet, these considerations were dealt with in follow-up research (Kwerel & Williams 1993; 2002), where it was argued that generic rules permitting market reallocations were efficient mechanisms for promoting competition and innovation, given the alternative of government rule makings.

Overall, the template for analysis is an important first step in understanding how arguments over spectrum allocation should proceed. In some of what has followed, this has been the basic format (Hazlett & Munoz 2009; Hazlett, Munoz & Avanzini 2012). In much of the debate of unlicensed v. licensed allocations, it has not, however. Here we describe some of the problematic arguments made.

a) *Unlicensed spectrum is uniquely worth billions annually to the U.S. economy.*

Thanki (2009), in a consulting report commissioned by Microsoft, produces a relatively careful study attempting to measure consumer surplus associated with unlicensed band uses, focusing on the marginal gains yielded in three specific applications:

-- residential wi-fi networks
-- hospital WLANs
-- radio frequency identity chips used in the clothing industry

The paper attempts to give estimates of the consumer value added by three estimates (low, medium and high) for these three application categories. They summarize their findings by aggregating their three mid-point valuations, which suggest that the applications generate, as of 2009, about $25.4 billion annually.

TABLE 2. VALUATIONS FOR UNLICENSED DEVICES (THANKI 2009, FIG. 27)

Scenarios (2009 – 2025)	Low	Medium	High
Economic value generated by home Wi-Fi ($ billions per year)	4.3	8.4	12.6
Economic value generated by hospital Wi-Fi ($ billions per year)	9.6	12.9	16.1
Economic value generated by clothing RFID ($ billions per year)	2.0	4.1	8.1
SUM OF ANNUAL ECONOMIC VALUE ($ billions per year)	**16.0**	**25.4**	**36.8**

Source: Perspective analysis

The paper contains the proviso that unlicensed bandwidth supports many more applications, and argues that excluding them from the quantification exercise biases the valuation estimate downwards. Conceding that values for licensed spectrum use in the U.S. are “an order of magnitude greater than the combined value created by the three unlicensed uses we have assessed” (Thanki 2009, 35), the paper attributes the gaping difference to the fact that

> [T]he two numbers cannot be directly compared. Whereas the $200 billion figure encompasses the vast majority of licensed usage of spectrum, our estimate of $16-37 billion only looks at the small part of the increasingly large unlicensed ecosystem… (Id.)[15]

Yet this asymmetric caveat tilts the analysis. Consumer surplus associated with

---

[15] Curiously, the paper notes that the “total economic value generated by licensed uses of the spectrum is substantial… around $250 billion per year” (Thanki 2009, 22). This may be a reference to the 2008 forecast for 2010 consumer surplus in the U.S. cellular market of $263 (Entner 2008).

U.S. cellular service, using first-half 2009 data, has been estimated by Hazlett, Muñoz & Avazini (2012), at about $212 billion per year. The estimate relies on actual consumer demands, revealed by market purchases, highly likely to be more reliable than the inferences made from the indirect evidence of consumer gain used in studies of unlicensed spectrum. Yet, because of the historical nature of the price and quantity (minutes of use) data, the aggregate value forecast clearly and substantially under-estimates total social surplus. As the authors noted in presenting the evidence, the "historic price-quantity pairs [were] lower- bound estimates of current conditions with respect to consumer demand" (Hazlett et al. 2009, 99). The method yielded "conservative forecasts of the value currently delivered by wireless network services in the United States" (Id.).

Further limiting the licensed valuations are the exclusion of a wide variety of popular services accessing licensed spectrum, such as satellite television. DBS (direct broadcast satellite) not only serves video to over 30 million U.S. households, generating about $22 billion in annual revenues, but limits pricing by cable and telco video providers, accounting for billions more in pure consumer surplus. Other excluded applications offered by networks using licensed spectrum include machine-to-machine (M2M) applications, including tracking devices, RFID applications (as with unlicensed spectrum), On-Star (emergency call service) and Kindle. In addition, the "increasingly large *licensed* ecosystem" includes gains emanating from mobile network devices, such as the Apple App Store or Google Play. Indeed, massive investment in network infrastructure – for handsets, base stations and backhaul (linking cell sites to global voice and data networks) – provides additional benefits not fully captured from cellular (voice and text) service revenues. Smartphone sales alone, increasing from under $12 billion in 2009 to a projected $51 billion in 2015 (see Figure 1), generate substantial additional (and off-net) benefits that complement, and have in large measure been driven, by investments to enhance to the utility of licensed spectrum. Capital spending to expand and densify mobile networks, expanding the infrastructure platform, amounted to $34.397 billion in 2013 – nearly $300 per household per year.[16] (The capital expenditures are accounted for, if much less than fully, in separate calculations of mobile consumer surplus. Duplication of value estimates is, indeed, embedded in Thanki's suggestion that home wifi values are additive with other services, including mobile data offload, that largely flow through the home networks already accounted for.) When the magnitude of these economic activities are symmetrically included in the comparison, it is highly probable that valuations for licensed spectrum continue to dominate those for unlicensed.

---

[16] Roger Entner, *Every Way You Look At It: US Carriers Spend More in Capex Than Their EU Peers*, Recon Analytics (June 9, 2014).

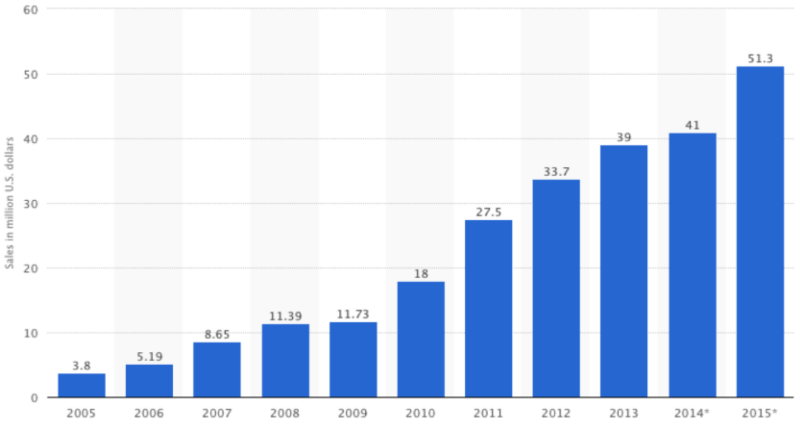


FIG. 1. U.S. SMARTPHONE SALES, 2005-2015 (SOURCE: STATISTICA)

Despite the lack of value comparisons, and the complete lack of marginal spectrum valuation, the purpose of the Thanki (2009) study, of course, was ostensibly to do exactly that. This is revealed in its stated purpose, when the study sought to present the flip side to the view (attributed to Rysavy 2008) that:

The overall value from unlicensed usage of spectrum is an insignificant fraction of that generated by licensed usage

Therefore, further licensed allocations of spectrum, especially at lower frequencies, will create the most economic value[17]

The findings about the value of applications using unlicensed spectrum is thereby undertaken, with the conclusion rendered that the services supplied are not “an insignificant fraction of that generated by licensed usage.” In this, the study aptly illustrates the *Broadcast TV Spectrum Valuation Fallacy*, ignoring input substitution and relevant spectrum margins, while offering the value of certain wireless services as unadjusted proxies for the value of the spectrum allocation they are seen to utilize.

Nonetheless, this study has been successful in influencing the policy debate. MLE (2011, 2), for instance, rely on its findings to open their argument that “[t]here is considerable evidence that unlicensed spectrum has huge economic value.” The authors go on to argue for additional unlicensed allocations based on this unilateral claim. Nowhere does the Milgrom paper attempt, let alone establish, that marginal spectrum allocations are valued more highly under one regime than the other.

---

[17] Thanki 2009, 20.

b) The widespread use of wifi for mobile data offloads presages a paradigm shift, with markets abandoning licensed spectrum, embracing unlicensed spectrum.

Given the decided valuation superiority given to cellular network services, Yochai Benkler argues that the service being delivered over cellular networks is really not relying on licensed spectrum. He writes that

> customers who buy wireless data service from Verizon or AT&T are not getting their service delivered exclusively over licensed spectrum. If 92% of data to tablets and 42% of data to handsets is delivered over Wi-Fi, and customers pay for carriage of bits, not for "use of spectrum," a more reasonable approach would be to take the money customers pay for mobile data carriage and equipment and apportion it based on the amount of traffic carried (Benkler 2012, 99; footnote omitted).

This approach follows Milgrom et al., who offer the view that if half of U.S. mobile network data traffic is offloaded to fixed networks via WLANs, then half the value of the service should be attributed to wifi rather than the WWAN.[18]

It is a dubious economic analysis to assign values not based on demand but on physical attributes. The analogy would be to attribute 95% of the value of a diamond ring to the band on which it is set, based on the fact that the band accounts for that proportion of its weight. The problem in wireless is seen immediately when considering, for instance, the current advertisements for wifi in Buicks. The car comes with factory-installed WWAN connectivity, which is then distributed to radios used in the moving vehicle via a WLAN. Since all the traffic will flow through both the WWAN and the WLAN, the contribution to value is given (under the Milgrom-Benkler rule) as fifty-fifty, despite the fact that the one opportunity (WWAN connectivity) creates the opportunity of interest. A WWAN connected car without wifi could yet generate mobile network access (and for multiple devices using wires or pico cells) while a WLAN connected car without WWAN access does not.

Here it is straightforward to see why the handoff from the WWAN to the WLAN is called *offload*. Demand generated by one network, and the ecosystem thereby created (including services, portable devices and apps), preserves its capacity by distributing traffic – where fixed networks are fixed nearby – to another. The network connection of the device used in a stationary location is typically less challenging; corded connections have supplied these links for decades. The point of wire*less* transmissions is, in fact, to reduce reliance on these *wires*. So, in constructing an asserted proxy for valuing the service, it is arbitrary and capricious to consider a transport link sending a bit equal to

---

[18] "A ballpark estimate for the value created by 3G data transmission might be the data-related revenue of mobile carriers; in 2010, this exceeded $50B. If mobile phone users use Wi-Fi to transfer roughly half this amount of data, it suggests an annual economic value being created in excess of $25B" (MLE 2011, 18).

any other.[19]

When the key variable, distance, is introduced into the offload measurement, the value proxy for wifi collapses. If the average wifi data offload at a residential location or Starbucks hotspot is fifty feet, and the average mobile data transmission through the cellular network links through a base station one mile from the device, a 50-50 mix of the traffic load turns into a *99-1 distance-weighted* mix in favor of the cellular network.

We are not suggesting that this is a good way to approximate the relative contributions of licensed and unlicensed spectrum in delivering mobile services. Quite the reverse: we show how tenuous are the measures produced by others, and how dependent they are on unwarranted assumptions. The economic valuation question must, to be useful, focus on the marginal value of spectrum rights, and include administrative costs and benefits as well. That those valuations are difficult to infer from physical or technical metrics observed by regulators is why economists have long argued for market mechanisms. It is a transactional advantage – meaning that allocational choices tend to be better, being better informed and more easily subject to efficient adjustment – that exclusive rights to spectrum access can (and have been, in limited cases) defined, packaged and sold by governments to responsible economic agents. These agents tend err less in the construction of valuation metrics and be corrected more quickly, given strong incentives to profit by using bandwidth most productively.

This key aspect of efficiency is illustrated in a recent example. With liberal licenses auctioned by the FCC, 2002-2008, Qualcomm initiated a mobile television service in the United States. Using licenses allocated the UHF spectrum formerly set aside for TV broadcasting (Channel 55), the company launched MediaFlo in 2007, a service delivering 15-20 video channels (including CNN, CBS, CNBC, Fox News and Comedy Central) for subscribers. Reception was integrated into certain handsets sold by AT&T and Verizon, or on stand-alone receivers. Qualcomm spent about $1 billion on the foray, including technology development, network broadcast facilities, and FCC licenses.

By October 2010, however, the company found the innovation unprofitable. It was quickly undone. In particular, the FCC licenses were sold to AT&T, which acquired

---

[19] Even this proxy rule is calculated asymmetrically. If the value supplied by $50 billion in expenditures in mobile data service is attributed $25 billion to wifi (and then to unlicensed spectrum) on the grounds that half the bits in question will travel via wifi, the calculation omits many other transmission paths involved. The wifi router delivers data to and from a broadband network, encompassing both the last mile facility supplied by an ISP and the longer links forming the Internet's network of networks. Cellular base stations, in turn, pass data to backhaul transport facilities that connect to still larger voice and data networks. While cellular operators organize and largely construct these latter, the wifi offload to the fixed network is inexpensive precisely because so little additional investment is made. It is also true that much of the offload of data from mobile devices is made to fixed networks, as well as the cloud, via wires – USB power cords that both recharge and sync Apple devices, for example. While it is impossible to identify the "correct" calculation that reflects the marginal gains supplied by the associated networks, it is dubious to assert that wifi deserves a valuation reflecting the flow of traffic through just two of many more networks.

them in December 2011 (after a year-long FCC license transfer approval), and the underlying bandwidth repurposed to supply capacity for the wireless network's 4G system, supporting "our goal of expanding mobile broadband deployment throughout the country" in the words of the FCC.[20]

As demands for cellular data were growing rapidly, the new deployment was devoted largely to accommodating those demands, clearly driven by the increasing popularity of mobile *video*. Hence, the visible separation of application from the spectrum allocation over which it flows. TV programs provided by terrestrial TV systems broadcasting on the TV Band were transmitted over the same spectrum by MediaFlo and are now distributed via mobile broadband networks. But because liberal, exclusive spectrum rights replaced traditional, rigidly defined broadcast license, a smooth migration to a market generating greater social product has occurred. That the flexible use rights, owned by a profit-maximizing agent, remain ready to be deployed differently as opportunities evolve fortifies the efficiencies yet demonstrated.

We note a key analytical point: the transitional process described in the *Broadcast TV to Media Flo to 4G* spectrum reallocation is not well summarized by comparing the popular use of the resulting applications. Indeed, in important respects, the basic video application was unchanged. In another sense, "usage" may have declined in the switch-over from broadcasting to mobile TV. Yet, because market competitors were able to seek new efficiencies in market organization consumer welfare was enhanced. While there exists supporting evidence that 4G generates higher social product than did the frequencies allocated Qualcomm's licenses previously, the key evidence is found in the market for frequencies. Qualcomm outbid others for access rights and deploy them to supply the innovation of their choice. In the event, the new product flopped, but the effort was productive in:

- transitioning away from a relatively low-valued use (TV broadcasts);
- testing a possible beneficial innovation (mobile TV);
- promptly forcing a reconciliation of the innovation's costs and market prospects, displacing one service with another of evidently higher value:
- producing a market test to gauge, and reveal, efficiencies (if broadcast TV or MediaFlo were the superior social opportunity, AT&T would not likely have outbid others for the spectrum rights in question); and
- avoiding administrative hold-ups that long delay such transitions in centralized allocation decision-making.

This search for efficiency in inputs and outputs is the empirical reality suggested by the argument made for spectrum markets since Leo Herzel (1950, 1951) and Ronald Coase (1959) stated the case. These transactional characteristics foster value discovery, seizing upon dispersed information known to private agents but not to central

---

[20] As quoted in Eric Engerman & Chris Strohm, *AT&T's $1.93 Billion Qualcomm Airwave Purchase Wins Approval*, BLOOMBERG BUSINESS (Dec. 23, 2011).

administrators, and offer institutional advantages over government-directed processes riddled with externality problems. In particular, because neither errors nor delays redound to the financial harm of government spectrum allocators, incentives for efficient performance are relatively weak. The FCC has itself confronted the problem, documenting that standard spectrum allocations (for licensed spectrum) take 6-13 years, imposing large (net) social costs, and arguing that "market-based mechanisms" be created to improve performance (FCC 2010).[21] It is important to contrast such outcomes with the salubrious dynamics unleashed via liberal licenses as in the MediaFlo example. Technology investments flowed to market, unblocked; an experiment to test demand for an innovative product was performed; the opportunity costs of the spectrum utilized appeared to outweigh the net benefits generated; decision makers were keen to monitor this information; spectrum was quickly reallocated to a higher and better use. In the space of about five years, services supplied on UHF TV Channel 55 were twice liquidated and the bandwidth repurposed -- despite the 14-month FCC approval process for the Qualcomm-AT&T license sale.

We also note the contrast between this experience and what MLE (2011) offer as a recent success in unlicensed spectrum allocation: TV White Spaces. The FCC announced its intention to open this spectrum for unlicensed devices in Dec. 2002. Over the ensuing years, during which UHF Channel 55 has been transitioned twice and now hosts millions of AT&T 4G subscribers (supplementing other bandwidth), FCC-approved white space devices (WSDs) yield no signs of consumer life.[22] As of April 2015, five years after FCC rules were finalized and over twelve years since the Commission announced its intention to permit WSDs, there were only 538 fixed wireless radios registered (and thereby approved for use) in the TV Band, and not a single mobile radio.[23]

Despite this lack of output, no reconfiguration of the deployment will soon be possible. Regulators have committed to (a) protecting incumbent use, terrestrial TV broadcasting, while (b) sprinkling in unlicensed use rights in restricted areas around them. While the WSDs are burdened by restrictions and overhead requirements, as were devices accessing unlicensed PCS frequencies set-aside in the 1990s (when new radios were mandated to include listen-before-talk capability so as to protect incumbent microwave users), reallocating the bandwidth to alternative uses will prove an arduous,

---

[21] The FCC's Incentive Auction is one such suggested policy outcome, as it relies heavily on the revelation of demand by bidders in both the reverse (TV station) and forward (mobile) license auctions. See, also, Hazlett (2014).

[22] "[T]this technology hasn't seen much deployment," writes one industry source, calling WSDs "a niche within a niche." It blames the lack of progress on the late availability of industry standards and that "stakeholders are waiting to see how much TV white spaces spectrum is freed up in the upcoming broadcast spectrum incentive auction." The first asserted lag is endogenous to the unlicensed approach, as presently constituted, which requires the government to set spectrum sharing rules and procure bandwidth. The second is misleading, as the FCC has made it clear that at least about 174 MHz (of the 294 MHz TV Band) will remain available for broadcasters and WSDs post-auction. Joan Engebretson, *AT&T vs. TV White Spaces: Parsing a Blog Post*, TELECOMPETITOR (April 7, 2015).

[23] According to the Google Spectrum database (visited April 21, 2015). See also, NAB (2015), 12.

lengthy task. In U-PCS, 1915-1920 MHz was allocated in 1993.[24] After a decade wherein there was no use of the bandwidth, as no devices were approved by the FCC, regulators began to consider shifting the bandwidth to flexible use licenses in 2003.[25] The reallocation came to fruition with the auctioning of H-block licenses, allocated 10 MHz (1915-1920 MHz and 1995-2000 MHz) in February 2014 for about $1.56 billion.

Nonetheless, Milgrom et al. conclude that the unlicensed TV Band allocation is socially useful because:

> The substantial benefits associated with unlicensed spectrum today, in particular the large value attributable to ubiquitous technologies such as Wi-Fi and Bluetooth, suggest strongly that additional unlicensed spectrum would continue to contribute to social welfare (MLE 2011, 19).

This rendition of the *Broadcast TV Spectrum Valuation Fallacy* has the added advantage of revealing an unjustified leap from inframarginal performance to asserted marginal gains. A categorical "great success" judgment is transformed into a consumer welfare contribution for additional allocations, opportunity costs – and administrative process costs – ignored. The danger is clear; hype springs eternal:

> One application is "Super Wi-Fi", which is expected to increase the range of Wi-Fi by a factor of two to three as well as allow Wi-Fi to go over hills and through walls. Super Wi-Fi achieves these benefits using white spaces – the unused bands of spectrum between used channels – in the lower frequency television bands between 54 MHz and 698 MHz (MLE 2011, 19-20; footnote omitted).

Thus far, the predictions of social benefit – cheap to construct, no penalty for error – have proven overly ambitious. Lauded as "wi-fi on steroids,"[26] substantial spectrum has been reserved for the application with, as of yet, a lack of offsetting benefits. Meanwhile, were the 174 MHz (or more) reserved for broadcast TV/white spaces allocated to liberal overlay licenses, it is clear that great economic gains would ensue. TV station transmissions could be reconfigured or, more completely, moved entirely to cable, satellite, and broadband delivery platforms. Mobile carriers, bidding over $41 billion for AWS-3 licenses allocated 65 MHz in the 1.7/2.1 GHz bands in February 2015, reveal intense demand, at the margin, for exclusive bandwidth. Over and above all the opportunities afforded by unlicensed bands allocated at 900 MHz, 1.9 GHz, 2.4 GHz, 3.65 GHz, 5 GHz and for the TV Band white space devices, the AWS-3 frequency space supports very highly valued new wireless services – as per the

---

[24] *Amendment of the Commission's Rules to Establish New Personal Communications Service, Second Report and Order,* 8 FCC Rcd 7700 (1993).

[25] AWS Third Report and Order, 18 FCC Rcd 2223 (2003), ¶¶ 46-49.

[26] Nate Anderson, *WiFi on steroids? First "WhiteFi" prototypes hit testing stage now that the FCC has opened empty TV channels to wireless networking*, ARS TECHNICA (Aug. 27, 2009); *White-space puts Wi-Fi on steroids: A second wireless revolution is starting, thanks to television's switch to digital*, THE ECONOMIST (Nov. 17, 2011).

investments of firms that must pay for the privilege of developing them. It is highly likely that at least as much would be forthcoming in winning bids for liberal licenses allocated the TV Band (174 MHz, encumbered with stations), with consumer surplus estimates conservatively estimated at northwards of $500 billion, in present value terms.

Yet, to be clear, the counter argument to MLE (2011) is not that 4G is superior to wi-fi, as measured by consumer welfare estimates. The fundamental consideration is that, with liberal license rights, competitive forces can discover the most productive uses for bandwidth. Hence, even were the valuation metrics in evidence actually reversed, wireless deployments based on WLAN technologies like wi-fi would be more efficiently supplied, taking all market and administrative costs and benefits into account, over time, with demands revealed through competitive bidding for the rival allocations. MLE (2011) argue that this is impractical, that transactions costs associated with the aggregation of demands for services supplied via unlicensed bands demands are prohibitive. We turn to that discussion just below.

Before doing so, however, we note the import of the admission. That government spectrum allocation is said to be necessary for unlicensed spectrum to operate is to attach that costs of that system of "command and control" to such allocations. Alternatives are available; with liberal licenses, control over bands are delegated to institutions that operate without the administrative handicaps, transactional expense, and rent-seeking opportunities that accompany the top-down FCC spectrum allocation regime. This is a hugely important consideration in any institutional comparison of rival approaches, one that is assumed away in the MLE (2011), Benkler (2012) and other analyses (although articulated in Benjamin [2003]).

c) Auctions for unlicensed fail due to demand aggregation problems.

In an interesting 2008 FCC paper, economists Mark Bykowsky, Mark Olson, and William Sharkey created and tested an auction format designed to allocate spectrum across licensed and unlicensed regimes. Their approach allowed companies desiring access to exclusive access rights to bid against firms wanting additional bandwidth for unlicensed devices. The trick was, while letting every bidder state a price it would pay for the rights being assigned by the regulator, bids for unlicensed frequency use would be aggregated. This reflected the fact that, post auction, all such bidders would benefit from access to the unlicensed band, realizing the gain that their bids were made to capture.

Hence, if four licenses, each allocated 20 MHz of spectrum space, were offered for sale and the top four bids (say, $5 billion, $4 billion, $3 billion, $2 billion) received were from firms seeking exclusive licenses, the four firms would each win one license. *Unless*, that is, additional bids for *unlicensed* bandwidth were received that totaled more than $2 billion – say, three bids of $800 million. These bids, jointly, would win the fourth block, moving the allocated 20 MHz into an unlicensed allocation. In laboratory testing, using experimental economics techniques, the authors found that the bidding scheme was largely successful in producing efficient allocations, although some free riding emerged with respect to unlicensed bids. This problem emanated from the restrictive contractual context assumed, however, and is solved by alternative forms (as shown in consortium bidding for FCC licenses, as in SpectrumCo's successful 2006 participation in the AWS-1 auction [see, generally, Milgrom et al., 2009], or by device licensing agreements as proposed in Kwerel & Williams 2002, 31).

The FCC has yet to adopt the Bykowsky-Olson-Sharkey spectrum allocation design. One fairly strong argument against it is that it is unnecessary, and unduly complicated, for the auctioneer to aggregate bids. Instead, private bidders wishing to use liberal exclusive rights for the creation of an "unlicensed" band are free to form joint ventures or other collaborative bidding arrangements, aggregating bids according to agreements struck among the parties. Indeed, as noted, consortia have already formed to participate in FCC auctions, winning billions of dollars in licenses. Rather than redesigning auction structures, public policy could be better served by eliminating regulatory barriers to the supply of "unlicensed" bandwidth by licensees. So, for instance, build-out requirements often imposed as license terms should be relaxed. Productive deployments not involving specific network infrastructure (as specified by the FCC) would then be permissible.[27]

Yet, even if a new format is unnecessary, the idea of using auctions to allocate spectrum access rights across regimes is a very helpful contribution. The FCC analysts pinpoint the basic informational problem plaguing regulators who "designate[s] spectrum to either licensed use or unlicensed operations… through an administrative process" (BOS, 25). Noting that the European Commission has condemned that process as "neither transparent nor objective,"[28] and that a prominent British economist has deemed it "arbitrary and unsatisfactory,"[29] policy makers are forced to rely on "the reported needs of interested parties" (BOS, 26). Auctions help in

> reducing the incentive that parties have to exaggerate the value they place on a given set of license rules involves creating a market for such rules in which participants bid to have their license rule needs met (Id.).

The endorsement for market-based demand revelation could go further: The bids registered for additional rights not only discipline the claims made by rival parties seeking alternative outcomes, but define marginal values not yet revealed in infra-marginal transactions. By undoing an "the incentive that interested parties have to misrepresent their economic interests" (Ibid.), the FCC analysts attack a key distortion. Noting that truthful demand revelation invokes "incentive problems" discussed in economic theory (BOS, 2), the economists recommend competitive bidding to "substantially improve the efficiency of the licensing process and, thus, the economic benefit society receives from one of its most valuable resources" (Ibid.). Instead of having, say, Apple Computer ask the FCC for unlicensed frequencies to be set aside for

---

[27] CITE XXX

[28] Citing STUDY OF LEGAL, ECONOMIC & TECHNICAL ASPECTS OF "COLLECTIVE USE OF SPECTRUM" IN THE EUROPEAN COMMUNITY – FINAL REPORT, by Mott MacDonald Ltd., Aegis Systems Led., IDATE, Indepen Ltd, and Wik Consult (Nov. 2006), 13.

[29] Ibid., citing Professor Martin Cave (Cave 2006, 224).

local data networks in 1990, with lobbyists and bureaucrats deciding to set-aside bandwidth for an ill-fated U-PCS allocation, the process would be much improved were the company been given an opportunity to buy flexible use rights and deploy the bandwidth as desired. The latter approach internalize both the costs and benefits of the spectrum rights sought, and subjects proposed business plans to market test, rather than a political, test.

Milgrom et al. 2009 respond to the Bykowsky et al. proposal, dismissing it as unworkable due to a public goods problem with unlicensed spectrum. They note that "Auction markets work best when one can identify the relevant bidders in advance, bring them to the auction, inform them about what is for sale, and motivate them to bid" (MLE 2011, 25). But given the "diverse and evolving group of devices that use and benefit from unlicensed spectrum" – including computers, mobile phones, baby monitors and garage door openers "and the millions of consumers that use them," the set of potential "beneficiaries is too large and diverse to be identified, informed and motivated to bid…" (Ibid.).

This rationale for government set-asides of unlicensed spectrum fails on two counts. First, it is asymmetrically applied and, second, it advances a policy solution that does not remedy the "market failure" cited as justification.

The asserted lack of informed bidders among a "large and diverse" group of unidentified beneficiaries is also present in other markets where auctions work well (Milgrom 2004). It is routinely solved by the emergence of agents who specialized in gaining information about future demands and seek efficient means to aggregate and serve them, selecting inputs in competitive markets. When mobile operators acquire spectrum access rights, they act on behalf of "millions of consumers" and device makers who will benefit from the networks they help build. Sprint's PCS licenses purchased in FCC Auction No. 4, concluded in March 1995, today host valuable wireless services for tens of millions of subscribers who were entirely unknown to corporate bidders, and millions who were not yet born. When Apple sold $30 billion worth of iPhones to Sprint in 2012, the device maker likewise reaped gains from Sprint's 1995 bids, despite their having preceded the creation of the iconic smartphone by more than a decade.

Even were there a "market failure" to remedy (i.e., a lack of informed bidders), the idea that government should therefore pre-empt a competitive bidding process is a non sequitur. The information problem cited exists for government regulators as well as private firms; replacing auctions with administrative allocations does not reveal the missing information but forecloses whatever relevant data would be supplied by the demands that organized bidding would reveal. In fact, government does act as an agent for future beneficiaries in setting aside spectrum bands, however regulated. But it may do so in a competitive or monopolistic environment; in the former, the state bids against rival parties for resources, in the latter, it pre-empts all competing bids and operates unilaterally. The unilateral approach constitutes the "Gosplan" (Faulhaber & Farber 2003) so widely criticized by analysts and regulators themselves. So, to the degree that spectrum allocations are plagued by the absence of key information, eliminating competing demands *intensifies* the failure. The obvious policy solution to the "large and diverse beneficiary" problem asserted by MLE 2011, assuming its import, would be not to eliminate the auction but to have government participate in it.

The obvious example has been stated many times, albeit in confused form. The argument is given that spectrum need by allocated by the FCC for unlicensed allocations, just as public parks are set aside for non-commercial use by local, state or federal governments (FCC 2002a). The analogy of possible set-asides (one in land rights, the other in spectrum) holds. But the allocational *process* is distinct. In land, there exists no "Federal Land Commission" (Coase, Meckling & Minasian, 1995); society adopts the more efficient method for distributing use rights, assigning ownership rights by market bid. In that context, governments – aggregating future benefits for unidentified beneficiaries – compete against rival parties. Importantly, that auction process reveals key data concerning opportunity costs; governments do not acquire rights for which they are unwilling to outbid others.[30] In order to divert marginal increments of property to the provision of public parks, an explicit monetary decision must be made. This transparency facilitates the rational use of resources, offsetting valuation estimates unbounded by competing alternatives.

Whatever informational problems exist in forecasting future benefits from unlicensed (or any other) allocation of spectrum, reducing the opportunity for revelation of the relevant data does not solve the problem. When the state pre-empts competitive bidding to requisition and quarantine spectrum for a particular application, it sacrifices available information-generating mechanisms, reducing possible efficiencies aimed specifically at solving the problem stated. In fact, this reprises the original debate on government allocation of bandwidth, when FCC Chief Economist Dallas Smythe squared off against University of Chicago law student Leo Herzel in 1951-52. Herzel's arguments for competitive bidding for bandwidth, explicitly including that used for public purposes by government agencies, was so overwhelmingly compelling that it convinced Ronald Coase of the superiority of markets over administrative allocations (Coase 1993).

Market allocation of spectrum is not only compatible with the existence of unlicensed bands, it accommodates them more efficiently than do administrative set-asides. By shifting competition for unlicensed bandwidth from the FCC rule makings to FCC auctions, unlicensed advocates are led to quantify their demands while staking credible commitments in moving resources away from rival claimants. This resolves the externality problem inherent when regulatory decision makers rely on no-risk valuations put forth by interested parties. Companies arguing for more unlicensed frequency space, including Google and Microsoft, compete directly against firms such as T-Mobile and Verizon. To the degree that policy makers determine that market performance is

---

[30] This includes, by virtue of takings law, special legal arrangements whereby governments are given means to mitigate certain transaction costs. Because condemnation requires "just compensation," governments are forced to make rational trade-offs with respect to takings, mimicking a bidding situation stripped of opportunistic hold-up. See, generally, Epstein (1985).

hampered by transactional difficulties uniquely associated with the former, the straightforward remediation is to subsidize such demands. To this end, regulators can issue bidding credits to enhance the purchasing power of parties seeking to acquire unlicensed bandwidth, offer other incentives for industry consortia to bid (eliminating build-out requirements that discriminate against 'plug 'n play' network deployments not incorporating coordinated infrastructure construction), or fund public enterprises to bid openly for unlicensed bandwidth. In this manner, any underestimation of future demand due to the asserted public good problem will be offset, without losing the contribution to efficiency made by price discovery via competitive bidding.

d) Unlicensed bands generate greater innovation due to lower transactions costs.

It is indisputably correct that, with private property, there is a cost to defining, enforcing, and trading ownership rights (Demsetz 1967). Where society does not gain sufficient benefit to offset these costs, rights will typically not be so defined. Yet, there are often offsetting benefits, and there are no free lunches. Other forms of social coordination, including non-exclusive rights to access unlicensed frequencies, require resource expenditures. These include the opportunity costs that are associated with restrictions imposed on devices by regulators, delays in allocations that reduce or eliminate productive activity during an administrative processing period, and common interest tragedies that form when new, more efficient spectrum configurations emerge but are blocked due to regulatory gridlock.

The one-dimensional view put forward is that unlicensed bands uniquely entail transaction costs because they may be used "without having to negotiate permission from spectrum owners" (Benkler 2012, 89). In fact, all devices must be authorized by regulators. That process may take years, as in the TV white spaces proceeding. And once permission is granted in the form of *type acceptance* (a standards-based approval allowing innovation in certain dimensions) the imposed limits create "transaction costs" by blocking a range of otherwise useful technologies, services, or applications. Benkler (2012, 72) concludes that U-PCS was an "outright failure" or 3650 MHz unlicensed "largely anemic" (Ibid., 100), without noting that such outcomes reflect the transaction costs of coordinating economic activity using non-exclusive use rights. Liberal licenses created in PCS supported successful coordination that generated high levels of consumer welfare.

Owners of exclusive rights internalize transaction costs and, hence, attempt to minimize them in crafting sharing rules. This is seen in the way a mobile phone "negotiate[s] permission from spectrum owners." The bargaining process is not zero-priced, but is efficiently bundled. Standard contracts allow one transaction, for a post-paid contract or pre-paid card, to set terms of trade for thousands of phone calls, text messages and mobile data connections. Indeed, a roaming phone user accesses licensed frequencies around the globe controlled by carriers it has never negotiated with, seamlessly completing valuable connections. Roaming access options for local wi-fi connections also form, also allowing for transactional gains. But the efforts do not face unique barriers due to exclusive spectrum rights.

MLE (2011, 2) posit that "[unlicensed spectrum] provides a platform for innovation upon which innovators may face lower barriers to bringing wireless products to market, because they are freed from the need to negotiate with exclusive license holders." They further declare that "unlicensed spectrum can make the cost of setting up and deploying systems for local wireless transmission extremely low. There are no licensing fees to pay, no approvals to obtain, and no need for radio frequency planning" (Ibid., 15; footnote omitted). This ignores that unlicensed bands are regulated by administrators that can permit – or block – innovations, and who engage in "radio frequency planning" without the information or incentives promoting efficiencies elsewhere. Moreover, the view fails to incorporate the benefits of motivated rights holders, who seek to maximize spectrum value by coordinating investments and complements. This has led to the creation of a rich innovation ecosystem in mobile markets with respect to devices, services and applications. This process is materially advanced by a spectrum property rights system that substitute private, for-profit "frequency planners" for public ones.

Spectrum ownership rights often help organize the introduction of new products by allowing efficient transactions to occur. LightSquared, an innovative firm that attempted to introduce a new 4G LTE national network to compete with incumbent mobile operators by sharing L-Band satellite frequencies, is an excellent example. Originally, satellite licenses were issued to multiple operators, and interleaved. This meant that, while considerable bandwidth – at least 40 MHz – was available for productive use, the too-numerous borders created spillovers that rendered valuable services such as LTE effectively impossible to supply. By paying other licensees to trade, however, spectrum rights were rationalized. As the FCC noted, in approving the license transfers, the band could not have supported valuable innovative technologies if the private bargains had not delivered a cooperative solution making large, contiguous blocks of spectrum available.

Nonetheless, the emerging network was thwarted in 2012. Regulators, reacting to an interference dispute with GPS interests, using an adjacent band, revoked previous authorizations. As shown in Hazlett & Skorup (2014), the social gains from additional LTE competition was likely in the tens of billions of dollars, with costs (targeted fixes to upgrade GPS radios, remediating possible spillovers) only a small fraction. The net social loss in pre-empting the competitive foray sprang from the non-exclusive GPS band rights, a jumble of licensed (satellite) and unlicensed radios. With no responsible agent able to negotiate with Lightsquared (which offered to invest significant sums to buy new GPS equipment for impacted parties), the dispute was determined by administrative process and political rent seeking. The FCC bent to the protests of powerful incumbents supporting GPS – including airlines, the Federal Aviation Administration and the U.S. Department of Defense – in halting the 4G network. Blair Levin, the head of the 2009-10 National Broadband Plan Task Force, which had projected that 40 MHz of L Band spectrum would be released to the market by 2015, characterized the spectrum allocation choice (to protect GPS by effectively making adjacent frequencies a guard band) thusly:

> Something extraordinary happened last week. Our country reallocated 40 MHz of commercial spectrum. No Notice of Proposed Rulemaking from the FCC. No notice and comment period. No economic analysis. Not even a legal decision stating that that is what we are doing.[31]

This lack of transparency was a product of the lack of property rights. Had an effective owner of the GPS band been in place, that party – acting to maximize returns -- would have considered the relevant trade-offs: what damage the adjacent use might cause against what Lightsquared would pay to offset it. Many billions of dollars in annual gains were possible. But due to the truncated GPS Band use rights, fragmentary and non-exclusive, "negotiations" were only possible through the FCC. These agents do not internalize efficiency gains, but focus on other margins. As Levin describes:

> Through a complicated process—mostly out of the public eye—of K St. machinations, inter-agency battles, and congressional pressure, we as a country came to the unstated but clear conclusion that the GPS industry has a primary right to use the spectrum in the band owned by LightSquared.[32]

The view that exclusive ownership rights are costly (true) and that non-exclusive ownership rights supervised by regulators are free (false) flows from an asymmetric set of assumptions and evinces little understanding of the policy issues at hand. MLE produce just such a naïve, asymmetric analysis:

> Another drawback to property rights is that they can stifle third-party innovation: third-party innovators face a threat of *hold-up*. A company that comes up with a new mobile device or business model needs to convince the owner of the spectrum to let it develop its idea, and it may have to share a large fraction of the value that is created with the spectrum owner. If the new development threatens the owner's existing business, it is particularly unlikely to be allowed. And, if the innovation requires the assent and coordination of multiple spectrum owners, it is even more difficult to get the owners all to agree. The potential for this type of coordination failure is sometimes referred to as the *tragedy of the anticommons* (MLE 2011, 13; emphasis original; footnote omitted).

The passage relies on one set of asserted failures, for which it gives no examples. It ignores others, for which examples of "coordination failures" are many -- including the Lightsquared debacle, U-PCS, 3650 MHz and TV Band white spaces. But most interestingly, it includes a citation that is stunningly incomplete. The reference provided

---

[31] Blair Levin, *Remarks to the Minority Media & Telecom Council*, Minority Media & Telecom Council (Mar. 8, 2012).

[32] *Id.* Levin was referencing the fact that not only was the FCC limiting L Band use by Lightsquared to provide an interference buffer for GPS receivers using adjacent channels, but was protecting GPS radios that actually used L Band transmissions to better interpret locational information.

is to Michael Heller's classic 1998 article in the HARVARD LAW REVIEW, *The Tragedy of the Anticommons: Property in the Transition from Marx to Markets.* The essay described how, in post-Communist Moscow, resources that were privately owned nonetheless sat idle. The reason was that ownership rights, when highly complementary, may not be productively employed if owned by many different parties. The policy solution was to avoid rights fragmentation. One might infer that that would often favor exclusive spectrum rights, where bands are controlled by profit-maximizing firms or other organizations, over unlicensed allocations, where property rights are distributed to thousands or millions of uncoordinated parties. But the article did not apply the anticommons analysis to problems of spectrum allocation.

Ten years later, however, Heller wrote a book that did. In his 2008, THE GRIDLOCK ECONOMY, Heller spent a chapter on the topic. There he wrote, contra MLE 2011, that unlicensed bands generate transaction cost problems that can be largely remedied by exclusive, liberal spectrum rights. He took pains to note that "unlicensed spectrum can't operate outside of the law of commons tragedy" (Heller 2008, 86). He illustrated: "You and all your neighbors might be happy to sell your rights to garage-door opener spectrum and get in exchange access to some next-generation technology... But in a spectrum commons, there is no way to make that deal" (Ibid., 85-86). The lost opportunities may be unnoticed, as "the gridlock side of unlicensed spectrum remains invisible..."[33]

MLE 2011 make other errors. One factual misconstruction is to claim empirical support for their unlicensed-creates-innovation claim by noting that

> The FCC permitted spread spectrum techniques in unlicensed bands for the first time in 1985, laying the foundations for Wi-Fi, Bluetooth, and many other standards, and already the first WLAN products were appearing in the late 1980s. In contrast, the two key applications of spread spectrum techniques on licensed spectrum, namely code division multiple access (CDMA) and time division multiple access (TDMA), were not introduced in cellular networks until the mid-1990s. [MLE 2011, 16; footnotes omitted]

This juxtaposition, which footnotes to the work of one of the coauthors of this article, mixes up events. The introduction of spread spectrum radios occurred first in cellular bands, even as it took a bit longer (1988) for the FCC to allow any digital technologies to be deployed there. CDMA phone networks (2G) were launched, Dec. 1995 to April 1996, in Hong Kong, USA, and South Korea.[34] Wi-fi radios were not released until 1999 (Hayes 2008). The WLAN devices that MLE note, introduced earlier, did not use spread spectrum technologies and, in any event, did not become highly popular innovations achieving mass market acceptance. Success for wi-fi utilized key technical innovations in cellular phones – in particular, the technology enabling the high-frequency gallium

---

[33] Heller (2008), 86.

[34] CDG Resources: CDMA History; https://www.cdg.org/resources/cdma_history.asp.

arsenide chip-scale amplifiers. These power amplifiers in wi-fi radios "took a free ride on technological innovation in licensed bands" (Hazlett & Leo 2011, 1080).

In a far broader sense, Wi-fi's market opportunity came, similarly to the deployment of cordless phones, with the build-out of networks using privately owned spectrum. The telecommunications network signals coming into a home or business can be conveniently distributed locally with either wires or wireless transmissions. Given the short distances involved within the local area network, coordination challenges endemic in the coordination of wide area networks fade. Wireless (via unlicensed) came to rival, and largely supplant, wires given the advantages of "cordless" phones and computers. The wi-fi innovation complemented the emerging networks, but is mischaracterized as leading it. And the misleading nature of the omission expands when the property structure of the wide-area complement is noted. There is no trend away from exclusive rights evinced by the growth of broadband networks to which cordless PCs attach than there is a trend away from telecommunications networks incorporating cordless phones.

Fig. 2. Global Apple App Store Downloads (2008-2014)[35]

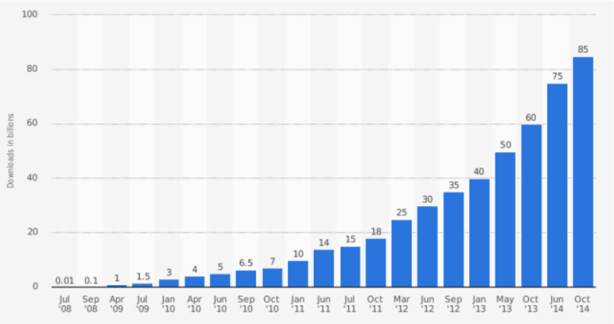
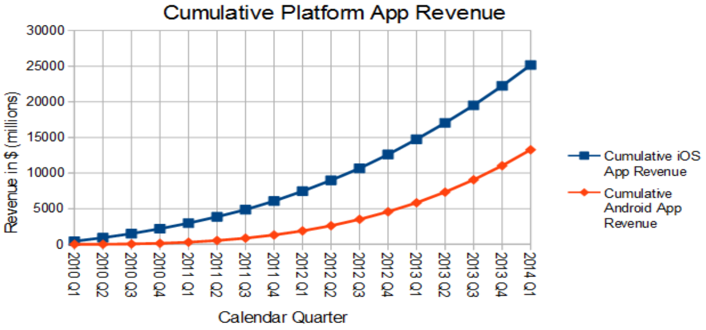


Fig. 3. Global Apple/iOS and Google/Android App Store Revenues[36]

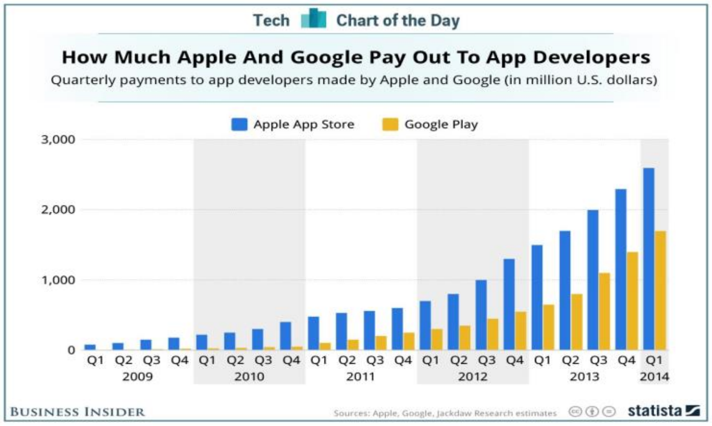


---

[35] Source: Statista.

[36] Mark Hibben, *Apple iOS Vs. Android: The Wealth Of Ecosystems,* Seeking Alpha (June 30, 2014).

FIG. 4. GLOBAL APPLE/IOS AND GOOGLE/ANDROID DEVELOPER PAYOUTS



In any event, measuring innovation is highly problematic. Many unlicensed devices have proven useful; many have not. Failure is part and parcel of a vibrant, innovative market, so not much can be said about success rates. Alternatively, liberally licensed spectrum has supported many popular services and devices, generating revenues far above wireless applications relying on rival allocations. In particular, emerging "mobile ecosystems," have become platforms for innovation supporting very substantial consumer welfare generation. Apple's App Store and Google's Google Place offer hundreds of thousands of software programs for wireless devices, generating billions of downloads – over 85 billion for Apple alone, since 2008 (see Fig. 2), and at least $38 billion in direct (non-advertising) revenues for Apple plus Google/Android (see Fig. 3). It speaks to the transaction cost issue that hundreds of thousands of transactions with third party developers have been "negotiated," and that the direct payout to these suppliers has totaled about $28 billion (70% of direct app revenues from customers), exceeding $4 billion in the first quarter of 2014 alone.

Champions of unlicensed allocations may argue that downloads on wi-fi connections account for a large portion of the physical transmissions, as do USB wires and other links. But there is little question that flexible use wireless licenses played the key role in enabling 2G, 3G and 4G network build-outs. This causative pattern – mobile licenses are issued, networks are built to complement the value of the spectrum allotted to such licenses, other elements of the ecosystem are constructed around these emergent platforms – is considered obvious in market analyses. In a recent paper, engineers considering the "wave of transformation [in] the mobile ecosystem," define the "drivers [as] attractive consumer tariffs for ubiquitously available mobile broadband access, development of increasingly mobile devices, and the emergence of an 'App Economy'" (Basole & Karla 2011, 313).

While measurements of consumer welfare better reflect actual social gains from innovation, the idea that these (potentially superior) quantifications be used as evidence on which policy maker allocate spectrum according to their reading of the data is underwhelming. That view, at bottom, constitutes the administrative allocation system of such widely noted systemic failure. Other tools can be used to reveal valuations more reliably, while creating more fluid conditions for the competitive reallocation of spectrum in secondary markets.

e) Unlicensed bands are supplanting licensed bands in terms of utility.

The idea the wi-fi competes with and may largely supplant mobile networks is a running theme in Benkler (2012). It is used to support the view that non-price allocations (favoring unlicensed) should be increased. But if the market premise is correct, the policy implication is just the reverse. The claim that more bandwidth would better accommodate additional unlicensed spectrum use reveals that scarcity still obtains, no matter the numerous claims made previously (see Benjamin 2003), and that valued opportunities are yet sacrificed by allocating more spectrum for one set of employments versus another. These social costs need be accounted for to achieve a rational use of resources, and the information supplied by prices established in competitive spectrum transactions hence *increases* in value.

The problem with imposing a favored allocation by fiat is seen in the embrace of unlicensed allocations by cable TV operators. Having danced with the possibility of entering mobile markets, acquiring AWS-1 licenses (accessing about 20 MHz, nationwide) in 2006, SpectrumCo (a consortium of four leading cable TV carriers) ultimately decided against the strategy, selling their licenses to Verizon in 2012.[37] These operators settled, in the alternative, on leveraging their fixed networks to offer nomadic wireless access via hotspots. Suppose cable's wi-fi strategy becomes successful, getting mobile customers of, say, Verizon, to drop their subscriptions and switch to Comcast's wireless service. Because Verizon acquires its licenses in market exchanges, it accounts for opportunity costs; Comcast, relying on set asides of the Federal Communications Commission does not. Hence, the competitive outcome is distorted.

Benkler points to the 2011 market entry by Republic Wireless as emblematic of the shift to unlicensed spectrum. Republic, held by a private firm that does not disclose subscriber counts but which in 2014 claimed a customer base "in the healthy six figures,"[38] offers low-cost mobile phone service by relying on a "wifi first" strategy. A subscriber making a voice call or data connection will be directed to a fixed network via wifi; if that is unavailable, the requested link is provided by a mobile network. In Republic's case, the network is Sprint's, which contracts with some 130 MVNOs (mobile virtual network operators).[39] Other MVNOs are attempting similar strategies, including a recent entry announced by Google (using both T-Mobile and Sprint mobile networks). Benkler argues that wifi-over-unlicensed is competing with and evidently eclipsing mobile-over-licensed, which is used as a backup. The reverse could be claimed: mobile

---

[37] Marguerite Reardon *What $3.9 billion Verizon/cable spectrum deal means to you (FAQ);* CNET NEWS.COM (Aug. 25, 2012).

[38] Scott Moritz & Olga Kharif, *Wi-Fi alternatives threaten cell carriers; Upstarts can provide service at lower cost,* BLOOMBERG NEWS (March 3, 2014).

[39] Ibid.

operators are enabling use of wifi hotspots.[40]

The reality is that, as from the beginning, mobile networks are constructed and operated with a mix of wires and wireless assets, utilizing fixed and mobile components. As has been joked, "mobile networks are actually stationary – it's only the subscriber who moves." To the extent that WLAN nodes accessing unlicensed bands are increasingly helpful in bringing wireless subscribers to the fixed network, the evolution of the network architecture continues. Yet, to encourage efficient choices in the allocation of spectrum resources, relative demands for the alternative use models are inherently of interest. Should additional bandwidth be configured to protect local wi-fi traffic, or should greater mobile bandwidth be diverted from alternative uses to accommodate the coverage component for mobile usage?

Champions of unlicensed allocations argue that categorical assertion of the importance of hotspots decides the issue; it does not. What matters is the panoply of marginal valuation factors discussed at the outset of this paper. To point to increased use of hotspots as full and complete evidence for additional allocations is to commit the *Broadcast TV Spectrum Valuation Fallacy*.

In fact, however, the premises of the argument – that wi-fi is a net substitute for mobile wireless – is rejected by regulators. WWANs and WLANs have been found by both the Federal Communications Commission and the U.S. Department of Justice Antitrust Division to inhabit distinct service markets. While substitute uses are easy to understand – as a mobile phone subscriber switches to wi-fi (or USB cords) for data-intensive downloads undertaken at home or office – the net economic impact between the wireless link technologies appears to be complementary. In considering the proposed merger of AT&T and T-Mobile (the second and fourth largest wireless carriers; the government filed suit to block the merger, which was then abandoned) in 2011, antitrust authorities defined the market to include the four national U.S. cellular carriers (the merging parties plus Verizon and Sprint). Wi-fi hotspot services, networks, or "wi-fi first" MVNOs were excluded (DOJ 2011). A similar determination was made by sector regulators (FCC 2011, pars. 35-36).

It may serve a rhetorical strategy to characterize wi-fi access as primary and mobile access as secondary, but the theme is undermined by common use of the term "wi-fi offload" and, in any event, carries little economic meaning. AT&T offers mobile service to about 100 million U.S. subscribers, incorporating its more than 30,000 hotspots,[41] at least 50,000 cell sites,[42] and 40,000 "small cells."[43] There is no evident

---

[40] We note the irony that, whereas scholars such as Tim Wu (2007) argued that mobile operators were suppressing access to wi-fi for anti-competitive reasons, a new argument has emerged that mobile carriers are facilitating wi-fi access by providing a back-up service filling in for wi-fi coverage gaps.

[41] Joan Marsh**,** *The White Space Black Hole,* AT&T Public Policy Blog (April 2, 2015); http://www.attpublicpolicy.com/fcc/the-white-space-black-hole/.

[42] When AT&T Wireless merged with Cingular in 2005 the new entity (now called AT&T Mobility) emerged with 50,000 cell sites. *AT&T Mobility*, WIKIPEDIA (visited April 27, 2015). Company officials claim that the total continues to grow, but the authors could find no more recent citation.

migration away from investments leveraging licensed spectrum. Moreover, the firm bid $18 billion in early 2015 to acquire rights for marginal bandwidth, specifically 20 MHz of AWS3 frequencies.[44] This revealed preference for exclusive rights suggests provides direct evidence of high productivity.

To move the right amount of spectrum into one set of employments versus another – to where the social product is maximized – requires such information as to marginal valuations. These are revealed in market bidding, which occurs in primary and secondary market transactions where exclusive, flexible rights are exchanged. Under current regulatory procedures, however, prices for additional unlicensed bandwidth are held at zero, distorting choices. Not only is demand revelation sacrificed, it enables rent seeking opportunities. Regulators are lobbied by firms to apply one set of rules or another; the firms loyally promote their corporate interests, but the evidence they proffer on social values is considerably less trustworthy – as per economic theory, regulatory experience, and the stated premises of the FCC's BOS (2008) study – than actual transactions. Sacrificing such information reduces choices to politically driven administrative determinations, misallocating spectrum not only between liberal licenses and unlicensed bands, but between competing unlicensed set-asides.

Currently, rent-seeking battles wage between firms like Comcast and General Motors over use of the unlicensed 5.9 GHz band. Under existing rules, the permitted applications are tied to vehicle telematics, including those used in collision-avoidance and driverless cars. Under proposals put forth by Comcast and other parties, at least some of the 100 MHz band would be set aside for WLAN use (extending the 5.8 GHz band).[45] Auto makers argue "safety first," and warn that reducing the extant allocation would endanger lives. Champions of additional unlicensed bandwidth for wi-fi argue that, at the margin, some of the band could be opened to wi-fi hotspots, facilitating new services while having *de minimus* impact on car safety. Moreover, the 5.9 GHz auto telematics allocation has been in place for many years and yet car companies have made scant use of it. As one writer explains: "Nobody is saying that better access to YouTube videos should take precedence over life-saving technology. But tech companies argue that a network of connected cars is still many years away and the auto industry is barely using its dedicated airwaves. Meanwhile, convention-goers in Las Vegas can't get a decent Wifi signal."[46] In fact, on some margins cat videos on YouTube *should* take precedence over this "life-saving technology." The difficulty is in finding that margin via administrative fiat rather than via spectrum input prices.

Similar conflicts are raging with respect to U-LTE, with a dispute over how

---

[43] Pico or femto cells further subdividing licensed spectrum into smaller cell sites within macro cells. The 40,000 count is an end-of-2015 projection by AT&T. Phil Goldstein, *AT&T exec: We'll be adding 1,500 to 3,000 cell sites per year for 'foreseeable future,'* FIERCE WIRELESS (May 21, 2014).

[44] Phil Goldstein, *Analysts: AWS-3 auction helps AT&T catch up to Verizon in spectrum ownership in major markets,* FIERCE WIRELESS (Feb. 2, 2015).

[45] Joann Muller, *Should Talking Cars Share Coveted Airwaves With Wifi Providers?* FORBES (Feb. 22, 2014).

[46] *Id.*

unlicensed bands are used to complement mobile networks,[47] and in the way in which the 5 GHz bands are allowed to serve WISPs (wireless Internet Service Providers). WISPs object to an FCC-proposed power limit reduction in use of the 5.8 GHz band which, while helping to protect hotspots and other local applications, imposes burdens on rural ISPs. One "WISP, which serves certain rural regions of northern California, said the combined effect of narrowing the band and the lowering power levels could result in 700 to 1,000 of its subscribers' losing service…"[48] Perhaps such harmful effects are worth the benefits provided by the change, perhaps not. Without bidding for inputs revealing tradeoffs, regulators will have stories about relative marginal valuations but be lacking hard evidence of the relevant costs and demands.

Claims about wi-fi popularity are put forward as if they end to such tradeoffs. One exuberant effort begs comment. Benkler quotes Hazlett (2001) on the efficacy of using unlicensed spectrum, given Quality-of-Service issues; the same passage excites Benkler sufficient for him to use it twice in the same article (2012; 75, 114). In the first, it appears with an impressive foreign language zinger: "After all, to quote the most vocal critique of open wireless policy, with open wireless, as with the Internet, '[c]lassically, the brain surgeon cannot read the life-or-death CT-scan because the Internet back- bone is clogged with junk e-mail.' *Eppur si muove.*"[49]

And yet it does not. The confusion stems from Benkler's misdirection. He leads his readers to believe that the QoS concerns related to WLANs, to which he sparred by noting the emergent use of wi-fi: "As early as 2008, it was already clear that hospitals were buying and deploying open wireless technologies, in particular Wi-Fi, as the core wireless technology for in-hospital *medical grade*, *mission-critical wireless networks*" (emphasis original). But wi-fi was not in the 2001 discussion, and local hotspots or "in-hospital" applications not then in debate. The passage quoted responded to Benkler's 1998 argument that wide area networks were being created using unlicensed bands. These were advanced as competing with, perhaps eclipsing, networks relying on exclusive spectrum rights. Specifically, Benkler noted the business model pursued by Metricom, offering subscribers Internet access (at pre-broadband speeds) using the 900 MHz ISM band. It was a brave prediction, and its empirical fate bears noting.

It is not determinative that Metricom long ago went bankrupt (twice, in fact[50]); a lively, innovative field often sees pioneers exit. The important fact is that Metricom was not overtaken with additional entrants into the WISP niche market using "free spectrum" to out-compete rivals, but by (among others) mobile wireless operators deploying exclusive spectrum rights acquired via expensive purchases. According to FCC data for Internet Service Provider subscribership, as of year-end 2013, total "Fixed Wireless" subscribership – which includes both service providers relying on licensed and

---

[47] Amy Schatz, *FCC Plans a Vote on New Spectrum Sharing Plan*, RECODE (March 27, 2015).

[48] Tammy Parker, *Wi-Fi advocates clash with automakers, WISPs over FCC's 5 GHz plan*, FIERCE WIRELESS (Aug. 17, 2014).

[49] The Latin admonition ("yet it moves") is either from the (likely apocryphal) story regarding Galileo recanting his recantation, or the title of the sixteenth episode of hit NBC series, *The West Wing*.

[50] *Ricochet (Internet Service)*, WIKIPEDIA (visited April 27, 2015).

unlicensed spectrum (but does not give a breakdown) – came to 858,000. Mobile broadband subscribership, all of which relies on licensed spectrum, came to 197 million; cable, DSL and fiber fixed links, about 100 million. See Table 3. The idea that investments to organize emerging broadband networks could be more efficiently undertaken with "spectrum commons" has thus far proven false.

TABLE 3. SUBSCRIBERS FOR VARIOUS ISP SERVICES (2009-13, THOUSANDS)[51]

	2009	*2010*		*2011*		*2012*		*2013*	
Technology	*Dec*	*Jun*	*Dec*	*Jun*	*Dec*	*Jun*	*Dec*	*Jun*	*Dec*
Total	**136,294**	**157,017**	**182,065**	**206,131**	**230,201**	**243,359**	**262,564**	**284,692**	**293,397**
Total Fixed	**79,994**	**81,764**	**84,521**	**86,575**	**88,317**	**89,945**	**92,511**	**93,986**	**96,032**
aDSL	**30,987**	**30,759**	**31,470**	**31,611**	**31,330**	**31,173**	**30,974**	**30,657**	**30,690**
sDSL	**225**	**191**	**167**	**159**	**148**	**139**	**132**	**117**	**108**
Other Wireline[1]	**719**	**755**	**795**	**778**	**795**	**770**	**796**	**769**	**772**
Cable Modem	**42,439**	**43,923**	**45,334**	**46,698**	**48,263**	**49,664**	**51,646**	**52,760**	**54,009**
FTTP[2]	**3,980**	**4,441**	**4,993**	**5,477**	**5,898**	**6,300**	**6,733**	**7,250**	**7,745**
Satellite	**1,116**	**1,144**	**1,176**	**1,204**	**1,190**	**1,217**	**1,454**	**1,623**	**1,849**
Fixed Wireless	**527**	**551**	**587**	**649**	**693**	**682**	**777**	**810**	**858**
Mobile Wireless[3]	**56,300**	**75,253**	**97,544**	**119,556**	**141,883**	**153,414**	**170,053**	**190,706**	**197,365**

Of course, this is not the only prediction that was offered suggesting that unlicensed spectrum would render licensed spectrum obsolete. In 2002, omitting any mention of the business model of the by then bankrupt Metricom, Benkler promoted the idea that ad hoc wi-fi meshes could form, displacing network-centric deployments and supporting efficiencies unavailable with licensed spectrum. With ad hoc formations, consumers would link wi-fi radios without coordination provided by an ISP. As seen in the intervening years, wi-fi meshes – discussed in the technical literature since the early 1980s – have not made much impact. While mesh architectures have been deployed, both on licensed and unlicensed bands, they have almost entirely been the work of network deployments in which an enterprise or ISP managed the process. In this model, the network operator can also contract for spectrum rights, removing a key argument for unlicensed bandwidth set aside for non-exclusive use by regulators.

Given the lack of progress in consumer adoption of ad hoc mesh networks, Benkler's 2012 article extolling the virtues of additional unlicensed allocations scant mention of them. Yet here, again, the failure of advertised technologies to overcome coordination problems merits appreciation. Benkler (2012, 95), failing to note the predictions he appears to have abandoned, complains that Hazlett "continues to make the same arguments, sometimes almost verbatim" supporting markets in spectrum, and "ignores the actual experience of equipment markets in the past decade, [which] should lead to some skepticism…" In fact, the marketplace evidence in wireless devices, competitive bidding for licenses, applications and services is exactly why the economic theory of spectrum markets has become more compelling since propounded by Coase (1959), and why defenses of traditional administration allocation become less so.

---

[51] FCC 2014, Table 5: for ISPs delivering at least 200 kbps, one way.

IV. CONCLUSION

FCC to Decide in Battle for TV Spectrum

-- WALL STREET JOURNAL headline (Aug. 18, 2008)

Considerable evidence has accumulated on the challenges of radio spectrum allocation policy. The traditional regulatory approach, still commonly employed, has proven expensive relative. Without valuation information that might otherwise be gleaned from competitive bidding, it forces administrative choices to made across the various regimes, licensed and unlicensed, that are employed to coordinate wireless activities. In this process, as captioned by the *Wall Street Journal*, policy makers pre-empt the market and make fundamental rule choices.

There are alternative policy paths. In general, regulators have seen their advantages. While often compromised in practice, it is standard for policy makers in the U.S. and internationally to pronounce *technological and service neutrality* as best practices. Instead of centrally planning specific technologies, business models or applications, and then requisitioning spectrum inputs, it has become evident that allowing flexible use of radio spectrum can generate large social gains. That way market forces are left to discover efficiencies, including those delivering experimentation and innovation. When regulators attempt to determine outcomes, these gains are deterred.

In both licensed and unlicensed spectrum, considerable progress has been made towards liberalization over the past three decades. But the underlying spectrum allocation still holds vast amounts of spectrum hostage to traditional rules and restrictions. In moving to reallocate under-utilized frequencies, making them available for potentially more productive use, public policy is too often captured by the template of the past: determining, based on administrative conclusions, how radio spectrum should best be utilized. This forecloses the rich flows of information revealing market demands. And, as practiced, often results in a conflation of the value of given wireless applications with the optimal deployment of radio spectrum inputs. That can result in distinct and enormously costly misallocations, as in the *Broadcast TV Spectrum Valuation Fallacy*.

Spectrum reallocation is seriously impeded by the “shipping and handling” costs of the administrative process. It is appropriate to examine alternatives that better reveal marginal valuations and which distribute incentives for diverse agents to contribute to positive sum transactions. The FCC (2010, 81) described its proposal for an Incentive Auction, as a “market-based way to reassign spectrum, shifting a contentious process to a cooperative one.” Whatever the outcome of that particular approach, the quest to discover new and improved means for moving spectrum into the most productive employments is an important step in the right direction.

Most fundamentally, that is not because the apps provided in licensed spectrum are more valuable than the apps provided in unlicensed. It is because market-based mechanisms that reveal superior information about relative values, and allow for adjustments to be made by well-incentivized actors not constrained by administrative spectrum allocation rules, can accommodate efficient activities with special force. Were the parties to be arguing for more unlicensed allocations, or for different types of unlicensed rules, to bid against parties with different arguments, demands for the conflicting approaches could be made visible. And new bandwidth could be made available, without the debilitating burdens of deadening regulatory delay or tragedy of the anticommons, to support the most valuable. Public parks, like unlicensed bands, are best provided under conditions of transparency and with preferences revealed. It is not "anti-wifi" to suggest that liberal licenses offer pronounced transaction cost advantages any more than the transition from broadcast to cable – or the reallocation in the FCC's Incentive Auction – are anti-television.

Categorical pronouncements that the world is embracing wi-fi and that, therefore, licensed allocations should be pre-empted by unlicensed bands fit neither economic theory, empirical assessments of consumer welfare generated in wireless, nor the experience gained by studying regulatory history. MLE (2011) argue that unlicensed frequencies were "garbage bands" that became socially valuable due to the non-exclusive rules applied to them by the FCC, but ignore that they became "garbage bands" not due to nature but from regulatory rules. Assigning fragmented, overlapping property rights that are difficult to rationalize often yield common interest tragedies (Heller 2008), blocking efficient change in response to shifting technological options, business innovations, and consumer demands. The view omits the fact that "garbage bands" have been rehabilitated via overlay rights, allowing private agents to trade (such that incumbents can be incentivized to substitute for spectrum inputs, freeing bandwidth for entrants) as licensed PCS successful transitioned where unlicensed PCS failed. And it sidesteps the key consideration that such bands, made operational by limiting power levels and technical choices so as to reduce conflicts, inherently block other possible usage configurations. These deterred deployments may well be more valuable than those that are protected, the emerging outcome in the TV Band white spaces proceeding.

To gauge both the value of marginal allocations, on the one hand, and the cost of opportunities foreclosed, on the other, it is necessary to take a more economic, and broader, view of the spectrum allocation process. Rather than associating particular applications with particular government allocations, a systemic approach should attempt to bring better information, greater fluidity, and more effective feedback mechanisms into the market. That can be done, but not if basic analytical and empirical tools are left unutilized.

V. REFERENCES

Basole, Rahul C., and Jürgen Karla. 2011. *On the Evolution of Mobile Platform Ecosystem Structure and Strategy*, BUSINESS & INFORMATION SYSTEMS Engineering 313.

Benjamin, Stuart Minor. 2003. *Spectrum Abundance and the Choice Between Private and Public Control,* 78 N.Y.U. LAW REVIEW 2007 (Dec.).

Benkler, Yochai. 1998. *Overcoming Agoraphobia: Building the Commons of the Digitally Networked Environment*, 11 HARVARD JOURNAL OF LAW & TECHNOLOGY 287 (Winter).

_____. 2002. *Some Economics of Wireless Communications*, 16 HARVARD JOURNAL OF LAW & TECHNOLOGY 25 (Fall).

_____. 2012. *Open Wireless vs. Licensed Spectrum: Evidence from Market Adoption*, 26 HARVARD JOURNAL OF LAW & TECHNOLOGY 69 (Fall).

Bulow, Jeremy, Jonathan Levin & Paul Milgrom. 2009. *Winning Play in Spectrum Auctions*, NBER Paper 14765 (March).

Bykowsky, Mark M., Mark A. Olson & William W. Sharkey. 2008. *A Market-based Approach to Establishing Licensing Rules: Licensed Versus Unlicensed Use of Spectrum*, Federal Communications Commission OSP Working Paper No. 43 (Feb.).

Cave, Martin. 2006. *New Spectrum-Using Technologies and the Future of Spectrum Management: a European Policy Perspective*, in E. Richards, R. Foster & T. Kiedrowski, eds., COMMUNICATIONS: THE NEXT DECADE, Ofcom (Nov.).

Cisco. 2009. *Wi-Fi versus Femtocell*, Cisco Slideshare (Dec. 11).

_____. 2015. *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2014–2019,* White Paper (Feb. 3).

Coase, R.H. 1959. *The Federal Communications Commission,* 2 J. LAW & ECON. 1.
_____. 1993. *Law and Economics at Chicago*, 36 J. LAW & ECON. 239 (April).

_____, William Meckling & Jora Minasian. 1995. *Problems of Radio Frequency Allocation*, Rand Corporation, Publication DRU-1219-RC (Sept.)

Cramton, Peter, Evan R. Kwerel & John R. Williams. 1998. *Efficient Relocation of Spectrum Incumbents,* 41 J.L. & ECON. 647 (Oct.).

De Vany, Arthur. 1998. *Implementing a Market-Based Spectrum Policy*, 41 J.L. & ECON. 627 (Oct.).

DOJ. 2011. *United States of America v. AT&T et al.: Complaint,* U.S. Department of Justice Antitrust Division, Case: 1:11-cv-01S60 (Aug. 31, 2011).

Entner, Roger. 2008. THE INCREASINGLY IMPORTANT IMPACT OF WIRELESS BROADBAND TECHNOLOGY AND SERVICES ON THE U.S. ECONOMY: A FOLLOW-UP TO THE 2005 OVUM REPORT ON THE IMPACT OF THE U.S. WIRELESS TELECOM INDUSTRY ON THE U.S. ECONOMY (Wash. D.C.: Cellular Telecoms. & Internet Association).

Epstein, Richard. 1985. TAKINGS: PRIVATE PROPERTY AND THE POWER OF EMINENT DOMAIN (Cambridge, MA: Harvard University Press).

Faulhaber, Gerald. 2006. *The Question of Spectrum: Technology, Management, and Regime Change*, 4 J. ON TELECOMM. & HIGH TECH. L. 123 (2005-2006).

_____ and Daniel Farber. 2003. *Spectrum Management Property Rights, Markets, and the Commons*, in Lorrie Faith Cranor & Steven Wildman, eds., RETHINKING RIGHTS AND REGULATIONS: INSTITUTIONAL RESPONSES TO NEW COMMUNICATIONS TECHNOLOGIES (Cambridge, MA: MIT Press).

FCC. 2002a. Federal Communications Commission, *Spectrum Policy Task Force Report* (Nov. 15).

_____. 2002b. Federal Communications Commission, *In the Matter of Additional Spectrum for Unlicensed Devices Below 900 MHz and in the 3 GHz Band, Notice of Inquiry*, ET Docket No. FCC 02-328 (rel. Dec. 20, 2002).

_____. 2010. Federal Communications Commission, NATIONAL BROADBAND PLAN (March), Chapter 5: *Spectrum*.

_____. 2011. *Staff Analysis and Findings* (in Application of AT&T for transfer of T-Mobile USA licenses), Federal Communications Commission WT Docket No. 11-65 (Nov. 29).

_____. 2014. *Internet Access Service: Status as of December 31, 2013,* Federal Communications Commission, Industry Analysis and Technology Division, Wireline Competition Bureau (Oct.).

Fennell, Lee Anne. 2004. *Common Interest Tragedies,* 98 NORTHWESTERN U. LAW REVIEW 907.

Gilder, George. 1990. LIFE AFTER TELEVISION (New York: W.W. Norton).

Hardin, Garrett. 1968. *The Tragedy of the Commons*, 162 SCIENCE 1243 (Dec. 13).

Hazlett Thomas W. 2001a. *The Wireless Craze, the Unlimited Bandwidth Myth, the Spectrum Auction Faux Pas, and the Punchline to Ronald Coase's "Big Joke": An Essay on Airwave Allocation Policy*, 14 HARV. J.L. & TECH. 335 (Spring).

_____. 2001b. Thomas W. Hazlett, *The U.S. Digital TV Transition: Time to Toss the Negroponte Switch*, AEI-Brookings Joint Center for Regulatory Studies Working Paper No. 01-15 (Nov.).

_____. 2008a. *Optimal Abolition of FCC Allocation of Radio Spectrum*, 22 JOURNAL OF ECONOMIC PERSPECTIVES 103 (Winter).

_____. 2008b. *A Law and Economics Approach to Spectrum Property Rights: A Response to Professors Weiser & Hatfield*, 15 GEORGE MASON UNIVERSITY LAW REVIEW 975 (June).

_____. 2014. *Efficient Spectrum Reallocation with Hold-Ups and Without Nirvana* George Mason Law & Economics Research Paper No. 14-16 (June 10).

_____, & Matthew L. Spitzer. 2006. *Advanced Wireless Technologies and Public Policy*, 79 SOUTHERN CALIFORNIA LAW REVIEW 595 (March).

_____, Robert Muñoz & Diego Avanzini. 2012. *What Really Matters in Spectrum Allocation Design*, 10 NORTHWESTERN JOURNAL OF TECHNOLOGY & INTELLECTUAL PROPERTY 93 (Jan.).

Heller, Michael. 1998. *The Tragedy of the Anticommons: Property in the Transition from Marx to Markets*, 111 HARVARD LAW REVIEW 621 (Jan.).

_____. 2008. THE GRIDLOCK ECONOMY: HOW TOO MUCH OWNERSHIP WRECKS MARKETS, STOPS INNOVATION AND COSTS LIVES (New York: Basic Books).

Kwerel, Evan R., and John R. Williams. 1992. *Changing Channels: Voluntary Reallocation of UHF Television Spectrum*, Federal Communications Commission OPP Working Paper No. 27 (Nov.).

_____. 1993. *Moving Towards a Market for Radio Spectrum*, 16 REGULATION 53.

_____. 2002. *A Proposal for a Rapid Transition to Market Allocation of Spectrum*, Federal Communications Commission OPP Working Paper No. 38 (Nov.)

Milgrom, Paul. 2004. PUTTING AUCTION THEORY TO WORK (Cambridge: Cambridge University Press).

_____, Jonathan Levin & Assat Eilat. 2011. *The Case for Unlicensed Spectrum* Paper presented at the National Science Foundation (Oct. 12, 2011); also National Bureau of Economic Research Paper No. __.

NAB. 2009. *In the Matter of Spectrum for Broadband A National Broadband Plan for Our Future, To the Broadband Task Force, Comments – NBP Public Notice #26, The Association for Maximum Service Television, Inc. and the National Association of Broadcasters*, Federal Communications Commission GN Docket Nos. 09-47, 09-137 (Dec. 22).

_____. 2015. *In the Matter of National Association of Broadcasters Petition to Amend Sections 47 C.F.R 15.711(b) and 47 C.F.R 15.717 Regarding Changes to Certain Rules for Unlicensed Operations in the Television Bands, Repurposed 600 MHz Band, 600 MHz Guard Bands and Duplex Gap and Channel 37: Emergency Motion for Suspension of Operations and Petition for Rulemaking*, Federal Communications Commission Docket No. __ (March 19).

Negroponte, Nicholas. 1995. BEING DIGITAL (New York: Vintage).

Rysavy, Peter. 2008. *Mobile Broadband Spectrum Demand* (Dec.).

Shelanski, Howard A. 1997. *The Bending Line Between Conventional "Broadcast" and Wireless "Carriage,"* 97 COLUM. L. REV. 1048 (May).

Spiller, Pablo T., and Carlo Cardilli. 1999. *Towards a Property Rights Approach to Communications Spectrum,* 16 YALE J. ON REG. 53.

White, Lawrence J. 2000. *"Propertyzing" the Electromagnetic Spectrum: Why It's Important, and How to Begin*, MEDIA LAW & POLICY 19 (Fall).

Wu, Tim. 2007. *Wireless Carterfone*, 1 INT'L J. COMMUNICATION 389.

Before the
Federal Communications Commission
Washington, D.C. 20554

In the Matter of	)	
	)	
Amendment of the Commission's Rules with Regard to Commercial Operations in the 3550-3650 MHz Band	)	GN Docket No. 12-354
	)	
	)	

Reply Comments of

William Lehr
Massachusetts Institute of Technology

PALs as Options to Exclude GAA

August 1, 2014

William H. Lehr
Research Scientist
Massachusetts Institute of Technology
32 Vassar Street (32-G814)
Cambridge, MA 02139

Reply Comments of William Lehr
Massachusetts Institute of Technology

PALs as Options to Exclude GAA

August 1, 2014

Table of Contents

1. Introduction and Summary

I, William H. Lehr,[1] respectfully submit these reply comments in response to the Further Notice of Proposed Rulemaking in Regard to Commercial Operations in the 3550-3650 MHz Band ("*FNRPM*").[2]

---

[1] Economist and Research Scientist, Computer Science and Artificial Intelligence Lab (CSAIL), Massachusetts Institute of Technology; Contributor to the PCAST (2012) report on spectrum policy. See Attachment 1 for further biographical details. All opinions expressed herein are the author's alone.

[2] See Further Notice of Proposed Rulemaking, In the Matter of Amendment of the Commission's Rules with Regard to Commercial Operations in the 3550-3650 MHz Band, Before the Federal Communications Commission, GN Docket No. 12-354, Released April 23, 2014, available at:

First, I agree with many of the other commenters in this proceeding in applauding the Federal Communication Commission's (FCC's) efforts to expand commercial access to radio frequency spectrum by crafting rules for the Citizens Broadband Radio Service (CBRS) in the 3550-3650 MHz (3.5GHz) band,[3] and in moving forward to implement the recommendations of the *PCAST Report*.[4] As a spectrum policy analyst and economist, and as a contributor to the *PCAST Report*, I support the FCC's efforts to implement the three-tiered model for sharing spectrum to be managed by a dynamic Spectrum Access System (SAS) that was advanced in the *PCAST Report*.[5] This is an important step toward expanding spectrum sharing opportunities and enhancing economic incentives to use spectrum efficiently over time.

The FNPRM's three-tiered framework envisions adding two new tiers of commercial users to the Incumbent Access tier: Priority Access License (PAL) users; and Generalized Authorized Access (GAA) users. PAL users would bid for licenses that would grant them an assured degree of interference protection;[6] whereas GAA users would access spectrum on a license-by-rule basis, but without additional interference protection rights.[7] A key question that arises is how PAL and GAA users might coexist and share spectrum.

---

http://transition.fcc.gov/Daily_Releases/Daily_Business/2014/db0425/FCC-14-49A1.pdf, hereafter "*3.5GHz FNPRM*"). The opinions expressed herein are those of the author alone.

[3] Many commenters approve the FCC's efforts to expand commercial access to spectrum in the 3.5GHz band for shared use by multiple tiers of users, with sharing to be managed by a Spectrum Access System; however, there is significant heterogeneity in the commenters views regarding the PAL license terms, the extent to which GAA or PAL access should be preferred, the design of the SAS, and other issues. For example, see the comments filed by AT&T, Google, Microsoft, Nokia Solutions and Networks, Pierre de Vries, T-Mobile, Verizon, Wireless Innovation Forum, and others. (These comments were filed as "Comments of X," on July 14, 2014, in response to the *3.5GHz FNPRM*, note 2 *supra*. Hereafter, these will be referenced solely as "Comments of X filed in the *3.5GHz FNPRM* on July 14, 2014" unless another docket is being referred to.)

[4] See PCAST (2012), "Report to the President: Realizing the Full Potential of Government-Held Spectrum to Spur Economic Growth," President's Council of Advisors on Science and Technology (PCAST), July 2012, available at: http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast_spectrum_report_final_july_20_2012.pdf.

[5] I also support the many commenters who agree that the initial specification of exclusion zones designed to protect Federal incumbent users are too large, having been derived based on modeling scenarios that are inconsistent with current thinking of how the CBRS will be used (e.g., with more realistic modeling of radio system performance); and that power levels limits should be relaxed to allow more scope for commercial deployment.

[6] The FNPRM proposes allocating PALs in 10MHz chunks on a Census Tract basis, for one-year licenses. Some commenters argue in favor of much more granular assignments (e.g., see Google Comments to *3.5GHz FNPRM*, filed July 14, 2014); while others argue in favor of larger license areas and longer license lives (e.g., see T-Mobile or Nokia Solutions and Networks Comments, separately filed to *3.5GHz FNPRM*, July 14, 2014).

[7] GAA users would not have a claim of interference protection from other authorized users, and would be required to avoid interfering with PAL or incumbent users. However, all users of the spectrum may

Although many commenters support the FCC's overall approach in this proceeding, they disagree with respect to how PAL and GAA access to spectrum might best be managed.[8] Some parties have emphasized the importance of the PAL tier,[9] while others have focused on the importance of the GAA tier.[10] Some parties have called for greater reliance on a dynamic SAS to manage interference on the basis of more realistic models of the actual interference environment,[11] while others argue in favor of more static licensing terms (with larger territories and longer lived licenses than proposed in the *3.5GHz FNPRM*).[12]

Whether a spectrum user needs and is willing to pay for the interference protection afforded by a PAL, or whether GAA rights are sufficient for the user will depend on the usage context. The usage context will be a function of an individual user's wireless technology, business model, and market conditions (including the associated interference environment) that will evolve over time.[13] In this sort of complex decision-making situation, rife with asymmetric, uncertain, and private information, it is desirable to allow market forces greater scope for managing when and where PAL and GAA users may co-exist.

In these reply comments, I focus narrowly on how this may be accomplished by modifying the payment rules for PALs. This will help expand opportunities and improve economic incentives for both PAL and GAA users to invest in and make use of the CBRS. In *FNPRM* paragraph 127, the FCC seeks comments regarding how the payment rules for PAL bidders might be modified to

---

expect to be protected from illegal uses of the spectrum, which presumably would include non-compliant GAA, PAL, or incumbent devices.

[8] For example Google, the Wireless Innovation Forum, and Pierre de Vries argue in favor of managing spectrum access on the basis of interference threshold requirements, rather than geographic license territories, which would enable greater scope for GAA co-existence with PAL users. In contrast, AT&T, Verizon, T-Mobile argue in favor of expanded PAL protection rights and against GAA operation in active PAL territories (see separate comments to *3.5GHz FNPRM*, filed July 14, 2014).

[9] See Comments Nokia Solutions and Networks to *3.5GHz FNPRM*, filed July 14, 2014.

[10] See Comments of Microsoft to *3.5GHz FNPRM*, filed July 14, 2014.

[11] See Comments of Google to to *3.5GHz FNPRM*, filed July 14, 2014.

[12] See Comments of T-Mobile to *3.5GHz FNPRM*, filed July 14, 2014. The *3.5GHz FNPRM* proposes that PALs should grant access to 10MHz of spectrum and be awarded for 1 year and on the basis of Census Tracts (see paragraph 6, 49).

[13] For example users with interference-sensitive applications (e.g., hospitals) or seeking to deploy multiple cells over a larger serving area (e.g., LTE mobile network operator) may feel they need PAL interference protection for their usage case, whereas users of interference-tolerant applications (e.g., latency/noise tolerant data services or devices) or mass market consumers deploying cells for home networks may find GAA spectrum preferable. Also, the value of interference protection is likely to be higher in congested markets (urban rather than rural markets) and may be expected to change over time (either becoming more valuable as aggregate commercial usage increases over time or less valuable as technologies enabling more intensive sharing continue to improve and other spectrum becomes available).

provide additional incentives "for the productive use of spectrum" by allowing winning bidders to delay final payments until "initiation of service" in a specific PAL.[14]

A simple way to do this, within the context of the existing SAS and licensing framework, is to view the PAL as an *option to exclude GAA usage*. PAL licensees would acquire the right to exclude GAA access, owing P1 when the PAL is awarded, and then owing P2 when the licensee elects to exercise the option at some later time before the license expires. Until such time as the PAL option is exercised, GAA usage would be permitted in the PAL license territory.

In Section 2, I further elaborate how this might work in practice; and in Section 3 discuss the benefits such an approach offers. Section 4 addresses some of the concerns I anticipate might be raised; and Section 5 concludes.

2. PALs as *Options to exclude GAA*

As noted above, a principal difference between a PAL and GAA is that PAL users pay for a right to interference protection. Whether such protection requires exclusion of GAA users depends on the particular usage context. Viewing the PAL as an option to exclude GAA users from some point onward during the life of the license allows adjustment to the particular usage context.

In its simplest implementation, the licensee would *owe* P1 when the PAL is acquired (following the auction) and P2 when the "option to exclude" is exercised, for a total payment of P1+P2.[15] Two extreme cases may occur. First, the licensee may elect to exercise the option immediately, owing P1+P2. Were all licensees to do this, it would preclude any GAA operation in PAL spectrum. Second, the licensee may elect never to exercise the option. In that case, GAA devices would be permitted to operate in the spectrum for the entire duration of the license. Were all licensees to do this, GAA devices would be permitted to operate in all commercially accessible 3.5GHz spectrum.[16] Enabling the PAL option expands market opportunities for GAA access in a way that is responsive to the needs of individual PAL users that may vary by market and over time.[17]

For example, the most obvious case is that of a PAL user that does not expect to be operational in a license territory immediately upon receipt of the license, but anticipates requiring excluding GAA users once the PAL user goes operational. In that situation, both the PAL and GAA users

---

[14] See FNPRM, paragraph 127.

[15] The licensee owes the payment when the obligation for the payments takes place. The actual payments may be further split in time between down payments and final payments, if desired. This will impact valuations, but the added complexity of such further refinements is ignored here.

[16] GAA devices also may be permitted to operate in Incumbent Access spectrum. The focus here is on whether GAA devices are allowed to operate in PAL spectrum or only GAA spectrum.

[17] For example, in the same geographic regions or over time, different PAL users may have different tolerances to potential interference and preferences for excluding GAA devices. This means that the share of PAL spectrum open to GAA usage may decline more or less gradually over time.

can benefit: the GAA users gain from having expanded access to spectrum; and the PAL users gain by being able to delay (and thereby reduce[18]) their payments for interference protection.

Alternatively, the PAL user may recognize that they do not need to exclude GAA access over the entire territory, and may expect that the need to exclude GAA access may never arise unless the density of GAA usage exceeds some congestion threshold. Such a licensee may find the option to exclude valuable even if the licensee believes it is relatively unlikely that the option will be executed, and taking advantage of the option reduces spectrum access payments while still providing contingent interference protection.[19] Furthermore, it is worth noting that interpreting the PALs as options does not reduce the obligation of GAA to avoid interfering with PAL users, nor the effectiveness of other enforcement mechanisms that may be required to ensure this occurs.[20] Indeed, the option to exclude GAA is a complementary interference management tool, available to PAL licensees (who may individually choose to invoke the option) and the FCC (which may modify the attractiveness of exercising the exclusion option by the design of the PAL licenses and auction/pricing/payment mechanisms used to award PALs).

Finally, a user who never expects to invoke the exclusion option (and thus incurring P2) and does find the interference protection afforded by just paying P1 sufficiently valuable might simply elect to use GAA spectrum.[21]

The basic schema is simple to implement and involves just two modifications. First, the payment needs to be split between P1 and P2; and second, there needs to be a process for implementing the option in the SAS and connecting that with the FCC mechanism for collecting payments.

With respect to how to split the payment, there are multiple options, but as a start, the simplest approach might be to split the total PAL auction bid into two equal-sized payments. A winning bidder (with a bid of P for a PAL) would expect to owe ½ P when the license is awarded and ½ P when the licensee elects to exercise the option to exclude. The opportunity to delay payment would provide winning bidders with an economic incentive to avoid excluding GAA users unless the benefits of such exclusion outweigh the costs of exercising.

---

[18] Taking into account the time value of money, and the contingency of potentially not needing to invoke the exclusion provision should interference protection requirements be met while still allowing GAA sharing in the band.

[19] The option provides a measure of insurance that is valuable against contingent harms, the expected value of which may vary by user or market environment.

[20] For example, interference claim thresholds for receivers, third party monitoring (including spectrum sensing), software/firmware requirements for PAL/GAA devices that may be adopted to ensure compliance with the SAS, or other interference management strategies may also be incorporated as part of the overall spectrum management framework. The adoption of "PALs as options" is compatible with other approaches for managing interference, and should ease adoption of these since it makes exclusion policies more flexible.

[21] It is possible that in the future, the growth of aggregate GAA usage might necessitate some mechanism for rationing GAA access. This might induce interest in purchasing PALs to ensure access even if the interference protection afforded by a PAL is not desired. Note, these comments are agnostic as to the amount of spectrum that may be reserved for GAA access and how aggregate GAA access is managed.

Integrating this into the SAS/FCC payment collection framework would be relatively easy. The licensee would register its intent to exercise the option to exclude GAA users with the SAS and the FCC as of a certain date (exercise date). The SAS would be updated to reflect that the affected PAL spectrum was no longer available for GAA use; and the FCC would know that P2 is due and initiate collection subject to its payment/collection rules. This minor adjustment to the SAS represents a minimalist approach to making the SAS more dynamic (a goal of spectrum policy!) and the allocation of spectrum between exclusive PAL and GAA use more responsive to market forces (economic efficiency in spectrum management!), while keeping the practical elements that need to be adjusted as simple as possible (minimal implementation costs!).

Obviously, the above schema could be made more complicated to better match the splitting of payments to the value of the option (P1) and the value of its exercise (P2).[22] The apportionment factor (z, set above at 50%) could be adjusted either up or down to front or back-load payments and might even be split into more than two payments. The split in payments might be adjusted based on market data or some other basis that may better inform the relative apportionment of value.

Discussing such refinements is beyond the scope of these reply comments, but the basic idea for this approach originates in the law & economics literature discussing the relative merits of different legal regimes for protecting against harmful externalities arising from the exercise of property rights.[23] Spectrum access rights are a form of property right,[24] and getting the

---

[22] These are Real Options, which are closely related to Financial Options, which are a subclass of Derivative Securities. There is a wealth of theory and practical experience in how these may be designed to carefully tailor the payments to shape incentives. For a discussion of these, see for example, Dixit, A. and R. Pindyck (1994), *Investment under Uncertainty*, Princeton: Princeton University Press, 1994; Trigeorgis, L. (1996), *Real Options*, Cambridge: MIT Press, 1996; or, Wilmot, P. (1998), *Derivatives: the Theory and Practice of Financial Engineering*, New York: John Wiley & Sons. The benefits of this theory and experience may prove useful in informing industry stakeholders' valuations of the PALs, however, taking account of the many options afforded for the design of the PALs is not necessary in order to implement the basic and simple idea. Keeping things simple may be preferable, at least initially.

[23] For example, see Calabresi, G., & Melamed, A. D. (1972), "Property rules, liability rules, and inalienability: one view of the cathedral," *Harvard Law Review*, 1089-1128; Coase, R. H. (1960), "The Problem of Social Cost," *Journal of Law and Economics,* vol. 3 (October) 1; Kaplow, L., & Shavell, S. (1996), "Property Rules versus Liability Rules: An Economic Analysis," *Harvard Law Review, 109*(4), 713-790; Ayres, I., & Talley, E. (1995), "Distinguishing between Consensual and Nonconsensual Advantages of Liability Rules," *The Yale Law Journal, 105*(1), 235-253; or Smith, H. E. (2004), "Exclusion and Property Rules in the Law of Nuisance," *Virginia Law Review, 90*(4), 965-1049.

[24] Earlier folks tended to associate property rights with exclusively licensed spectrum and contrasted that with unlicensed spectrum that was modeled as a common. In truth, both are property rights regimes with different definitions of what constitute the relevant property rights, and between exclusively-licensed spectrum and unlicensed there are a continuum of potential sharing (rights) models. For further discussion, see for example, CFP (2014), "Toward More Efficient Spectrum Management," MIT Communications Futures Program (CFP) Spectrum Working Group White Paper, March 7, 2014, Notice of Ex Parte submission to GN Docket 12-354, available at http://apps.fcc.gov/ecfs/comment/view;jsessionid=y1ppTXCR9CN27jGyTHkGh9bwb6b0CHPLwJn25YJvLsWWvQ2xNl2S!-448120223!-58662085?z=dyzsz&id=6017604194.

assignment of such rights correct in order to induce appropriate incentives for allocating and using spectrum efficiently has long been recognized by economists and scholars interested in spectrum management reform.[25] A valuable insight from Professor Ayers is that options theory can augment property rights and liability rules to render them closer substitutes for enforcing efficient allocation of usage rights, especially in decision-making contexts with imperfect information (asymmetric, private, uncertain).[26] Managing interference in today's complex world of wireless offers just such a situation where the optimal assignment of radio resources depends on a mix of factors that may be unknown to regulators and may change over time with market conditions.

3. Benefits of Modified Payment Terms

Regarding PALs as real options (guaranteed access to PAL spectrum with interference protection, plus right to exercise right to exclude GAA operations in the PAL spectrum as of option exercise date) offers multiple benefits, including:

- More efficient spectrum usage and expanded access for commercial users;
- Encourages participation of PAL and GAA commercial users by enabling better matching of PAL costs with network investment requirements and by expanding access for GAA;
- Simple to implement so low implementation costs;
- Reduces potential risk of spectrum hoarding by PAL; and,
- Flexibility and consistency with dynamic shared spectrum future

Each of these benefits is discussed further in the following sub-sections.

3.1. More efficient spectrum usage and expanded access to commercial users

The key motivation for this proposal is to enhance incentives and prospects for efficient use of our radio frequency spectrum. By allowing flexibility in whether GAA use is allowed in the PAL spectrum, the PAL Option approach expands the economically viable spectrum accessible to both GAA and PAL users.

This is the overall goal of the PCAST report's recommendations and the FCC's efforts to define rules for the new CBRS in 3.5GHz. If successful, the benefits of this approach may be extended to other bands.

This approach affords greater scope for economic market forces to help direct the efficient use of spectrum resources, and thus helps reduce the need for more detailed technical rules and

---

[25] For example, see Coase, R. H. (1959), "The Federal Communications Commission," *Journal of Law and Economics,* 2 (October) 1-40; Hazlett, T. W. (1998), "The Law and Economics of Property Rights to Radio Spectrum: Introduction," *Journal of Law and Economics, 41*(S2), 521-522; Werbach, Kevin (2004), "SUPERCOMMONS: Toward a Unified Theory of Wireless Communication," *Texas Law Review*, 82 (March 2004) 863-973; or Goodman, E. (2004), "Spectrum Rights in the Telecosm to Come," *San Diego Law Review, 41*, 269-404.

[26] See Ayres, I. (2010) *Optional law: the structure of legal entitlements*, University of Chicago Press, 2010.

specifications that the FCC policy process and the requirements of dynamic wireless markets are ill-equipped to frame appropriately. As such, this approach is more consistent with the policy goals of service and technical neutrality, and the desire to minimize regulatory-induced distortions.[27] The CBRS rules should facilitate the commercialization of both PAL and GAA business models, and leave it to market forces to the extent possible, to determine the appropriate balance between these two models in the marketplace. Both spectrum usage models are important to different groups of commercial actors, and denying adequate spectrum resources to either type of user will make the CBRS less attractive to that class of actors. As explained further below, the PAL Option approach can improve incentives for both PAL and GAA users to participate.

3.2. Encourage participation of PAL and GAA commercial users

3.2.1. Better matching of PAL costs with network investment requirements

The PAL Option approach enables PAL users to better match their spectrum acquisition costs to their network investment and operational requirements. After securing their spectrum rights, it will take time for PAL users to build out their networks and their operational requirements for interference protection may evolve over time.

This is likely to be especially relevant for PAL users like mobile network operators that anticipate using the spectrum as part of a wide-area service (i.e., larger than a single PAL), which may have different requirements in different parts of their serving areas.[28] The PAL Option approach allows PAL users to lower their payments, thereby reducing the cost of acquiring spectrum. In effect, the option allows more granular management of the spectrum resources and is complementary to and akin to the benefits of smaller license territories.[29] By lowering the cost of acquiring spectrum for efficient PAL operators, the option approach should encourage greater commercial participation.[30]

---

[27] While regulatory interventions in markets may be necessary, they are often a second-best solution since regulators typically have less good information than market participants and their tools have limited flexibility. A goal of service/technical neutrality in regulatory design is to minimize the potential adverse impact of regulatory distortions and allow greater scope for market forces to select appropriate business models and technologies. While we may aspire for commercial neutrality in spectrum policy, this goal is not fully achievable and technical rules are an unavoidable and necessary component of any real world spectrum management regime. However, introducing market-based tools such as envisioned in the PAL option approach may help ameliorate the potential rigidities and problems associated with such technical rules.

[28] Spectrum and interference protection requirements will vary market-by-market based on customer density (rural/urban) and the spectrum assets that the mobile operator already has. The next generation of mobile systems based on 4G LTE are designed to flexibly manage and integrate diverse spectrum resources.

[29] See Lehr, W. and J. A. Musey (2013), "Right-sizing Spectrum Auction Licenses: the case for smaller geographic license areas in the TV Broadcast Incentive Auction," white paper, available at http://apps.fcc.gov/ecfs/document/view?id=7520959686.

[30] The presumption is that the value of the option to exclude GAA is based on the operational requirements of the PAL operator, and not on some other strategic motivation (e.g., to foreclose GAA

3.2.2. Expanded access for GAA

A key feature of the PAL Option approach is that it expands access for GAA spectrum beyond what would otherwise be available. A number of commenters have noted that GAA operation could occur in many parts of a PAL without causing any harmful interference[31] to the PAL user, either because the PAL user is not yet operating, is operating in a small portion of the PAL, or because the PAL user can tolerate co-existence with GAA in the same spectrum.[32]

The option to share PAL spectrum is separable from any efforts to reserve some minimum amount of 3.5GHz spectrum for GAA access as proposed in the *FNPRM*, but this expanded access may make it easier to tailor a solution for ensuring adequate spectrum access for both GAA and PAL users.

3.3. Simple implementation

As noted above, very little is needed to implement this approach. Fundamentally, it may be accomplished by a slight modification in the payment terms for implementing the PALs and readily incorporated via functionality already anticipated in all but the most conservative and least-dynamic visions of the SAS. At a minimum, the SAS is expected to be able to provide guidance to CBRS band users (incumbent, PAL, and GAA) on which frequencies are available in which license areas for use, and this database information needs to be updatable on a relatively coarse (not necessarily real or near-real-time basis) that is fully consistent with reasonable parameters for the timing of option exercise notification.[33]

Would-be PAL bidders may take advantage of the full wealth of option theory to craft their valuation and bidding strategies, but little additional complexity is required to implement the basic form of this idea as proposed here. However, this does not preclude more complex refinements, if desired. Ample ideas for some of the ways this idea might be expanded exist in the Options literature.

3.4. Reduced risk of under-utilized 3.5GHz spectrum

A concern is that PALs might be acquired that are subsequently under-utilized. This may be because the bidders have not yet deployed active systems, because they have acquired excess resources for future growth and peak traffic needs, or because they value having the spectrum be

---

competition by limiting GAA access to spectrum). As discussed further below, the PAL Option approach also offers benefits to the extent that such other strategic motivations might be a concern.

[31] For example, see Google Comments to *3.5GHz FNPRM* filed July 14, 2014.

[32] For example, the expected GAA usage may not pose a sufficiently significant risk of harmful interference for the PAL user or the PAL user's usage may be localized in time, geo-space, or direction.

[33] For example, it is certainly reasonable to require that exercising of the Exclusion Option have a required lead time for notification before becoming effective (unless invoked immediately) in which to accommodate the orderly closure and exclusion of GAA usage in the PAL. This lead-time may be measured in days, weeks, or longer; but need not be expected to take effect in very short time periods.

free of radio signals. This last reason may be because they are interested in the PAL spectrum for guard bands, or potentially, because they have a strategic desire to deny spectrum resources to potential competitors.[34] In each of these cases, different industry stakeholders may have conflicting notions of what constitutes under-utilized spectrum.

Indeed, if no one elects to build, deploy, and use PAL/GAA compliant devices, the spectrum will remain under utilized and the FCC's efforts to create a viable CBRS will be considered a failure. However, it remains uncertain how utilization of the CBRS spectrum will proceed.[35]

The PAL option approach provides an economic incentive to the licensee not to preclude GAA utilization of the spectrum unless there is a real economic value to doing so (although that value may be a private value to the licensee). By opening additional spectrum for GAA users, the PAL option approach helps promote GAA usage.

Additionally, the PAL option approach is superior to other approaches that might be considered for ensuring spectrum is utilized. First, it is important to point out that the technical utilization of spectrum is often not the best metric of whether the level of spectrum utilization is economically or socially optimal. For example, radio telescopes that are trying to extract weak signals from space prefer spectrum with very low noise from other transmitters. This is true for other radio services and system designs as well. Prospective customers and investors for wide-area systems may demand that sufficient excess spectrum resources be available in advance of need and in anticipation of peak traffic demands (i.e., appear under-utilized from the perspective of real-time measurements of spectrum occupancy) to justify investing in or subscribing to the commercial services that are expected or already operating in the spectrum. Regulatory rules that embed build-out requirements, technical utilization (spectrum occupancy) standards, or otherwise seek to manage spectrum utilization directly are likely to be hard to implement effectively.[36]

The PAL option allows PAL and GAA users more scope to design their systems and services, and to differentiate their business models to compete more aggressively in wireless markets.[37] Some wireless system operators may choose to invest more in small cells and more capable radio systems that enable higher spectral efficiency, while other operators may use more spectrum

---

[34] Whether it is even economically rationale to hoard spectrum resources as suggested in Note 30 *supra* is debatable, however, this is a concern that has been voiced by participants in spectrum reform discussions.

[35] At this stage, there remains significant uncertainty regarding which uses/business models/technologies will be most successful; how fast traffic will grow; how much co-existence between the three tiers of users will be economically viable; etcetera.

[36] Build-out or technical occupancy rules with too much detail and specificity are likely to over-constrain PAL user flexibility in choosing technologies, wireless architectures, or business models; and rules that are too loose may be ineffective.

[37] This is analogous to the differentiated offers for broadband service from DSL resellers. They may differentiate their services based on their offered quality-of-service (jitter, latency) and its variability. A service that is under-provisioned in the sense that it is more vulnerable to congestion (dropped packets, increased latency during peak periods) may be less expensive. Which is the more attractive broadband service (taking into account price and service quality) may depend on who the customer is (e.g., and their willingness to pay for higher quality service).

resources to economize on efficiencies elsewhere in the overall system. Ultimately, the goal of efficient spectrum usage is to maximize the value of the wireless services that make use of the spectrum, not to maximize the energy in any particular band.

By helping to expand access to economically viable (lower cost, additional) commercial spectrum for PAL and GAA users, the PAL option approach helps promote competition that is not biased in favor of a particular business model. *Ceteris Paribus*, this offers the most effective mechanism for deterring spectrum hoarding strategies that might be designed to limit competition.

3.5. Flexibility and consistency with dynamic spectrum access

As recognized in the *PCAST Report*, enabling continued growth in wireless services of all kinds requires expanding commercial access to spectrum resources. It is neither feasible nor economically desirable to meet all of this demand with new allocations of dedicated spectrum. Fortunately, the design of modern wireless systems makes it unnecessary to rely on dedicated spectrum resources. The future of wireless is to share spectrum much more intensively, and there are many technologies, business models, and regulatory regimes that make this possible.[38] For example, a key feature of the 4G LTE technology that mobile network operators are currently in the process of adopting around the world is its ability to provide fine-grained, dynamic control of spectrum frequency resources.

Advances in smart networking and radio systems, including in such technologies as cognitive and software defined radios, signal processing, and advanced Internet architectures[39] are enabling a more capable wireless future. These advances will help support new services (for public safety, for Internet of things, for smart grids, for mobile cloud computing, etc.) and will help accommodate the exponential growth in wireless traffic that is expected.

Realizing this future depends on designing systems that are more flexible. A key goal of spectrum management is to decouple radio services and higher level user services from the underlying spectrum resources to make such services more robust, flexible, and capable. It is already the case that wireless telephony is no longer bound to specific frequencies. This unbundling of spectrum resources, long-lived spectrum infrastructure (like towers and even base station hardware radios), and wireless services enables more scalable investment, expanded options for mix-and-match ways to provision and differentiate services, and increased opportunities for competition.

The PAL option approach is consistent with this trend toward a more dynamic, flexible, and shared spectrum management regime. In the future, it is not inconceivable that spectrum

---

[38] Exclusive licensed, unlicensed, and the new three-tiered regime planned for the 3.5GHz CBRS band are all spectrum sharing regimes that allow spectrum resources to be dynamically allocated to different users/uses on a granular basis (in time, power, space, code, etcetera). See CFP (2014), note 24 *supra*.

[39] For example, the author is part of the research team involved in the NSF-funded MobilityFirst Future Internet Architecture project that is focused on expanding mobility support in the Internet (see http://mobilityfirst.winlab.rutgers.edu/).

resources might give rise to a rich selection of financial securities for managing risk and allocating value across investors, system operators, and end-users. This might be analogous to the range of financial securities associated with other of the raw commodity resources on which our economy depends (e.g., oil, electric power, grains, etc.). This securitization and commoditization of basic resources is a key component of our global trade and economic production systems.[40]

The PAL option approach is a simple first step toward building a more robust SAS ecosystem that is consistent with such an evolutionary path, and helps "future-proof" the regulatory process (without committing anyone to the bolder vision such a trajectory implies).

4. Potential Challenges and Issues to Resolve

The purpose of this reply comment is to suggest a simple change to the interpretation of PALs that I believe would offer a range of important benefits, however it is likely that others may be able to suggest additional enhancements or refinements to this concept that are worth addressing. And, there are likely to be a range of challenges to adopting the PAL Option approach and questions that would need to be addressed before moving forward. In this section, I briefly touch on a few obvious issues that might be of concern.

4.1. Feasibility of Enforcing GAA Exclusion Option

A key concern for all users of the CBRS spectrum is the efficacy of the SAS in protecting compliant access. Incumbent and PAL users with explicit interference protection rights are justified in being concerned that those rights are honored. Similarly, investors and users of GAA devices also have a vested interest in ensuring that whatever framework is eventually adopted provides the spectrum access they can reasonably expect.[41] In Chapin & Lehr (2007), we explained how important it is for the commercial success of dynamic spectrum access for stakeholders to trust the sharing regime.[42]

The PAL Option approach is consistent with the most conservative and easiest to implement version of the SAS-managed 3-tiered scheme being proposed for the CBRS band. If it proves infeasible to effectively switch PAL spectrum from allowing GAA operations to excluding GAA operations, then it seems implausible that any sort of GAA exclusion or shared operation would be possible. Consequently, these comments assume (and the author believes that is a reasonable

---

[40] The securitization of raw commodities allows businesses to manage supply and demand risk, trading off forward and spot purchases/sales to best accommodate their special needs. Options are a well-understood tool.

[41] Because users of GAA spectrum will not have a legal right for interference protection from other legal users of the CBRS spectrum and will have an obligation to avoid causing interference to Incumbent or PAL users with protected access rights, the GAA spectrum users access will be uncertain. An effective SAS regime will help reduce the uncertainty and its attendant investment costs.

[42] See Chapin, J. and W. Lehr (2007), "The path to market success for dynamic spectrum access technologies," *IEEE Communications Magazine*, May 2007.

assumption) that the enforcement of the PAL exclusion option would be practically and technically feasible as described here.

4.2. Impact on Commercial Access Licensing Revenues

The current proposal is to auction PALs. Because these are anticipated to be for one year and based on Census Tracts (or potentially smaller geographic units), this will entail auctioning many more spectrum licenses, each of which will be less valuable (per license), than has been tried previously.

Some may be concerned that allowing PAL bids to be split into two payments, P1 and P2, will result in lower total license (auction) proceeds being collected. This is certainly a possible outcome. However, it is also possible that implementing the PAL option may increase license proceeds. This might occur if the PAL option encourages participation from PAL and GAA users that results in more confidence (higher future expectations of growth) and self-fulfilling faster growth of a rich commercial ecosystem of CBRS users (devices, applications, and services). Anticipation of such a larger market could induce higher valuations in PALs today.

Whether this is true or not, however, seems like the wrong question. The goal should be to promote efficient spectrum usage, not raise spectrum proceeds. Even without this proposal, it is conceivable that demand for PALs will be limited and the price per PAL and number of PALs purchased will be small. This would leave more spectrum available for GAA users. Alternatively, it is conceivable that demand for the spectrum will be strong and that concerns over acquiring adequate interference protection sufficiently intense that PALs will be scarce and prices may be relatively high (although presumably significantly less expensive on a MHz-POP basis than what is expected in the broadcast spectrum incentive auctions). Uncertainty over the value of PALs and concerns over the relative scarcity of PAL and GAA spectrum may deter ex ante investment in developing business models and technologies to invest in using the spectrum. The PAL option approach provides a market-based way to help manage this scarcity better.

4.3. Threat of Option Allowing Inexpensive Mechanism for Restricting GAA Access

Another potential concern might be if PAL prices are sufficiently low that it is inexpensive for bidders to warehouse spectrum with initial P1 payments, and the mere threat of later invoking the GAA exclusion option proves a sufficient deterrent to investment in GAA devices that the GAA ecosystem of devices and services fails to succeed. Under this scenario, enabling the PAL Option might enable a lower cost way for potential bidders to foreclose GAA access and usage.

On further reflection, this scenario does not seem overly concerning. First, future radio technologies will need to tune to all of the frequencies in the band, with the SAS indicating which frequencies are available in which license areas. Second, the FCC has committed to ensuring at least a minimum level of spectrum access for GAA devices (even if GAA access to PAL spectrum is fully excluded, which could be the case without the PAL option or with the PAL option if all licensees choose to exercise immediately). Thus, if the GAA model works anywhere, it should only be more viable and work better if the PAL Option enables additional GAA spectrum resources.

At this point, it is unclear what the appropriate price for a PAL should be, although the hope is that because aggregate spectrum scarcity should be reduced, the restrictions on PAL usage rights, the short license terms, and small geographic areas, that PAL prices will be low and easily affordable for both large mobile network operators and small users like hospitals.

4.4. Mispricing of Payments

Closely related to the above concern is the concern that the mechanism for splitting the payment between the initial option purchase price, P1, and the exclusion exercise price, P2, will fail to appropriately mimic the economic value of the two rights.

The proposal to split the payment in half is obviously ad hoc but has the advantage of being simple and relatively neutral in its assumptions. More complex mechanisms might be suggested and future adjustments using market or modeling data could be utilized.

At a minimum, this approach implies a non-trivial benefit for postponing exercising the exclusion option until it proves sufficiently important to the PAL licensee. Whether it provides too great or too little an incentive to invoke exclusion will depend on both the mechanism used to allocate payments and the aggregate expected payments.

4.5. Nits and Details

In addition to the above questions/challenges, I can imagine a host of other nits and details that will need to be addressed. To suggest the range of possible questions, but not to try and identify all such questions, here are a few:

4.5.1. Exclusion on subsequent license terms and Reversibility

How does exercise of the exclusion option in a year impact the option to exclude in future year licenses? For example, if the licensee purchases a PAL for year 1, 2, and 3 and executes the exclusion option in year 1, does that mean that the exclusion option for year 2 and 3 are assumed executed at the commencement of the new term? Is it possible to re-introduce GAA sharing in PAL spectrum once the exclusion has been invoked? The choice of how to treat sequential licenses will impact initial valuations regardless of whether the PAL option approach is adopted or not, and the PAL option just provides a more nuanced way to think about this problem which needs to be addressed in any case.[43]

Some might argue that the types of PAL systems that are most likely to choose to invoke the GAA exclusion option are likely to be long-lived and the need for exclusion would be

---

[43] For example, assuming that the exercise price is fixed at ½P (as proposed earlier), then the value of exclusion might be decreasing as the license approaches termination (assuming that the value of exclusion is constant per unit of time) or increasing (e.g., if congestion is increasing sufficiently rapidly), which may make it more or less likely that an option would ever be exercised after some proportion of the license were past. The treatment of exclusion rights on subsequent licenses to the same spectrum would impact this calculus, but similar sorts of concerns apply to bidding on multi-year licenses and are not unique to the PAL option approach.

monotonically increasing in the level of traffic, and hence the desire to reverse the exclusion in subsequent periods would be a rare occurrence.

On the oft chance that that is not the case, Professor Ayers "option law" framework posits creating both call and put options to allow for more dynamic flexibility in how resources are allocated.[44]

4.5.2. Separate trading of Option and Spectrum access rights

Some might wonder about whether it should be possible to separately acquire or trade the PAL access option and exclusion options. Potentially something along the lines of a carbon trading scheme might be possible, that would allow users to trade or acquire exclusion rights to create resources for GAA access.

Certainly in real world markets, such repackaging of rights is both feasible and generally regarded as contributing to overall efficiency. Figuring out how this might work and impact the management of sharing in the CBRS is beyond the scope of this comment, but at least initially, this might add to implementation cost.

4.5.3. Other stuff

There are likely to many more nits and details that will occur to folks, some of which may be easily answered and others that may require more thought and further adjustments.[45] Fundamentally, the PAL Option approach is intended to be a simple first step to inject a helpful and scalable tool for introducing economic incentives into the CBRS management framework.

Its adoption will impact the need and efficacy of other aspects of the CBRS framework (e.g., how much spectrum to reserve for GAA? whether GAA operation in PALs should be permitted at all? What technical rules or build-out requirements are needed to promote efficient spectrum utilization? Etc.). While the intent is to propose a modification that either complements or offers a better substitute for alternatives under discussion, its full evaluation will depend on a more complete articulation of the CBRS rules framework (length of licenses, territory size, power limits, etc.).

It is precisely in this context of a partial and evolving set of rules that introducing economic incentive mechanisms may prove most valuable.

---

[44] See Ayres (2010), note 26, *supra*.

[45] For example, as presented, PAL licensees may invoke their exclusion option individually. What are the interaction effects across PAL spectrum on the SAS and on PAL and GAA users in adjacent spectrum? Answering this question may depend on the granularity of the licenses and SAS management capabilities.

5. Concluding remarks

These reply comments offer the suggestion that the FCC modify its payment rules such that PALs are viewed as a real option, such that the payment is split into two parts: P1 for the right to use PAL spectrum and the interference protection it affords, which is owed upon the award of the license; and P2 to exercise the option to exclude GAA use of the spectrum from the time the option is exercised for the remaining duration of the license. Until the PAL option to exclude is exercised, GAA is permitted in the PAL spectrum.

This PAL Option approach provides a simple way to introduce economic incentives and market forces into the CBRS framework for managing shared commercial access in the 3.5GHz band. This will enhance economic incentives of both PAL and GAA users to participate in the band and use spectrum efficiently, will promote competition, and is consistent with the long-term trajectory for the evolution of wireless services and spectrum usage.

6. Attachment 1: About the author

Dr. William Lehr is a telecommunications/Internet industry economist and policy analyst with over twenty years of experience in academic research and industry consulting.[46] He is currently a research scientist in the Computer Science and Artificial Intelligence Laboratory (CSAIL) at the Massachusetts Institute of Technology (MIT). Dr. Lehr's research focuses on the economic and policy implications of broadband Internet access, next generation Internet architecture, and the evolution of wireless technology. Dr. Lehr has written extensively on spectrum policy matters and advised policymakers in the U.S. and abroad on wireless and spectrum management issues. Dr. Lehr was an invited expert participant in the PCAST report.

Dr. Lehr holds a PhD in Economics from Stanford, an MBA in Finance from the Wharton School, and MSE, BA, and BS degrees from the University of Pennsylvania.

---

[46] For more information, please visit http://people.csail.mit.edu/wlehr.

DRAFT – PLEASE DO NOT CITE OR DISTRIBUTE

COMPETITION BY PENALTY DEFAULT

Kristelia A. García*

In music licensing, powerful music publishers have begun – for the first time ever – to withdraw their digital copyrights from the collectives that license those rights, in order to negotiate considerably higher rates in the private market. At the beginning of the year, two of these publishing companies commanded a private royalty rate nearly twice that of the going collective rate. One view of this result is as a coup for the free market: constrained by consent decrees and conflicting interests, collectives are simply not able to formulate and command a true market rate. Another view of this result is as a pathological form of private ordering: powerful licensors using their market power to coerce an extortionist rate from a hapless licensee. While there is no way to really know what the "market rate" looks like in a highly regulated industry like music publishing, the anticompetitive consequences for artists, licensees, and consumers alike are self-evident. In industries with a tendency toward natural oligopoly – such as telecommunications and content licensing – network effects, parallel pricing and tacit collusion can work to eliminate meaningful competition from the marketplace. The resulting lack of competition threatens to stifle innovation in both the affected, and related, industries. Contrary to conventional wisdom, this Article posits regulation – not antitrust – as the optimal means of opening the market and maintaining competition.

Normally, where a market can operate in a workably competitive manner, the remedy for anticompetitive behavior can be found in antitrust law. In music licensing, however, some concerning behaviors – including both parallel pricing and tacit collusion - do not rise to the level of antitrust violation; as such, they cannot be addressed by antitrust law. This is no small irony. At one point, antitrust law served as a check on the power of the collective licensing entities by establishing a consent degree that governed their behavior. Due to consolidation of content, the collectives that are being circumvented – but whose conduct is governed by consent decree – now pose less of a competitive concern than do powerful copyright holders acting privately. The case of intellectual property, where regulation defers competition for creators and inventors for a limited period of time, is particularly challenging for antitrust. The paradoxical solution, I propose, is more regulation – but not just any regulation.

In prior work, I have examined the phenomenon of private ordering around an unpalatable statutory license, or "penalty default license," as the inevitable result of inherent inefficiencies and bounded uncertainty. Similarly, the proliferation of private ordering in music licensing – where rights are traditionally licensed by collectives – challenges the conventional view of collectives as the optimally efficient licensing mechanism, and instead demonstrates that they suffer many of the same inefficiencies as statutory licenses. While regulation is conventionally understood to restrict new entry and to interfere with competition in the marketplace, this Article will posit that a certain type of regulation – specifically, penalty default regulation that works to penalize the absence of competition – can actually work to open markets and to maintain competition in the marketplace.

---

* Associate Professor, University of Colorado Law School; Director, Silicon Flatirons Center for Law, Technology, and Entrepreneurship.

Using the recent rate court proceeding between Pandora and ASCAP as a starting point for the analysis, this Article suggests that regulation can actually encourage competition. In much the same way that penalty default rules encourage more efficient information exchange between otherwise unequal parties, and penalty default licenses encourage more efficient private ordering between otherwise unequal parties, this Article suggests that penalty default regulation – i.e., regulation that penalizes a lack of competition – can serve the same market-opening function as antitrust. While not without its drawbacks – primarily, an increase in the cost of private action – penalty default regulation in music licensing corrects anticompetitive behavior while ensuring ongoing access to content, continued innovation in distribution, and fair payment to artists. This is particularly relevant in such current debates as net neutrality, where regulation is working to maintain an open Internet, apart and and away from the control of powerful incumbents.

INTRODUCTION

On March 30, 2015, Jay-Z and 16 of his closest artist-friends came together at a press release event in New York City to announce what was widely billed as a "revolutionary" new music streaming service. The service, called Tidal, was touted as artist-owned, and promised to pay artists more than other streaming services. The pesky logistics around Tidal's business model were foregone in favor of such random displays as Madonna propping a leg across a table while she signed a mysterious declaration, and Alicia Keyes quoting Nietzsche, but this much is clear: The only way Tidal is going to get $19.99/month from customers is if the participating artists – which appear to include Jack White, Beyonce, Daft Punk, and Kanye West, among many others – actually withdraw their content from all other streaming services, thereby allowing them to create scarcity and command a higher profit. If this seems undesirable from a consumer perspective, we should recall that this result is precisely what intellectual property ("IP") does: it protects the rights holder from competition for a limited period of time.[1] The real concern – and the subject of this Article – is the potential for stifling innovation in the burgeoning music streaming industry, and for locking us into the current technology. This, I will argue, is not what IP does (or at least not what it should do). In industries with a tendency toward natural oligopoly – such as music licensing – network effects, parallel pricing and tacit collusion can work to eliminate meaningful competition from the marketplace. The resulting lack of competition threatens to stifle innovation in both the affected, and related, industries. Contrary to the conventional understanding of regulation as a means of restricting entry and thwarting competition, this Article posits that regulation – and not antitrust – can actually encourage competition in intellectual property.

From a regulatory (if not a contractual) perspective, the withdrawal scenario that I've posited is a real possibility. Take the recent ASCAP withdrawals, for example. At the beginning of the year, two of the nation's three major music publishers[2] withdrew their digital copyrights from ASCAP, [3] the collective that administers those rights, in order to license them privately in the market. Presumably, publishers and songwriters had not attempted withdrawal of these rights before for a couple of reasons: First, because they weren't allowed to (ASCAP's governing documents didn't allow for it); and second, because they didn't want to. The idea that a music

---

[1] Concededly, not a limited enough time for some. See, e.g, __.

[2] A "music publisher" is a company that owns the copyrights on various musical compositions, and licenses the use of those compositions to such entities as radio stations, sports stadiums, filmmakers, restaurants, and record labels – traditionally through a collective, but more recently via private negotiation.

[3] ASCAP is a "performance rights organization," or "PRO." PROs are specialized collectives that have handled the collection and administration of so-called public performance royalties – i.e. royalties incurred from the play of a song on terrestrial or digital radio – for the last century.

publisher might forego the ease and convenience of collective licensing of its IP rights in order to do all of the work itself is counterintuitive. Conventional wisdom justifies collectives first on the basis of their ability to reduce transaction costs, and second, for their consolidation of bargaining power. The purported circumvention of ASCAP by two of the major music publishers undermines these justifications, and suggests that they no longer hold sway in the new, digital age. Instead of requiring hundreds or thousands of individual negotiations, a digital radio service today can obtain upwards of 80% of all of the music publishing rights that they need from a mere three companies. And if those three companies cooperate on rate – explicitly or tacitly – the remaining rights holders will eventually be forced to take that rate as well in order to remain competitive.

The digital rights withdrawals required multiple amendments to ASCAP's governing documents, effectively removing the withdrawing publishers from the auspices of the consent decree that governs ASCAP. This marked the first ever attempt at withdrawal and private ordering in the music publishing industry, and resulted in a negotiated royalty rate nearly double ASCAP's going rate.. One view of this result is as a coup for the free market: constrained by an antiquated consent decree and faced with conflicting member interests, ASCAP artificially depresses the "market" rate for music licensing. Another view of this result as a pathological form of private ordering in which two powerful companies wielded their considerable market power to coerce a supra-competitive rate from a hapless licensee.

Unfortunately, we can't know what the market rate looks like in a regulated industry like music publishing because there has never been a true "market" for music licensing in any economically meaningful sense.[4] For this reason, this Article takes no issue with the rate obtained, but rather with the potential for unchecked anticompetitive consequences for artists, licensees, and consumers. Notably, the two music publishers at issue – Sony/ATV Music Publishing, LLC ("Sony/ATV") and Universal Music Publishing Group ("UMPG") – own roughly 50% of all digital music copyrights. Without this content, an Internet radio service would be hard pressed to compete for listeners. For this reason, one such service, Pandora, challenged the privately obtained rate as coerced. Further, the withdrawal wasn't necessary from a licensing perspective: ASCAP's license to the publishers' content is non-exclusive, meaning the publishers could have readily competed with ASCAP in the market to license their songs without withdrawing. But the withdrawal worked to reduce the field of competition – afterward, there was only one option for licensing that content, and that was through that publisher itself, sans consent decree.

In prior work, I've offered qualified encouragement of private ordering as a means of increasing efficiency in the statutory licensing context. This Article reaches a different conclusion in the collective context, where collectives are governed by

---

[4] While it's true that a songwriter can technically license her own songs, this is not logistically possible (at least in an analog world), and so everyone uses the collectives. The establishment of the industry arose more or less simultaneously with the creation of the collectives.

consent decree, while powerful individual rightsholders are not. Currently, the two largest collectives – ASCAP and BMI – operate under consent decrees that govern their operation. These consent decrees originated from concerns about anticompetitive behavior – including supra-competitive pricing and barriers to entry – that might result from the consolidation of a large number of copyrights in only a handful of entities. These concerns apply equally to individual music publishing companies, which also consolidate copyrights en masse, but the consent decrees do not.[5]

Normally, where a market can operate in a workably competitive manner, the remedy for anticompetitive behavior can be found in antitrust law. In music licensing, however, some concerning behaviors – including both parallel pricing and tacit collusion - do not rise to the level of antitrust violation; as such, they cannot be addressed by antitrust law. This is no small irony. At one point, antitrust law served as a check on the power of the collective licensing entities by establishing a consent degree that governed their behavior. This discrepancy has resulted in a situation where the collective that is being circumvented now poses less of a competitive concern than do individual, powerful copyright holders acting privately. This counterintuitive development is due, in part, to the inapplicability of extant consent decrees to individual copyright holders (even where, as in the case of the withdrawing publishers, the "individual" holds hundreds of thousands of copyrights). Network effects – that is, [ ] – also play a role, and have exacerbated the effects of consolidation in the music industry by removing the threat of meaningful competition from the marketplace. This allows individual companies – acting alone or in tacit collusion with similarly situated competitors – to engage in parallel pricing of performance rights, or to withhold rights altogether, thereby potentially barring entry to prospective licensees, all while avoiding antitrust scrutiny. The case of intellectual property, where regulation defers competition for creators and inventors for a limited period of time, is particularly challenging. It is this unique potential for anticompetitive behavior that does not meet the definition of antitrust "harm" – and therefore is not addressed by that body of law – that is the focus of this Article.

In prior work, I have examined the phenomenon of private ordering around an unpalatable statutory license, or "penalty default license," as the inevitable result of inherent inefficiencies in the statute, and bounded uncertainty.[6] Similarly, the proliferation of private ordering in public performance rights licensing – where rights are traditionally licensed by collectives – challenges the conventional view of collectives as the optimally efficient licensing mechanism, and instead demonstrates that they suffer many of the same inefficiencies as statutory licenses. While regulation is conventionally understood to restrict new entry and to interfere with competition

---

[5] [See Sharon's note on "regulation by consent decree" for an interesting footnote.]

[6] *See generally* Kristelia A. Garcia, *Penalty Default Licenses: A Case for Uncertainty*, 89 N.Y.U. L. REV. 1117 (2014).

in the marketplace, this Article will posit that a certain type of regulation – specifically, "penalty default regulation" that penalizes the absence of competition – can actually work to open markets and to maintain competition in the music licensing marketplace.

Using the recent rate court proceeding between Pandora and ASCAP as a starting point for the analysis, this Article suggests that regulation can actually encourage competition. In much the same way that penalty default rules encourage more efficient information exchange between otherwise unequal parties, and penalty default licenses encourage more efficient private ordering between otherwise unequal parties, this Article suggests that penalty default regulation serve the same market-opening function as antitrust.

So why should we care about all of this anticompetitive behavior on the part of music publishers anyway? Because it represents a very real threat to innovation in both the creation and distribution of content. To start with, if publishers go after the most successful companies in a space – like they did with Pandora in the music streaming space – we run the risk of a situation in which the companies that control an essential input (in this case, music) threaten to stifle innovation. In addition, these content owners can also block access to the essential input altogether. Without a statutory license, we no longer have guaranteed access to content, and so nothing to stop a content owner who wishes to start its own distribution company from denying its content to competitors. Perhaps most importantly, without a statutory license that guarantees artists a cut of royalties,[7] creators are potentially denied a cut of private deals, with predictable consequences for incentives.

There are several possible responses to these anticompetitive concerns: do nothing, rely on antitrust, or look to penalty default regulation. Doing nothing credibly threatens to stifle innovation in what Professor Timothy Wu has called "copyright's communication policy,"[8] or, the industry for the distribution and dissemination of copyrighted content. In a burgeoning market such as digital streaming, this has the added potential of locking us into a subpar technology.

As for antitrust, it has proven largely impotent in the music licensing context: Neither parallel pricing nor tacit collusion (as opposed to actual collusion) constitute violations of the Sherman Act, yet both carry significant competitive downsides. While parallel pricing is not in and of itself against the antitrust laws, it is indicative of a highly concentrated market. In addition, the consent decrees under which the collectives operate cannot be extended to cover individual publisher behavior. The FTC, through its repeated determination not to intervene in mergers that have reduced the music publishing industry to a mere three entities, has effectively foreclosed a finding of anticompetitive behavior on behalf of the individual publishers: ASCAP's consent decree requires all grants from rights holders be made on a non-exclusive basis, thereby allowing a party to opt out and negotiate privately.

---

[7] E.g., 114(g)(2)

[8] [cite]

This assumes that the individual publisher, acting alone, does not have the ability to act anticompetitively.

Of course, this assumption is wrong. The nation's largest music publisher, Sony/ATV, currently enjoys roughly 30% market share, and may well be able to exercise unilateral monopoly power in some circumstances. When working in tacit collusion with Universal Music Publishing Group ("UMPG"), the combined market share is raised to nearly 50%. Unfortunately, antitrust law does not provide a remedy for breaking up monopolist (or oligopolist) firms unless and until they engage in predatory conduct. Even if one firm is not found to be monopolistic on its own, two or more firms may tacitly collude to set prices, or to bar entry to a new service by denying content altogether. It is well established that antitrust law does not address the oligopolist problem of tacit collusion. This is why the maintenance of structural competition is so important. Where it is not maintained – as in the case of the music industry – it cannot be rebuilt. This means that even if the FTC were to recant its position, the mergers cannot be undone, nor can the market realities that exacerbate the effects of consolidation in the music industry.

In part, this is because the music industry is a two-sided market: it is both a "media" industry and a "platform" industry. As a media industry, the music industry supplies cultural capital in the form of songs. As a platform industry, the music industry connects artists, distributors, and consumers through a variety of technologies, including smartphones and online streaming services. The supply-side network effects inherent in music licensing further amplify the effects of consolidation by removing the threat of meaningful competition from a market that is already highly concentrated. In short, contemporary competition policy in copyright fails because it assumes robust competition despite all positive indications to the contrary.

Some commentators have suggested that antitrust's impotence in IP owes also to its overly narrow analysis; it ignores "the marketplace of ideas."[9] There is also some evidence that the existence of a regulatory scheme – such as we have in copyright – reduces the role of antitrust, not least of all because intellectual property regulation protects creators and inventors from competition for a limited period of time.[10] In addition, antitrust cases are notoriously difficult to prove, and conventional antitrust mechanisms hold no promise in a natural oligopoly environment. Even if we had conduct rising to the level of an antirust violation, it has been further suggested that antitrust is especially ill-suited to high-tech industries like music licensing where antitrust enforcers' ability to understand and predict industry evolution are especially limited.

In lieu of antitrust, this Article advocates for the utilization of penalty default regulation to produce the same market-opening and competition-maintaining effects that antitrust does. There is precedent for this model. Telecommunications has

---

[9] [cite]

[10] we see this now in the net neutrality debate

recognized the potential for pathological private ordering in oligopolistic markets. Regulation in telecommunications assumes a baseline that tends toward natural oligopoly, and so allows for private ordering only where robust competition can first be shown. Otherwise, regulation operates to ensure ongoing access to the relevant input(s) for all prospective consumers or licensees able and willing to meet the statutory requirements and to pay the statutory rate. Because this regulation does not necessarily represent a market rate – or, indeed, as high a rate as private ordering might obtain – it acts as a penalty default. If a company wants to engage in private ordering, it must first petition to show the existence of robust competition in the marketplace. While such "penalty default regulation" cannot create a robust competitive market where none exists, it can prevent a few powerful firms from unduly raising the price for an essential input, or from barring new entry, to the disadvantage of both consumers and innovators in the space. As with the telecommunications industry, the underlying assumption here is that the government has a greater responsibility for checking anticompetitive behavior in the music licensing industry owing to its role in the granting of exclusive property rights via copyright.

In the music licensing context, penalty default regulation does three things: First, it converts existing, circumventable statutory licenses – i.e., [Sections 114 and 115] – into mandatory (i.e. non-circumventable) statutory licenses. Second, it establishes a new, mandatory statutory license for the public performance rights that are currently handled by collectives. Those collectives can either cease to exist, or one (or more) of them may be designated as an official collective for the new statutory license for performance royalties (similar to SoundExchange's designation as official collective for sound recording royalties). Finally, it establishes a procedure through which a rights holder wishing to forego the statutory license and engage in private ordering, may petition to do so via a showing of robust competition in the market for music licensing. This structure reflects the legislative intent behind the interconnection agreements in the Telecommunications Act of 1996 (the "Telecommunications Act").[11] Building on existing literature in the areas of copyright, antitrust, regulation, and behavioral economics, this Article advocates utilization of penalty default regulation to maintain competition in industries where antitrust does not operate.

To be clear, I do not tread lightly into a recommendation for regulation, in music licensing or elsewhere. In addition to increasing the cost of private ordering, regulation – even penalty default regulation – forecloses the development of a true free market. This can lead to regulatory gaming, and to lobbying-related abuse. There is also the problem of statutory rate-setting. Ramsey pricing suggests that price discrimination (if that is indeed what we have here) may actually be efficient. Who is a governmentally-authorized entity like the Copyright Royalty Board ("CRB") to set rates anyway? Based on what? And yet, in addition to ensuring access to content, and curbing anticompetitive behavior by private parties, penalty default regulation is the

---

[11] 47 U.S.C. §[5 - __]

only means of ensuring access to content for all licensees and of guaranteeing direct payment to all artists.[12]

The analysis proceeds in four Parts. Part I begins with a music licensing primer, then turns to an analysis of the recent rate court proceeding between Pandora and ASCAP to illustrate the emergence of a pathological form of private ordering in an industry historically dominated by collective licensing. Part II discusses the paradox revealed by this narrative; namely, the fact that the licensing collectives that are being circumvented – once a cause for antitrust concern due to their consolidation of content, but whose conduct is now governed by consent decree – now pose less of a competitive concern than do powerful copyright holders acting privately. It describes the rise and fall of the PROs as the preferred means of licensing copyrighted content. Part III makes the case for music as unique among other forms of content in order to explain the failings of antitrust in music licensing. Borrowing from prior work on penalty defaults, Part IV suggests that contrary to the conventional view of regulation as competition-reducing, penalty default regulation can encourage competition in a way that antitrust cannot.

While not without its drawbacks – primarily, an increase in the cost of private action – penalty default regulation in music licensing corrects anticompetitive behavior while ensuring ongoing access to content, continued innovation in distribution, and fair payment to artists. It may well be the case that technology will eventually obliterate the need for statutory licensing of content altogether, much as it has already done in the context of landlines, but in the meantime, the proposed model improves greatly on the status quo. The Article concludes by suggesting ways in which penalty default regulation can work to open market entry, and to keep undue power away from incumbents. This is particularly pertinent in such current debates as net neutrality, where regulation that is keeping the Internet open to new entry, and away from the control of powerful incumbents.

---

[12] I intentionally sidestep the ongoing debate about copyright's status as a form of regulation versus as a property right. For the limited purposes of content licensing, at least, the role of regulation is undeniable. Not only do the copyright laws contain mandatory statutory licensing for cable and satellite television, but they also ultimately entrust royalty rate setting to governmental agencies, including the CRB, and rate court. In this context, a rate court is a court with exclusive jurisdiction to oversee licensing arrangements between a PRO like ASCAP, and potential licensees. The rate court gets its jurisdictional authority from the consent decree. Among other challenges raised, the wisdom (and legality) of delegating a single court – and indeed, a single judge – as the sole arbiter of rates for an entire industry has been questioned. [cite]

To participants at the Cable Academic Workshop:

Thanks for your willingness to review and provide feedback on this project. This paper is still in its nascent stages. My goal, which is probably too ambitious, is to put the FCC's tentative steps toward regulating IP network interconnection into historical context, identify potential regulatory pitfalls based upon the agency's prior efforts in the telecommunications industry, and suggest a framework under which it might accomplish its regulatory objectives without overreaching.

For over two decades, interconnection markets have largely flourished in a mostly unregulated environment. But over the past two years, disputes between networks over the terms of traffic exchange have grown, both in quantity and in visibility. This increase, driven in part by the rise of IP-based video and the consequent disruption this growth has caused to existing IP traffic flows, has prompted calls for various regulatory responses designed to minimize the risk of anticompetitive behavior and the negative effects of these disputes to consumers. In its 2015 Open Internet Order, the FCC indicated that it would provide limited review of interconnection agreements, as part of its overall responsibility to police against unjust and unreasonable practices and to assure that broadband providers fulfill their obligation to provide consumers with access to all or substantially all Internet endpoints.

This paper will critically examine the public policy implications of recent interconnection disputes, examining the ways in which interconnection raises issues similar to, and distinct from, those animating the net neutrality debate. It will also put IP interconnection disputes into historical context, asking what we can learn from the Commission's previous response to similar issues in the telephone and cable space. Tentatively, the paper concludes that the FCC should conduct limited review of interconnection disputes using antitrust principles, consistent with its displacement of the FTC as one of the two competition authorities overseeing the broadband industry—an approach that seems consistent with the direction suggested in the Open Internet Order.

What follows is an earlier draft of a paper I wrote on this topic last year, when the FCC was first contemplating regulation in this space. Most of the argument below is aimed at countering the suggestion by some that the FCC should mandate public disclosure of interconnection agreements. I anticipate this paper will build upon the work below, although the nondisclosure point should be a minor issue in the final article.

Regulating Interconnection

Daniel A. Lyons[*]

Introduction

---

[*] Associate Professor of Law, Boston College Law School. This paper was supported in part by Broadband for America and by the Boston College Law School Fund. Thanks to Fernando LaGuarda, Crystal Lyons, David Olson, and David Young for their helpful comments and suggestions.

Several commenters have requested that the Federal Communications Commission expand the Open Internet inquiry to encompass interconnection transactions at the Internet's "core." These calls followed announcements that Netflix would pay to connect its servers directly to the Comcast and Verizon networks, rather than relying on third-party networks to deliver its content to consumers. Spurred by concerns about content providers paying broadband providers to carry their traffic, these commenters have sought greater transparency in the interconnection market. Some have called upon the Commission to require that all such interconnection agreements be filed with the agency and their terms made public. But left open is the question of whether this greater transparency would benefit consumers.

The Commission has quite properly explained that the interconnection market lies outside the scope of the current docket. The Open Internet NPRM has focused upon the extent to which broadband providers can prioritize different types of Internet traffic over the so-called "last-mile" networks that carry information from the Internet to consumers. Whatever the merits of the proposed rules, they are distinct from the question of how various companies connect to each other in purely business-to-business transactions to form the network of networks known as the Internet. Chairman Tom Wheeler has correctly explained that interconnection is "not a net neutrality issue" and as a result, says a Commission spokesman, "[p]eering and interconnection are not under consideration in the Open Internet proceeding."

More fundamentally, interconnection is a competitive and innovative market that displays few if any signs of market failure justifying regulatory intervention. Network providers have negotiated connection agreements to exchange traffic since the advent of the Internet—sometimes for free and sometimes on a paid basis. Content providers can choose from a plethora of Internet service providers to carry their traffic from their servers across the Internet to consumers. Internet transit prices have fallen precipitously each year since broadband's inception, and many networks operate on razor-thin margins. Netflix's interconnection agreements are simply the latest in a series of examples of the market constantly evolving in response to changes in technology and demand. With a healthy, competitive ecosystem, regulatory interference risks ossifying the process and halting network providers' ability to adapt to new content and applications.

The proposed requirement that networks publicly disclose the terms of their interconnection agreements could harm that competition. It is a basic tenet of economic and industrial organization literature that sharing competitively sensitive information among rivals can facilitate tacit collusion. The Supreme Court, antitrust authorities, and even this Commission have stressed that disclosure of price and cost information can be particularly harmful to competition, especially in markets that display significant barriers to entry. Even absent collusion, price sharing can negatively impact firms' willingness to discount and therefore can lead to supracompetitive prices. Because of this potential effect on competition, the Commission should reject calls to mandate public disclosure of the terms of private business-to-business interconnection agreements.

I. Interconnection Markets are Robust and Highly Competitive

A. Overview of the Interconnection Market

Traditionally, the Commission has focused primarily upon the residential broadband market, and to a lesser extent, the market for commercial end-user broadband access. For example, the 2010 Open Internet rules applied only to "broadband Internet access service," which the Commission defined as "a mass-market retail service…that provides the capability to transmit data to and receive data from all or substantially all Internet endpoints" that is "marketed and sold on a standardized basis to residential customers, small businesses, and other end-user customers such as schools and libraries."[1] Similarly, the 2010 National Broadband Plan focused upon six long-term goals largely aimed at connecting homes and communities to broadband networks, consistent with Congress's directive that the agency "ensure that every American has 'access to broadband capability.'"[2] For years, end-user broadband service has been sold primarily as subscription-based model within which the consumer purchases a publicly-advertised monthly plan for Internet access, which buys either unlimited monthly service or unlimited service up to a monthly limit, with a per-unit overage charge for exceeding the customer's allotted consumption.

But upstream into the Internet ecosystem, the interconnection market is much more complex and dynamic. Commentators often describe the Internet, accurately, as a "network of networks." Interconnection agreements stitch this network together. Professor Christopher Yoo describes the interconnection market as a "collection of 35 thousand autonomous systems bargaining with one another through arms-length transactions" to shuttle traffic among the Internet's end-points.[3] As one might expect, these agreements inevitably contain wide variations in the terms under which parties interconnect and exchange traffic with one another.[4] Interconnection agreements can run hundreds of pages, governing a wide range of conditions, and are typically covered by non-disclosure agreements that reflect the competitively sensitive nature of those terms.

1. Transit Service

---

[1] *Preserving the Open Internet: Broadband Industry Practices,* 25 FCC Rcd. 17905, 17932 (2010).

[2] CONNECTING AMERICA: THE NATIONAL BROADBAND PLAN at xi, xiv-xv (2010).

[3] Christopher S. Yoo, THE DYNAMIC INTERNET: HOW TECHNOLOGY, USERS, AND BUSINESSES ARE TRANSFORMING THE NETWORK at 55 (2012).

[4] *Id.*

Much of the concern about Netflix's interconnection agreements stem from the misconception that Internet content providers typically pay nothing to deliver their traffic to the Internet. In fact, these providers often purchase connectivity from one or more Internet transit providers. In simplified form, Internet transit service is a business relationship whereby a network sells Internet access.[5] A content provider signs an agreement with a transit provider, which agrees to deliver the client's content to all Internet destinations. The transit provider then enters into interconnection agreements with other networks upstream to provide those pathways to any Internet destination.

Internet transit is typically sold on a metered basis, using the 95th percentile measurement method.[6] Through this method, the transit provider measures the amount of traffic to or from the customer every five minutes for a month. At the end of the month, each of these samples is converted to a megabit-per-second figure and the samples are rank-ordered from largest to smallest. The 95th percentile figure is used to represent the customer's monthly volume, and is multiplied by the transit agreement's per-Mbps unit price to calculate the customer's monthly bill.[7] Many transit providers provide a unit-price discount if the customer agrees to a minimum guaranteed amount of monthly traffic, known as a "commit."[8]

2. Peering

As an alternative to purchasing transit service, two networks may agree to enter into a peering agreement. Peering is a business relationship in which two companies agree reciprocally to provide access to each other's customers.[9] Peering is often done on a settlement-free basis, meaning that the two partners agree to exchange traffic without billing one another based on the

---

[5] William B. Norton, THE INTERNET PEERING PLAYBOOK: CONNECTING TO THE CORE OF THE INTERNET at 28 (2013).
[6] *See id.* at 30-32.
[7] *Id.*
[8] *Id.* at 32.
[9] *Id.*

traffic flow. Many settlement-free peering agreements are between networks of comparable size, where the flow of traffic in each direction is roughly equal and therefore the transaction costs of metering would be greater than the net monthly payout. But some peering agreements are on a paid basis, whereby one party agrees to pay the other as a condition of peering.

Peering and transit are related but distinct products. Peering provides a company access only to the peering partner's end-user customers. It does not guarantee that the traffic exchanged will be forwarded on to Internet points that are not within the peering partner's network. As Norton explains, "Internet Transit is a service that provides access to the global Internet, while Internet Peering simply provides a more direct path for a subset of the traffic."[10] Peering may be advantageous because it is cheaper than the equivalent service, or because direct access to the peering partner's customers reduces the number of "hops" between end-points and therefore improves the quality of the transmission. But because transit exists as an alternative to peering, peering prices are disciplined by transit markets as the next-best alternative.

3. Other Innovations: Content Delivery Networks and Server Farms

Content providers may also choose to rely upon a content delivery network rather than transit or peering to deliver their traffic to the Internet. Like transit providers, CDNs sell content providers access to the Internet. But rather than arrange transport across the Internet from the content provider's servers to consumers, CDNs maintain a distributed network of servers around the country, and enter into transit or (typically paid) peering agreements with end-user broadband networks. The CDN maintains a copy of the client's content on each server, and when requested, will deliver the content from one of its servers directly to consumers. Because the content is stored closer to its destination and traverses fewer interconnections, CDNs can be a high-quality, low-cost alternative to transit.

---

[10] *Id.* at 66.

Some content providers have also begun building server farms to cache and distribute their content locally. These server farms act like company-owned content delivery networks, interconnecting directly with broadband providers and bypassing the public Internet completely. Like CDNs, server farms shorten the pathway from server to consumer and thus reduce the possibility that congestion will reduce the quality of the transmission, and also may provide cost savings compared to traditional transit. Professor Yoo notes that Google, Yahoo!, and Microsoft have used server farms to bypass the public Internet for roughly one-third of their total traffic.[11]

B. Interconnection Markets are Competitive

As noted above, content providers have a wide range of options to choose from when deciding how to deliver their content to consumers. Some transit providers have a nearly global network footprint, while others operate more regionally and rely more heavily on interconnection agreements to route traffic to end-users.[12] Some providers offer only transit, while others provide a variety of complementary services as well.[13] And as noted above, some content providers may find peering or CDN delivery to be a competitive alternative to traditional transit service.

It is also worth noting that content providers and transit providers need not rely upon a single interconnection agreement to process their traffic. Rather, many content and transit providers will maintain multiple pathways through which traffic can reach end users, a practice known as "multi-homing." By securing multiple pathways to an end-user, multi-homing helps a provider offer greater reliability and reduces the market power that a single network may otherwise wield over the flow of Internet traffic.[14]

---

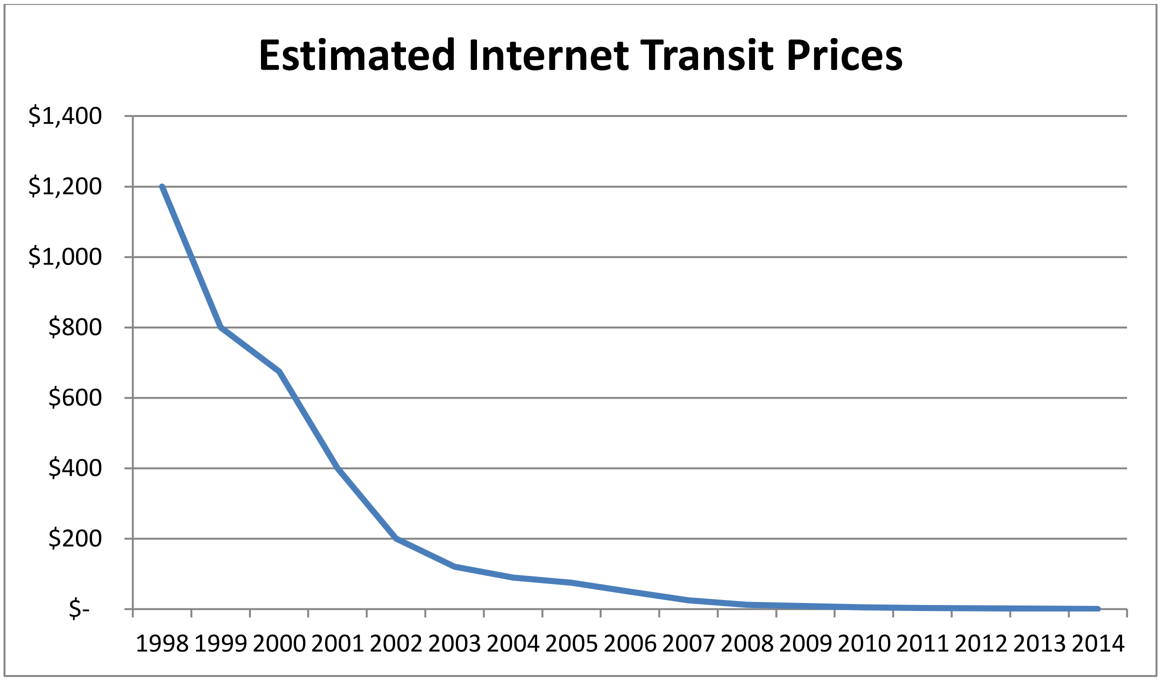
[11] Yoo at 68.

[12] See Dan Rayburn, *How Transit Works, What it Costs & Why It's So Important*, Feb. 24, 2014, available at http://blog.streamingmedia.com/2014/02/transit-works-costs-important.html.

[13] *Id.*

[14] See Yoo at 62-64.

Price trends demonstrate the competitive nature of interconnection markets. Though pricing schedules are often protected by nondisclosure agreements, there is a general consensus that competition has driven down Internet transit prices precipitously and continuously each year since the Internet's inception. Interconnection consultant William Norton calculates, based on informal surveys, that the average per-Mbps price for non-commit transit service has fallen from $1200 in 1998 to $12 in 2008 and $0.94 in 2014—an average rate of over 30 percent each year.[15] While Norton notes that the individual data points are only "rough indications" of price, the "trend is unmistakable, and no one would disagree."[16] Similarly, research firm TeleGeography estimates that American transit prices have fallen 26% annually from 2007 to 2012, and the rate of decline is increasing.[17]



Source: William B. Norton, INTERNET PEERING PLAYBOOK 2013

---

[15] Norton at 34.
[16] *Id.*
[17] See TeleGeography Press Release, *IP Transit Prices Steepen*, Aug. 2, 2012, available at http://www.telegeography.com/products/commsupdate/articles/2012/08/02/ip-transit-price-declines-steepen/.

Like the transit market, CDN prices are not generally made public, but studies suggest that CDN prices are falling at rates roughly comparable to transit prices. Streaming Media Analyst Dan Rayburn estimates that CDN pricing is down 20-25 percent from 2012 to 2013, and he expects even greater declines in 2014 and 2015.[18] Given that transit and CDN services are quasi-substitutes, one should not be surprised to see similar pricing trends in both markets.

C. Interconnection Markets are Dynamic and Evolving

Moreover, the array of interconnection services has evolved over time in response to the growth in the volume and diversity of Internet users, content, and applications. As Professor Yoo has explained at length, when the Internet backbone was privatized in the 1990s, the Internet reflected a hierarchical structure similar to the traditional telephone network: last-mile networks serving end-users contracted with regional ISPs, each of which in turn contracted with a private backbone provider to carry traffic to the Internet. These backbone providers interconnected at one of four public peering points to exchange traffic with one another and route it back down. Traffic thus flowed from the sender's last-mile provider up to a public peering point, over to another backbone provider, then back down again to the last-mile provider serving the recipient.[19]

But as these standard pathways became congested, networks entered into alternative agreements such as private peering and secondary peering agreements to exchange traffic without having to traverse the path through a public peering point. Today, the old up-over-down hierarchy has been replaced by a lattice of interconnecting networks that provide multiple potential pathways for traffic to get from one point to another. The existence of multiple pathways for any given Internet transmission helps alleviate congestion and makes the

---

[18] See Dan Rayburn, The State of the CDN Market, May 2014, available at http://www.streamingmedia.com/dansblog/2014CDNSummit-Rayburn.pdf.
[19] Yoo at 58-59.

marketplace more competitive, as "the presence of alternative paths to connect to the Internet naturally limits every market participant's ability to raise price."[20]

Notably, these innovative new interconnection arrangements coincide with disruptive changes in interconnection markets or in adjacent content and application markets. Norton explains that private peering among cable providers became necessary following the bankruptcy of the cable industry's primary Internet Service Provider, @Home, in 2001, coupled with the rise of peer-to-peer networking traffic that made settlement-free peering more cost-efficient than transit.[21] Similarly, it is no coincidence that some providers of bandwidth-intensive video content such as Netflix are significant customers of CDN providers, and that others such as Apple and Google are leaders in the server farm market. This content is unusually susceptible to Internet congestion, which drove demand for alternatives to traditional public Internet transit services.

D. Placing the Netflix Interconnection Agreements into Context

Netflix's direct interconnection agreements reflect this broader evolutionary trend and show how the competitive marketplace can adapt to meet content providers' needs. It is a mistake to suggest, as some commenters have, that Netflix's interconnection deals with Comcast and Verizon reflect a new cost to the company. It would be more accurate to say that the company has shifted some transit costs from one network to another. Rather than paying third-party transit providers and content-delivery networks to carry traffic to Comcast or Verizon networks, Netflix appears to have entered a paid peering relationship to interconnect directly to those networks. In essence, Netflix has cut out the middleman.

Dan Rayburn suggests that the move was a rational response to changes in the Internet ecosystem. For over a year, transit provider Cogent Communications has been locked in a public

---

[20] *Id.* at 63 (citing sources).

[21] William Norton, Evolution of the U.S. Peering Ecosystem, available at http://drpeering.net/white-papers/Ecosystems/Evolution-of-the-U.S.-Peering-Ecosystem.html.

peering dispute with Verizon. Cogent's increasing volume of Verizon-bound traffic exceeds the capacity of the existing connections between the Cogent and Verizon networks.[22] Cogent and Verizon reached an impasse over the cost of upgrading these connections and the transit fees, if any, that Cogent should pay for the increased volume. In the meantime, Cogent's customers, including Netflix, began suffering delays due to the dispute. Rayburn suggests that a similar dispute is likely responsible for the decline in the quality of Netflix quality to some Comcast customers.[23] Rather than suffer collateral damage from this interconnection dispute, Netflix made the reasonable decision to interconnect directly with Comcast and Netflix and reduce the number of hops from its content to its customers.

Nor are these the first direct interconnection agreements Netflix has signed, although they likely represent the first that were not on a settlement-free basis. But the company is hardly the first content provider to enter into a paid peering arrangement. As noted above, several prominent Internet content providers have built server farms to bypass parts of the public Internet, relying at least in part on direct interconnection with broadband providers to do so. In fact, Comcast has an entire business unit dedicated to selling interconnection services.[24] Amazon, Google, and Facebook are among the content providers who have "eliminated the middleman" and signed direct interconnection agreements with Comcast.

Nor should policymakers be unduly concerned that interconnection agreements allow broadband providers to exploit their unique positions in the Internet's architecture in ways that harm consumers. Telecommunications law has long been concerned with the control that last-

---

[22] See David Young, Unbalanced Peering, and the Real Story Behind the Verizon/Cogent Dispute, June 19, 2013, available at http://publicpolicy.verizon.com/blog/entry/unbalanced-peering-and-the-real-story-behind-the-verizon-cogent-dispute.

[23] See Dan Rayburn, *Here's How the Comcast & Netflix Deal is Structured, With Data & Numbers*, Feb. 27, 2014, available at http://blog.streamingmedia.com/2014/02/heres-comcast-netflix-deal-structured-numbers.html.

[24] See Lance Ulanoff, *The Comcast-Netflix Deal: Fact vs. Fiction*, Feb. 26, 2014, available at http://mashable.com/2014/02/26/comcast-netflix-net-neutrality/.

mile network providers maintain over so-called “bottleneck” facilities to consumers, because ownership of access networks may give rise to concerns about market power in access markets. But as an initial matter, it is worth noting that in this instance, both parties to an interconnection agreement occupy a strategically important position: while Comcast owns the final path to the end-user, Netflix owns the initial path from the company’s servers. With 35 million subscribers who are responsible for up to one-third of all North American Internet traffic during peak hours, Netflix’s bargaining position is far from ephemeral. When negotiating the terms of their interconnection agreement, both companies have leverage, because a holdout on either side could preclude a deal.

More generally, agreements that eliminate a middleman are typically efficient and welfare-enhancing. In this case, Comcast maintains interconnection agreements with a wide range of transit providers and CDNs, many of which Netflix could choose instead of direct interconnection to deliver traffic to Comcast consumers.[25] The fact that Netflix chose direct interconnection suggests that it found this to be the best alternative, either because the price was better or the quality of the service is worth any price premium.[26] Unsurprisingly, since the agreement was signed Comcast customers have seen improved performance from Netflix, because the data travels a more direct route to the customer’s house and is no longer dependent upon potentially overloaded interconnection bottlenecks. And other content providers who rely upon Cogent and other content delivery networks will likely see an improvement as well, as Netflix traffic no longer places such a significant burden on transit network links.

---

[25] Many of these transit and CDN providers deliver a broad mix of traffic to broadband networks, which makes it difficult for a broadband provider to use interconnection agreements as a proxy to price discriminate against certain content, such as video.

[26] Notably, many interconnection agreements contain a Service Level Agreement whereby the network guarantees a minimum level of quality, upon threat of financial sanction for noncompliance. Streaming media providers like Netflix, whose services are latency-dependent, may find SLAs valuable and may receive better SLA terms from broadband providers than intermediate transit providers.

Far from signaling a problem, Netflix's direct interconnection agreements demonstrate the robustness of the interconnection market. When the company's quality suffered as the result of an interconnection dispute, it selected a more efficient agreement from among myriad options to mitigate its exposure. Like multi-homing, secondary peering, and content delivery networks, this direct interconnection agreement shows how a fluid interconnection market can adapt in response to changes in external stimuli. As Professor Yoo explains, these innovations stem from network providers' experiments with new ways to reduce costs and improve the quality of transmission of Internet content.[27] Absent evidence of a market failure, the Commission should be reluctant to assume significant oversight of such a robust, competitive marketplace, because additional regulation may ossify the current environment and disrupt this virtuous cycle of innovation.

II. Disclosure of Interconnection Agreements May Harm Competition

The Commission should be particularly reluctant to adopt rules that require public filing and disclosure of private interconnection agreements between networks. The proposal appears to improve transparency in a market that the Commission has not yet investigated at length. But it may have the unintended consequence of harming competition among peering and transit providers by reducing barriers to collusion by larger networks. Even absent collusion, disclosure is likely to adversely affect prices by reducing companies' incentives to discount. Numerous well-documented examples illustrate the negative effect that disclosure regimes can have on prices and competition. If regulators have reason to suspect that a specific interconnection agreement is likely to harm consumers or competition, antitrust law provides a sufficient remedy to investigate potential issues without risking competitive harm from industry-wide disclosure.

A. Disclosure of Private Interconnection Agreements Can Facilitate Tacit Collusion

---

[27] Yoo at 69.

1. Antitrust Law Recognizes the Risks of Price Transparency

At first glance, the benefits of a price disclosure regime can seem enticing. The model of perfect competition assumes that buyers have perfect information as to firm prices and predicts that markets will move toward uniform, competitive prices for comparable goods.[28] Increased access to firm pricing can reduce search costs for consumers hunting for the best deal.[29] It also may reduce a seller's ability to price discriminate,[30] although one might note that price discrimination itself has ambiguous effects on competition.[31]

But increased price transparency can also have anticompetitive effects by facilitating the negotiation and enforcement of supracompetitive prices.[32] It is a "basic tenet in the economics and industrial organization literature" that "sharing information about cost, transaction prices, and other competitively sensitive information among rivals makes tacit collusion more likely."[33] For almost one hundred years, the United States Supreme Court has consistently recognized that "the exchange of price information among competitors carries with it the added potential for the development of concerted price-fixing arrangements which lie at the core of the Sherman Act's prohibitions."[34] "Regardless of its putative purpose," said the Court, "the most likely consequence of any such agreement to exchange price information would be the stabilization of industry prices."[35]

---

[28] See, e.g., Maurice E. Stucke, *Evaluating the Risks of Increased Price Transparency*, 19-SPG ANTITRUST 81, 81 (2005).

[29] *Id.*

[30] *id.*

[31] see Herbert Hovenkamp, ANTITRUST LAW ¶ 2340c, at 13.

[32] Stucke at 81.

[33] Joanna Shepherd, *Is More Information Always Better? Mandatory Disclosure Regulations in the Prescription Drug Market*, 99 CORNELL. L. REV. ONLINE 1 (2013). See, e.g., George J. Stigler, *A Theory of Oligopoly*, 72 J. POL. ECON. 44 (1964).

[34] *United States v. United States Gypsum Co.*, 438 U.S. 422, 457 (1978). See, e.g., *Am. Column & Lumber Co. v. United States*, 257 U.S. 377 (1921); *United States v. Am. Linseed Oil Co.*, 262 U.S. 371 (1923); *Maple Flooring Mfg. Assn. v. United States*, 268 U.S. 563 (1925); *Cement Mfrs. Protective Assn. v. United States*, 268 U.S. 588 (1925).

[35] *Gypsum Co.*, 257 U.S. at 457.

Federal antitrust authorities have also long warned about the potential anticompetitive risks of transparency among competitors. "A market typically is more vulnerable to coordinated conduct if each competitively important firm's significant competitive initiatives can be promptly and confidently observed by that firm's rivals. This is more likely to be the case if the terms offered to customers are relatively transparent."[36] While the sharing of information among competitors can be procompetitive, "in some cases, the sharing of information related to a market in which the collaboration operates or in which the participants are actual or potential competitors may increase the likelihood of collusion on matters such as price."[37]

The FTC/DOJ Antitrust Guidelines for Collaborations Among Competitors offers three red flags to help identify when information disclosure may facilitate collusion.

- **Information about price**: "Other things being equal, the sharing of information relating to price, output, costs, or strategic planning is more likely to raise competitive concern than the sharing of information relating to less competitively sensitive variables."[38]
- **Current information**: "Similarly, other things being equal, the sharing of information on current operating and future business plans is more likely to raise concerns than the sharing of historical information."[39]
- **Individual Company Data:** "Finally, other things being equal, the sharing of individual company data is more likely to raise concern than the

---

[36] Fed. Trade Comm'n & U.S. Dep't of Justice, HORIZONTAL MERGER GUIDELINES § 7.2 (2010).

[37] Fed. Trade Comm'n & U.S. Dep't of Justice, ANTITRUST GUIDELINES FOR COLLABORATIONS AMONG COMPETITORS § 3.31(b) (2010).

[38] *Id.*

[39] *Id.*

> sharing of aggregated data that does not permit recipients to identify individual firm data."[40]

Of course, the proposal to mandate disclosure of individual interconnection agreements raises all three red flags: it would reveal real-time information about prices and costs of transit on a company-by-company basis.

2. The Mechanics of Tacit Collusion

Price transparency helps overcome the two primary barriers to collusion. First, the open communication of prices reduces the uncertainty of negotiating a supracompetitive price.[41] Because overt communication about price collusion is prohibited by the Sherman Act, firms seeking to collude must overcome the difficulty of communicating indirectly to establish their target price. But as the *Container Corporation* Court explained, sharing current price data can solve this problem by signaling a target toward which others can move. In that case, suppliers of corrugated containers shared current price information upon request about the most recent price charged for a good.[42] The Court explained that "[t]he exchange of price information seemed to have the effect of keeping prices within a fairly narrow ambit" because "[k]nowledge of a competitor's price generally meant matching that price."[43] The result was a movement toward a stable, uniform price in violation of the Sherman Act.[44]

Once firms have established a collusive price, transparency also helps enforce the collusive agreement.[45] Here, as the Supreme Court has said, "[u]ncertainty is the oligopoly's greatest enemy,"[46] because of the difficulty of identifying and punishing cheaters. But price

---

[40] *Id.*

[41] See Stucke at 81.

[42] *United States v. Container Corp.*, 393 U.S. 333 (1969).

[43] *Id.* at 336-37.

[44] *Id.* at 334.

[45] Stucke at 81.

[46] *Brook Group v. Brown & Williamson Tobacco*, 509 US. 209, 238 (1993),

transparency eliminates that uncertainty and therefore facilitates enforcement: "If…every transaction is publicized immediately, all members of the industry will know when one has made a price cut, and each can retaliate on the next transaction. Knowledge that retaliation will be swift serves as a powerful deterrent to price cutting and therefore facilitates the maintenance of tacitly collusive prices."[47] Because market players know that any attempt at cheating will bring a swift response, they are less inclined to defect from the collusive price in the first place.

The general risk of tacit collusion is magnified by several structural factors inherent in the interconnection market. The first is concentration of competitively important players. Collusion is easier when fewer firms need to cooperate.[48] Though there are roughly 35,000 networks in the interconnection market, disclosure proponents argue that only a handful of them need to cooperate to control interconnection rates to end-user broadband networks. If they are correct, transparency would ease efforts by those broadband network providers to collude on a market price for interconnecting to last-mile networks. Second, there are significant barriers to entry.[49] Building and operating a broadband network requires significant upfront capital, which helps insulate the collusive scheme from the threat of competitive entry. Third, providers enter into regular and frequent interconnection agreements, most of which govern only a small portion of total traffic carried over a network. This makes cheating less likely because there is little benefit from departing from the collusive price in a single transaction, and competitors can move quickly to punish any defector.[50] Finally, players in the interconnection market are customers as

---

[47] F.M. Scherer & David Ross, INDUSTRIAL MARKET STRUCTURE AND ECONOMIC PERFORMANCE 348 (1990).
[48] See, e.g., John M. Kuhlman, *Nature and Significance of Price Fixing Rings*, 2 ANTITRUST L. & ECON. REV. 69, 71 (1968).
[49] *Id.* at 72.
[50] See James E. Hartley, *Market Definition in a Monopoly Case*, C695 ALI-ABA 99 (1991).

well as competitors; these multimarket contacts provide multiple pressure points with which to punish a cheater, which makes cheating less likely.[51]

Though interconnection markets differ somewhat from a typical wholesale transaction, the principles play out similarly. A broadband provider could attempt to communicate a collusive industry-wide interconnection rate to its rivals by insisting on that rate in a negotiation with a single content provider, and publicly disclosing the resulting agreement. Its likelihood of success would increase if the broadband provider is sufficiently large to command leverage in the negotiation, particularly if the negotiation is with a relatively small content provider. Once the interconnection price is publicized, rival broadband providers can use that price as an anchor or target to guide their own negotiations with content providers. As other interconnection agreements are made public, the would-be colluders could monitor the success of their efforts in real-time and adjust their proposed price targets if necessary through successive interconnection agreements in an attempt to achieve their shared objective. The multitude of potential content providers with which to partner allows the process to be highly iterative and eases the ability to use individual transactions to determine initial success and monitor ongoing compliance. Enforcement of the collusive price can be done in myriad ways, including by changing the terms of peering agreements between broadband networks to ensure cooperation with the collusive market for interconnection with upstream content.

Moreover, the complexity of interconnection agreements can mask tacit collusion. Interconnection agreements can run hundreds of pages and contain thousands of terms. It is unclear to what extent disclosure will increase transparency of the market to members of the public who lack detailed familiarity with such deals. But this complexity can mask attempts by

---

[51] See, e.g., Amalia R. Miller, *Did the Airline Tariff Publishing Case Reduce Collusion?* 53 J.L. & ECON. 569 (2010).

firms to communicate with one another in violation of the Sherman Act, though fine print and price quotes that lay deep within the text of these lengthy, complex agreements. Similar allegations lay at the core of the *Airline Tariff Publishing* case. The Federal Trade Commission alleged that certain airlines used an automated fare reporting system to coordinate fare increases by communicating via footnotes, start dates, and end dates that were publicly disclosed but were of little relevance to consumers or travel agents.[52]

B. Even Absent Collusion, Disclosure Can Negatively Impact Prices

Disclosure can also lead to rising prices without collusive action that would violate antitrust law. Even absent tacit collusion, transparency can have an anticompetitive effect based simply on the unilateral rational actions of market players. As the Court noted in *Brooke Group Ltd. V. Brown & Williamson Tobacco Corp.*, firms may set prices at "a profit-maximizing, supracompetitive level by recognizing their shared economic interests and their interdependence with respect to price and output decisions."[53] Particularly in concentrated markets, it is unsurprising to find that firms may set their prices based partly on strategic considerations about their competitors' behavior.[54] Absent some agreement among competitors, supracompetitive pricing that emerges from the unilateral actions of multiple market players does not violate the Sherman Act—though it has an adverse effect on customers and competition.

Price transparency undermines the likelihood that a particular firm will discount to gain a competitive advantage. As the Federal Trade Commission has explained, coordinated information sharing "can blunt a firm's incentive to offer customers better deals by undercutting the extent to which such a move would win business away from rivals."[55] Market participants

---

[52] *Airline Tariff Publishing Co.*, 1993-2 Trade Cas. (CCH) ¶ 70,409 at 71,167 (D.D.C.).
[53] 509 U.S. at 227.
[54] Stucke at 81.
[55] Horizontal Merger Guidelines § 7.2.

typically offer discounts in an attempt to gain market share away from rivals. But a company is less likely to offer such a discount if competitors can quickly learn the details of the agreement and move to match.[56] Because it would be unlikely that discounting would gain share, firms would be less likely to do so.

Transparency also decreases the incentives for companies to price goods aggressively. When a firm lacks knowledge about its competitor's prices, it has incentives to offer low prices in an attempt to beat the "unknown" deal.[57] But when rival pricing is no longer unknown, "the incentive to outbid unknown price terms disappears."[58]

In broadband markets, the planes of competition among broadband providers includes both interconnection price and quality of service. One could argue that because consumption of Internet content is somewhat non-rivalrous, broadband providers lack some of the incentives to price aggressively in response to their rivals like typical wholesalers do. If AT&T offers a low interconnection rate to Netflix, for example, it is unlikely that Netflix will shift some of its volume away from Verizon as a result. But transparency could affect Netflix's likelihood of securing non-price features that affect the quality of the product as delivered to end-user consumers, such as the capacity and location of interconnection ports. AT&T could bid aggressively by taking technical measures to assure that Netflix traffic is delivered with fewer interruptions over its network, which allows it to tout superior network quality to both content providers and end-user consumers. But those incentives would be retarded if public disclosure allowed rivals to move quickly to counter, because AT&T would secure no demonstrable long-term advantage as a result of these efforts.

---

[56] See Letter from Federal Trade Comm'n to Mark Formby, Representative, Miss. House of Representatives 9 (Mar. 22, 2011), available at http://www.ftc.gov/os/2011/03/110322mississippipbm.pdf (discussing anticompetitive effects of proposed state law mandating disclosure of price terms in pharmaceutical industry).

[57] Shepherd at 19.

[58] *Id.* (citing sources).

Notably, disclosure of interconnection agreements may also have anticompetitive effects on adjacent markets for content and applications. First, disclosure may make it easier for networks to price discriminate against particular content, because they could more easily identify the transit networks that targeted content providers use to deliver their traffic to the Internet, and can press for higher transit fees from those networks. Second, the disclosure of interconnection agreements will allow content and application providers access to competitively sensitive data about their rivals' transit costs, which can raise risks of tacit collusion in content and application markets. Third, content and application providers who are concerned about protecting this information may contract with networks that are not subject to disclosure rules, such as CDNs, or they may attempt to self-provision transit service to avoid disclosing cost information to competitors, even in situations where it would otherwise be uneconomical to do so.

C. Case Studies

Several empirical studies have established that mandatory disclosure of competitively sensitive information can be associated with higher prices.

1. Railroad Grain Contracting

The Staggers Rail Act of 1980 deregulated much of the railroad industry and allowed railroads to enter into privately-negotiated contracts with shippers and receivers.[59] Concerned that the railroads were price discriminating against small shippers, in 1986 Congress mandated the railroads disclose publicly the "essential terms" of any agricultural contracts.[60] These terms included price, the identity of the customer, the origin and destination of the shipment, the length of the contract, volume requirements, prior contracts between the parties, and effective date.[61]

---

[59] Staggers Rail Act of 1980, Pub. L. 96-448, 94 Stat. 1895 (1980).

[60] See Pub. L. 99-509, § 4051 (1986).

[61] *Id.*

An empirical study showed that this disclosure obligation had a significant and adverse effect on the price for railroad shipping.[62] Prior to the 1986 disclosure obligation, rates for railroad transportation of wheat in the Plains states was declining, a finding consistent with other studies testing the effect of deregulation on railroad rates.[63] But this trend reversed sharply after the disclosure obligations took effect in January 1987. After controlling for exogenous forces, the study found that rates rose between 10 and 13.7 percent.[64] The authors conclude that "contract disclosure and the increased reliance on posted tariffs facilitated rate coordination by the oligopolistic railroad industry, thereby leading to an increase in rail rates."[65] They note that this finding is consistent with earlier findings about rate disclosure in the inland barge industry.[66]

2. Ready-Mixed Concrete

In 1993, the Competition Council, Denmark's antitrust authority, gathered and published statistics on the transaction prices of individual firms for two grades of ready-mixed concrete in three regions within Denmark.[67] The Council took this action under authority granted to it by the Competition Act of 1990, which instructed the Council to combat suspected oligopoly collusion through measures designed to increase market transparency.[68] From October 1993 until December 1996, the Council sampled actual invoice prices from 18 production sites in three regions, and published this firm-specific price data quarterly in the hope of improving information for buyers (primarily building contractors).[69]

---

[62] See Stephen W. Fuller et al, *Effect of Contract Disclosure on Price: Railroad Grain Contracting in the Plains*, 15 WEST. J. OF AGRICULTURAL ECON. 265 (1990).

[63] *Id.* at 270-71.

[64] *Id.* at 271.

[65] *Id.*

[66] *Id*. (citing J.T. Hong and C.R. Plott, *Rate Fixing Policies for Inland Water Transportation: An Experimental Approach*, 13 BELL J. ECON. 1 (1982)).

[67] See Svend Albaek et al, *Government-Assisted Oligopoly Coordination? A Concrete Case*, 45 J. INDUS. ECON. 429 (1997).

[68] *Id.* at 429-30.

[69] *Id.* at 432.

As in the railroad example, the unintended consequence of this disclosure regime was to facilitate collusion and raise prices. According to one study, average prices of reported grades rose between 15 and 20 percent in the first year following publication.[70] Prices also converged significantly across firms serving the same market.[71] The authors considered, but rejected, several alternative explanations for this dramatic increase, including business upturn, capacity constraints, and input prices.[72] Ultimately, the authors conclude, "the evidence presented in this paper indicates that the Danish Competition Council, by providing reliable price reporting services, has unwittingly assisted firms in reducing the intensity of competition and thereby allowed them to increase prices."[73]

3. Telecommunications

The Federal Communications Commission has also previously considered acknowledged that "[o]ne of the basic prerequisites for [] anticompetitive behavior is knowledge of a competitor's prices."[74] Beginning in 1983, the Commission discussed whether tariffing of nondominant telephone companies "impair[ed] competitive pricing, and facilitate[d] collusive conduct."[75] It noted that forbearance "involves less disclosure to competitors of carriers' rates and tariff conditions than streamlined regulation" and consequently would "eliminate[] a potential vehicle for collusive conduct and facilitate[] price discounting."[76]

The Commission expanded on these thoughts in 1985, explaining that "[t]he continuation of tariffs for forborne carriers [] presents an opportunity for collusive pricing by competing carriers. Since carriers can ascertain their competitors' existing rates and keep track of any

---

[70] *Id.* at 429.
[71] *Id.*at 430.
[72] *Id.* at 440.
[73] *Id.* at 441.
[74] *Competition in the Interstate Interexchange Marketplace*, 5 FCC Rcd. 2627, 2644 (1990).
[75] *In re Policies and Rules Concerning Rates for Competitive Carrier Services and Facilities Authorizations Therefor, Fourth Report and Order*, 95 F.C.C. 2d 554, 554 n.1 (1983).
[76] *Id.*at 554 n.3.

changes in those rates by reviewing the filed tariffs, carriers may be encouraged to maintain rates at an artificially high level. Without forborne carrier tariffs on file, carriers may initiate price cutting or generally institute rates at a lower level to meet directly customer demand."[77] Although it ultimately concluded that the evidence was "inconclusive as to the issue of tacit price coordination among AT&T, MCI and Sprint," it concluded that to the extent that such coordination existed, it was "better addressed by removing regulatory requirements that may facilitate such conduct."[78]

D. The Possibility of Tacit Collusion Could Invite Costly Antitrust Scrutiny

Even if the disclosure regime ultimately has no actual anticompetitive effects, the proposed rule would impose substantial compliance obligations on the industry. Network providers must scrutinize interconnection agreements to assure that there are no antitrust concerns with disclosure, and that the agreements do not otherwise contain proprietary or competitively sensitive information. The disclosure obligation may also limit parties' flexibility when negotiating an interconnection agreement, because of concerns that any terms in the final agreement would be made public.[79]

Moreover, assuming the burden of these compliance costs provides no guarantee that the firm will avoid a costly antitrust investigation. As noted above, federal antitrust officials look skeptically at arrangements to share prices, particularly given the structural factors that mark interconnection markets. The routine exchange of such competitively sensitive information is likely to attract regular antitrust oversight and could trigger investigations of firms that are in fact

---

[77] *In re Policy and Rules Concerning Rates for Competitive Common Carrier Services and Facilities Authorizations Therefor*, *Sixth Report and Order*, 99 FCC 2d 1020, 1030 (1985).

[78] *In re Motion of AT&T to be Reclassified as a Non-Dominant Carrier*, 11 FCC Rcd. 3271, 3314-15 (1995). See, e.g., Peter K. Pitsch and Arthur W. Bresnahan, *Common Carrier Regulation of Telecommunications Contracts and the Private Carrier Alternative*, 48 FED. COMM. L.J. 447, 484 (1996) (proposing private contracts as alternative to public tariff, to avoid potential price coordination).

[79] See, e.g., Geoffrey Manne, *The Hydraulic Theory of Disclosure Regulation and Other Costs of Disclosure*, 58 ALA. L. REV. 473, 482-484 (2007).

innocent of wrongdoing. The fact that the Commission has mandated disclosure is not a complete defense if antitrust authorities suspect that parties are misusing the disclosure regime to illegally collude. Even a defendant cleared of any wrongdoing will incur substantial defense costs to clear its name.

Finally, given how unsettled the law is in this area, even a firm with no malicious intent may unwittingly incur liability. Claims that a particular exchange of competitively sensitive information violates antitrust law are decided under the rule of reason, which requires the court to consider a "number of factors" to "divin[e] procompetitive or anticompetitive effects."[80] Canvassing the history of such claims, the Second Circuit noted that "[t]he state of the law on this issue was not always so clear."[81] Numerous commentators have noted that it is not much clearer today.[82] Given the risk that innocuous disclosures may give rise to antitrust liability, it seems unwise policy to invite the proceeding and suffer the attendant compliance costs and judgment risks associated with exchanging competitively sensitive information.

III. Antitrust Law is the Superior Forum to Address Concerns about Anticompetitive Interconnection Practices

Antitrust law provides regulators and the public with sufficient tools to detect and investigate particular interconnection agreements or practices that officials suspect may be anticompetitive. The Federal Trade Commission and the Department of Justice Antitrust Division have overlapping authority to investigate, and if necessary prosecute, anticompetitive interconnection practices, just like they can, and do, in other areas of the economy. And private plaintiffs such as content providers or networks who are harmed by allegedly anticompetitive practices can avail themselves of federal or state competition law to seek civil remedies.

---

[80] *Todd v. Exxon Corp.*, 275 F.3d 191, 199 (2nd Cir. 2001).

[81] *Id.* at 198.

[82] See, e.g., Stucke at 81.

There are multiple advantages to using targeted ex-post judicial proceedings rather than broad ex-ante disclosure rules to combat anticompetitive conduct. Through the discovery process, the relevant plaintiff will get access to the terms of the interconnection agreement and whatever other information s/he needs to investigate the validity of the claim. But the judicial process provides myriad protections to minimize the risk that competitively sensitive information would be made public. For example, discovery may be permitted subject to confidentiality provisions or "attorney's eyes only" orders, backed by the power of a contempt proceeding. More generally, a targeted investigation is less likely than broad disclosure rules to be misused for anticompetitive purposes—and if the case proceeds to court, a neutral magistrate will preside over the case to assure that the proceeding is not abused.

Conclusion

Interconnection is a robust, competitive marketplace that has demonstrated a continuous ability to adapt in response to consumers' growing appetites for Internet-based products and services. Netflix's recent direct interconnection agreements with Comcast and Verizon reflect a broader trend toward alternatives to traditional transit service for delivery of significant volumes of Internet-based traffic, and should not, alone, raise public policy concerns. One may understand the impulse of some commenters to seek greater transparency in interconnection markets by mandating public disclosure of this and other interconnection agreements. But economic literature and almost a century of antitrust jurisprudence warn of the potential unintended consequences of such a rule. Disclosure of competitively sensitive information can create opportunities for tacit price collusion and even unilateral activity that raises prices to supracompetitive levels, as evidenced by prior experiments in the railroad grain and cement markets. Even without any actual anticompetitive effects, disclosure rules entail compliance costs and can lead to significant defense costs and potential liability. Traditional antitrust law provides a superior alternative to address specific instances of suspected anticompetitive conduct in the interconnection market.

Modularity as a Basis for a Dynamic Theory of Internet Regulation

Christopher S. Yoo

ABSTRACT

Recent debates over Internet policy have focused on the network's architecture. For example, the central justification for the Federal Communications Commission's Open Internet proceeding is that the Internet's architecture was critical to its past success and must be preserved. Unfortunately, policymakers and commentators have failed to provide any theoretical basis for determining whether any network configuration is optimal or when changed circumstances might justify a change to the architecture. The result is a regulatory approach that is unnecessarily static and reflexively accepting of the status quo and fails to provide a basis for distinguishing between architectural changes that are part of the network's natural evolution and those that are potentially anticompetitive.

This Article fills this void by providing an analytical framework for assessing network architecture based on modularity theory. It synthesizes modularity theory into five key concepts: near decomposition, interdependencies, abstraction/information hiding, requisite variety, and testing/integration. It then surfaces the tradeoffs inherent in the architectural decision by identifying the benefits and costs associated with modular architectures and analyzing the dynamics of architectural change. It concludes by using the framework to evaluate a series of recent policy applications, including the Telecommunications Act of 1996, network neutrality, the transition from IPv4 to IPv6, calls for opening the Apple iPhone's application programming interfaces (APIs), and the evolution of future Internet architectures.

An architectural approach based on modularity theory yields a number of important insights. It surfaces the inevitability of the tradeoff between generality and efficiency and how generality can obstruct certain types of innovation. It underscores how parallel experimentation can accelerate innovation at the same time that the lack of coordination can retard it. Moreover, like any design hierarchy, modular systems can resist technological change. Finally, it underscores the contingent nature of network architecture and provides heuristics for identifying when architectural change can be beneficial.

Towards an Architectural Theory of Internet Regulation

Christopher S. Yoo[†]

> [T]his Committee must keep in view a fundamental fact about the Internet: . . . the innovation and explosive growth of the Internet is directly linked to its particular architectural design. . . . If this Committee wants to preserve that growth and innovation, it should take steps to protect this fundamental design.
>
> - Statement of Lawrence Lessig Before the Senate Committee on Commerce, Science & Transportation[*]

INTRODUCTION

Recent debates over Internet policy have proceeded from the premise that the Internet's success has stemmed in large part from its architecture and that this architecture must be preserved if that success is to continue. The current debate over network neutrality provides an apt illustration of these dynamics. The Open Internet Order that the D.C. Circuit struck down in January 2014 lauded how the network "architecture enables innovators to create and offer new applications and services without needing approval from any controlling entity, be it a network provider, equipment manufacturer, industry body, or government agency," which in turn permits "new technologies to be developed and distributed by a broad range of sources, not just by the companies that operate the network."[1] Although the details of the revised Open Internet Order currently pending before the Federal Communications Commission has not yet been released, most expect that the new order will follow the same line of reasoning.

[†] John H. Chestnut Professor of Law, Professor of Communication, Professor of Computer & Information Science, and Founding Director of the Center for Technology, Innovation and Competition, University of Pennsylvania. I would like to thank Sandra Braman, Jonathan Cave, and attendees at the First Annual Conference on Law and Computer Science at the University of Pennsylvania, 40th Telecommunications Policy Research Conference, the Intellectual Property Scholars Conference, Tilburg Law and Economics Center, the Warren Center for Network & Data Sciences of the University of Pennsylvania, and the Conference on "Regulating the Evolving Internet Ecosystem," cosponsored by the Federal Communications Commission, American Enterprise Institute, the University of Nebraska College of Law for their comments on earlier drafts. I would also like to thank the Oscar and Miriam Handler Foundation, the New York State Bar Foundation, and University of Pennsylvania Law School's Center for Technology, Innovation and Competition for their financial support.

[*] *Net Neutrality: Hearing Before the Senate Committee on Commerce, Science & Transportation,* 109th Cong 54 (2006) (statement of Prof. Lawrence Lessig), *available at* http://www.gpo.gov/fdsys/pkg/CHRG-109shrg30115/pdf/CHRG-109Bshrg30115.pdf.

[1] Preserving the Open Internet, Report and Order, 25 FCC Rcd. 17905, 17910 ¶ 13 (2010) [hereinafter Open Internet Order]

Commentators generally agree that *modularity*, the partitioning of a system into subsystems that are structurally independent but work together, represents one of the key architectural principles around which the Internet was designed and to which it owes much of its success.[2] Computer scientists almost universally discuss modularity almost exclusively in laudatory terms and rightfully so.[3] The shift to modularity played a critical role in triggering the explosion in computer technology since the mid-1960s.[45] Indeed, as will be subsequently discussed at greater length, it represents the primary mechanism for making complex systems tractable. Furthermore, the current modular architecture has proven incredibly resilient and robust over a period spanning several decades.[6]

That said, recognizing the modularity provides substantial benefits still leaves a host of questions unanswered. For example, is modularity always the best choice? If not, under what circumstances is modularity optimal and suboptimal?[7] The absence of such a theory overlooks the fact that "neither modular designs nor interdependent designs are *inherently* superior."[8] On the contrary, engineering principles recognize that no one architecture does everything well and thus that every architecture necessarily involves tradeoffs. Assessments about the relative merits of particular modular

---

[2] *See, e.g.*, *Net Neutrality: Hearing Before the Senate Committee on Commerce, Science & Transportation*, 109th Cong 10, 12. 13 (2006) (statement of Vinton G. Cerf), *available at* http://www.gpo.gov/fdsys/pkg/CHRG-109shrg30115/pdf/CHRG-109shrg30115.pdf; YOCHAI BENKLER, THE WEALTH OF NETWORKS: HOW SOCIAL PRODUCTION TRANSFORMS MARKETS AND FREEDOM 100–03 (2006); LAWRENCE LESSIG, THE FUTURE OF IDEAS 92 (2001); BARBARA VAN SCHEWICK, INTERNET ARCHITECTURE AND INNOVATION 38–40 (2010); JONATHAN ZITTRAIN, THE FUTURE OF THE INTERNET—AND HOW TO STOP IT 31 (2008); Joseph Farrell & Philip J. Weiser, *Modularity, Vertical Integration, and Open Access Policies: Towards a Convergence of Antitrust and Regulation in the Internet Age*, 17 HARV. J.L. & TECH. 85 (2003).

Other major architectural principles include protocol layering and the end-to-end argument. For discussions of those concepts, see Christopher S. Yoo, *Protocol Layering and Internet Policy*, 161 U. PA. L. REV. 1707 (2013); and Christopher S. Yoo, *Would Mandating Network Neutrality Help or Hurt Broadband Competition?: A Comment on the End-to-End Debate*, 3 J. ON TELECOMM. & HIGH TECH. L. 23 (2004).

[3] *See, e.g.*, David Clark, *Forward to the First Edition*, *in* LARRY L. PETERSON & BRUCE S. DAVIE, COMPUTER NETWORKS: A SYSTEMS APPROACH ix, ix (4th ed. 2007) (observing that "[a]ll good computer scientists worship the god of modularity"); M.A. Padlipski, *A Perspective on the ARPANET Reference Model* 4, 8 (Request for Comments 871, 1982) [hereinafter RFC 871], *available at* http://tools.ietf.org/pdf/rfc871 (calling modularity a "buzzword" and noting that "'[e]veryone knew' modularity was a Good Thing").

[4] 1 CARLISS Y. BALDWIN & KIM B. CLARK, DESIGN RULES: THE POWER OF MODULARITY 221 (2000).

[5] *Id.* at 221.

[6] The protocols that form the essence of the Internet date from 1974. Vinton G. Cerf & Robert E. Kahn, *A Protocol for Packet Network Intercommunication*, 22 IEEE TRANSACTIONS ON COMM. 637 (1974). They became the exclusive core routing protocol on January 1, 1983. Jon Postel, *NCP/TCP Transition Plan* 2 (Network Working Group Request for Comments 801, 1981), *available at* http://tools.ietf.org/pdf/rfc801.

[7] *See* Carliss Y. Baldwin & Kim B. Clark, *Managing in an Age of Modularity*, HARV. BUS. REV., Sept.–Oct. 1997, at 84, 86 ("If modularity brings so many advantages, why aren't all products [and processes] fully modular?").

[8] BALDWIN & CLARK, *supra* note 4, at 258; *accord id.* at 257 (nothing that "there is no economic 'right answer' that works in all cases" when choosing "between an interconnected and a modular approach").

schemes require evaluating the forces resting on each side of this balance in any particular context.

Moreover, even when modularity is desirable, there are many possible ways to modularize a system. Any designer of a modular system must decide how many modules to create, which design elements should be assigned to each module, where the interfaces between modules should be located, and how the interfaces between the modules should be configured.[9] Identification of those determinants allows policymakers to create a dynamic theory of architectural design by asking an equally important question: How does one recognize the conditions under which a modular architecture should change?

Unfortunately, the considerations underlying these decisions remain poorly understood. As one commentator has noted, "the design tradeoffs inherent in abstracting from physical resources are rarely acknowledged in the computing literature."[10] Other commentators similarly observed that "there is little systematic research on how decision makers partition designs into the modules and what are the risks of partitioning incorrectly."[11] As a result, the literature tends to treat modular architectures as immutable artifacts without providing any heuristics to help determine when change might be beneficial.

This Article seeks to fill this void by exploring the theory underlying modularity and applying it to recent and current issues in Internet policy. Part I lays out the basic concepts that serve as the foundation for modularity theory. These include near decomposition, the role of interdependencies, abstractions/interfaces, hidden information, requisite variety, and testing/integration. Together these concepts underscore that the optimal structure of any complex system is determined by technological interdependencies and social considerations.

Part II describes modularity's benefits, including the way it makes complexity more manageable, increases the speed of innovation, facilitates outsourcing, and promotes flexibility. Part III surfaces the tradeoff inherent in modularity by analyzing its costs, including the sunk costs of designing a modular architecture, reduction in efficiency, restrictions on innovation, limitations on alternate institutional forms, lack of coordination, and excessive rigidity.

Part IV examines the dynamics underlying changes in modular networks. This framework makes clear that modular architectures are not natural

---

[9] Sendil K. Ethiraj & Daniel Levinthal, *Modularity and Innovation in Complex Systems*, 50 MGMT. SCI. 159, 162 (2004); *see also* Richard N. Langlois, *Modularity in Technology and Organization*, 49 J. ECON. BEHAVIOR & ORG. 19, 24 (2002). ("The real issue is normally not whether to be modular but *how* to be modular. Which modularization, which structure of encapsulation boundaries, will yield the best system decomposition?").

[10] Jean-François Blanchette, *A Material History of Bits*, 62 J. AM. SOC'Y INFO. SCI. & TECH. 1042, 1047 (2011).

[11] Ethiraj & Levinthal, *supra* note 9, at 160.

constructs. Instead, they depend on factors such as the nature of the interdependencies, technological heterogeneity, speed of technological change, and demand heterogeneity. As such, they should be expected to change over time.

Part V applies the framework to four major issues in telecommunications policy: access to unbundled network elements under the Telecommunications Act of 1996, network neutrality, the shift from Internet Protocol version 4 (IPv4) to version 6 (IPv6), calls for platforms such as the Apple iPhone and social media sites such as Twitter and Facebook to provide open access to their application programming interfaces (APIs), and proposals for future Internet architectures. These examples show how the theory underlying modularity can provide some insights into how best to craft Internet policy.

I. MODULARITY THEORY BASICS

Any assessment of whether a particular modular regime is socially desirable necessarily depends on having some theoretical framework for understanding why modularity exists and what benefits it provides. This Part lays out the basic concepts that provide the foundation for that theory, including near decomposition, the role of interdependencies, abstractions and hidden information in interfaces, requisite variety, and testing and integration.

A. Near Decomposition as a Solution to Complexity

The special challenges associated with developing complex systems are well illustrated by an anecdote related by Frederick Brooks in his celebrated book, *The Mythical Man Month*, which documented IBM's efforts to design System/360, the first fully modular family of computer systems.[12] Consistent with the prevailing conventional wisdom that everyone working on a project should have access to all of the available information about the project, the managers of the System/360 team insisted that every programmer maintain a formal workbook documenting all of the other parts of the system. After only six months, the workbook was more than five feet thick and required one hundred fifty pages of updates every day.[13] This led to Brooks's famous observation that adding more people sometimes slows down projects instead of speeding them up, because the communication cost associated with keeping more people informed can exceed the potential advantages of more man-hours and a greater division of labor.[14]

---

[12] FREDERICK P. BROOKS, THE MYTHICAL MAN-MONTH: ESSAYS ON SOFTWARE ENGINEERING (1975), *discussed in* Langlois, *supra* note 9, at 22.

[13] *Id.* at 76–77.

[14] *Id.* at 18–19.

The classic solution to this problem is implicit in Herbert Simon's landmark analysis of complex systems.[15] Simon suggested that one of the best ways to reduce complexity is to decompose the overall system into a series of subsystems in which interactions within subsystems are relatively frequent and stronger, while interactions between subsystems are relatively rare and weaker.[16] This allows subsystems to operate largely independently of one another in the short run, but to integrate into a larger complex system over the long run.[17] Such a system would only be *nearly* decomposable rather than *completely* decomposable, as the fact that all of the subsystems were part of a larger system meant that some interconnections subsystems would necessarily remain. That said, interactions among subsystems would remain relatively weak, but nonnegligible.[18]

Near decomposition of complex systems yields a number of advantages. As an initial matter, it makes complex systems easier to describe and comprehend.[19] By creating a larger number of intermediate forms that can constitute building blocks for the larger system, near decomposition permits experimentation to occur on a smaller scale instead of with the system as a whole.[20] The existence of stable intermediate forms also allows complex systems to evolve more rapidly.[21]

A related insight emerges from the field of software engineering, which has modeled complex tasks into a series of abstract sequential processes.[22] Subdividing a larger process into a series of smaller subprocesses greatly facilitated software engineers' ability to verify the accuracy of the system by allowing them to test portions of the overall system in isolation.[23] Similarly, design scholars have turned to near decomposition to deal with complexity of designs by organizing tasks into groups that are "interlinked, yet sufficiently free of one another to adjust independently in a feasible amount of time."[24]

B. Interdependencies as a Determinant of Module Boundaries

To say that complex systems are more easily analyzed and developed if broken down into smaller subcomponents does not provide much

---

[15] Herbert A. Simon, *The Architecture of Complexity*, 16 PROC. AM. PHIL. SOC'Y 467 (1962).

[16] *Id.* at 474, 477.

[17] *Id.* at 474.

[18] *Id.* Simon noted that to be partially decomposable, a complex system must be hierarchical. Systems that consist of a repeating structure, such as a linear polymer or a crystal such as a diamond, are too flat to permit partial decomposition. *Id.* at 469, 476. To the extent that such systems are regarded as hierarchical, they represent the trivial case of a hierarchy with a span of one. *Id.* at 471.

[19] *Id.* at 477.

[20] *Id.* at 470–71.

[21] *Id.* at 473.

[22] For a classic early description, see Edsger W. Dijkstra, *The Structure of the "THE"-Multiprogramming System*, 11 COMM. ACM 341, 343 (1968).

[23] *Id.* at 344.

[24] CHRISTOPHER ALEXANDER, NOTES ON THE SYNTHESIS OF FORM 43 (1967); *see also id.* at 81–83, 116–19.

information about how such a decomposition should be implemented. Consistent with Simon's observation that interactions should be frequent and strong within subsystems and rare and weak across subsystems,[25] Carliss Baldwin and Kim Clark argue that the key to understanding the "points of natural division" between the elements of a modular architecture is *task interdependencies*.[26] System architects must group tasks that have strong interdependencies with each other inside the same protomodule, and then must "systematically . . . sever all dependencies known to exist across the protomodules."[27] This implies that a well-designed module is "a unit whose structural elements are powerfully connected among themselves and relatively weakly connected to elements in other units."[28] The boundaries between modules are located at the "thin crossing points" where interdependencies are the thinnest.[29] Design theorist Christopher Alexander similarly advocated including variables whose interactions are very rich within the same subset permits the subsets to act relatively independently.[30] Organizational theorist James Thompson advocated grouping reciprocally interdependent tasks together and to rely on key positions to serve as bridges to other groups.[31]

Encapsulating highly interdependent tasks within the same module also reduces the costs of communication and coordination by reducing the need

---

[25] *See supra* note 15 and accompanying txt; *accord* Simon, *supra* note 15, at 477 (observing that in nearly decomposable systems, "[i]ntra-component linkages are generally stronger than intercomponent linkages").

[26] BALDWIN & CLARK, *supra* note 4, at 64.

[27] *Id.* at 70.

[28] *Id.* at 63; *accord* MEILIR PAGE-JONES, THE PRACTICAL GUIDE TO STRUCTURED SYSTEMS DESIGN 101–03 (1980) (arguing that the measure of good modular design is the extent to which it minimizes interdependence between modules to make them as independent as possible); Sendil K. Ethiraj et al., *The Dual Role of Modularity: Innovation and Imitation*, 54 MGMT. SCI. 939, 939 (2008) ("It is generally accepted that a modular design is based on a principle of encapsulating interdependencies within self-contained units called modules and minimizing reciprocal interdependencies between modules."); Eric von Hippel, *Task Partitioning: An Innovation Process Variable*, 19 RESEARCH POL'Y 407, 409, 411–12 (1990) (arguing that innovation projects should be partitioned to minimize problem-solving interdependencies, which occur when changes in the information relating to one task require problem-solving in tasks on the other side of the partition); *see also* Melissa A. Schilling, *Toward a General Modular Systems Theory and Its Application to Interfirm Product Modularity*, 25 ACAD. MGMT. REV. 312, 316 (2000) (noting that the scope of modularity is determined by a system's "synergistic specificity," which arises when "components of the system . . . require such extensive interaction . . . that any change in a component requires extensive compensating changes in other components of the system, or else functionality is lost.").

[29] Carliss Y. Baldwin & Kim B. Clark, *Modularity in the Design of Complex Engineering Systems*, *in* COMPLEX ENGINEERED SYSTEMS: SCIENCE MEETS TECHNOLOGY 175, 199 (Dan Braha et al. eds., 2006);; *accord* von Hippel, *supra* note 28, at 409. A related, but distinct, approach focuses not on current interdependencies, but rather on future technological opportunities, by identifying "difficult design decisions or design decisions which are likely to change" and making sure that those decisions are locked within a single module. D.L. Parnas, *On the Criteria To Be Used in Decomposing Modules*, 15 COMM. ACM 1053, 1058 (1972).

[30] ALEXANDER, *supra* note 24, at 43, 64–65, 121–24.

[31] JAMES D. THOMPSON, ORGANIZATIONS IN ACTION 58–60 (1967).

for cross-boundary communication,[32] with cheaper cross-group communication leading to a larger number of modules.[33]

The initial modular design must be based on the architect's expertise.[34] Establishing these basic design rules has the effect of privileging certain parameters and declaring certain parts of the design space to be out of bounds.[35] No matter how well system architects believe they understand the systems they are designing, they cannot fully understand the way the various components interact with one another until they gain experience with the system by experimenting with different solutions.[36] If done badly, the proposed modularization can cause the system to perform suboptimally[37] or may require a redesign of the architecture.[38] In the worst case scenario, the interdependencies cannot be solved without redesigning the basic architecture, in which case the modularization will fail.[39]

This means that modular systems are never designed in a vacuum. Redesigns of current modular systems are framed by the existing modularization. Even with respect to new modularizations, they must respond and adjust to the initial design proposed by the architect, which because of the lack of perfect prescience typically includes some inefficiencies that the architect did not foresee.[40] The recursive and restricted nature of this adjustment process means that modularization emerges more through a process of "improvisation, bricolage, and drift" than from "planification and control."[41]

---

[32] von Hippel, *supra* note 28, at 409–10.

[33] *See* Scott Schaefer, *Product Design Partitions with Complementary Components*, 38 J. ECON. BEHAV. & ORG. 311, 320 (1999).

[34] BALDWIN & CLARK, *supra* note 4, at 68.

[35] *Id.* at 68–69.

[36] *See id.* at 254 ("Given such a high degree of complexity, it is simply not possible for designers to know enough about the system to eliminate all uncertainty. Thus each new design is fundamentally an experiment. Its outcome may be guessed, but it cannot be known ahead of time."); Ethiraj & Levinthal, *supra* note 9, at 172 ("Designers engage in acts of creation, but unlike a divine creator, they lack omniscience. Choices of modules are guess about appropriate decompositions—decompositions that even in reality are only partial (i.e., nearly decomposable).").

[37] *See* Langlois, *supra* note 9, at 26 (noting that "freezing the design rules too early may result in an inferior modularization").

[38] BALDWIN & CLARK, *supra* note 4, at 70 ("Imposing a design rule when one is ignorant of the true underlying interdependencies can lead to design failure. . . . The best course of action, if possible, is to rescind the original design rule, which was based on insufficient knowledge of the critical interdependencies.").

[39] *Id.* at 86 ("In newly modularized design, when the modules are first brought together, problematic interactions usually appear and workarounds to those problems must be devised. In the worst cases, the workaround will feed back into the module designs, and the modularization itself will fail.").

[40] *See* Schaefer, *supra* note 33, at 325–26 (showing that the problem devising a truly optimal modular design partition is NP complete in that its complexity increases exponentially with the size of the problem and concluding "it would seem unlikely that a firm could ever hope to uncover an optimal modular design partition for a complex product").

[41] Blanchette, *supra* note 10, at 1055; *accord* Ethiraj & Levinthal, *supra* note 9, at 160, 162, 172.

C. Interface Design and the Role of Abstraction and Hidden Information

Dividing a production process into blocks of interdependent processes is only one step in modularizing a system. In addition, the architects must structure the interactions between the modules.[42] To do this, each module must be reduced to an *abstraction*,[43] which is a simplification of reality "that suppresses details of elements that do not affect how [modules] use, are used by, relate to, or interact with other elements."[44]

Once a system architect predetermines how the modules will interact with one another, the architect must ensure that the programmers working on each module refrain from introducing new interdependencies that fall outside of the design. Individual programmers who have access to all of the information about the entire system will be tempted to use that information to improve the performance of the module on which they are working. The problem, according to David Parnas's seminal work, is that doing so can "disastrously increase the connectivity of the system" by introducing connections between modules that violate the goal of keeping the intermodule interdependencies to a minimum.[45]

Parnas identified an ingenious way to ensure that that the architect retains control over the overall structure. Instead of making all of the design information available to everyone, Parnas argued that architects could maintain the architecture's integrity by maintaining strict control over the distribution of information across the system.[46] Specifically, architects could carefully design the interfaces connecting modules so that they include only the information associated with the interdependencies that other modules were permitted to take into account and to withhold information about the module's inner working from the other modules.[47] Information included in the interface is part of the *visible information* shared between modules; design parameters not included in the interface are part of the *hidden information*.[48]

Hiding information about the highly interdependent tasks that are supposed to be encapsulated within modules makes those tasks opaque to

---

[42] BALDWIN & CLARK, *supra* note 4, at 77.

[43] For a seminal statement about the benefits of abstractions, see Dijkstra, *supra* note 22, at 343; *see also* HERBERT A. SIMON, THE SCIENCES OF THE ARTIFICIAL 17–20, 100–01 (1969). For an earlier statement advocating that programs be built around generalized software modules, see W.C. McGee, *Generalization: Key to Successful Electronic Data Processing*, 6 J. ACM 1 (1959).

[44] LEN BASS ET AL., SOFTWARE ARCHITECTURE IN PRACTICE 21 (2d ed. 2003).

[45] David L. Parnas, *Information Distribution Aspects of Design Methodology*, 1 INFO. PROCESSING 71: PROC. IFIP CONG. 71, at 339, 342 (1972).

[46] *Id.*

[47] *Id.* at 344; *accord* Parnas, *supra* note 29, at 1056 ("The second decomposition was made using 'information hiding' as a criterion. . . . Every module in the second decomposition is characterized by its knowledge of a design decision which it hides from all others. Its interface or definition was chosen to reveal as little as possible about its inner workings." (citation omitted)).

[48] *Id.* at 73–75.

other modules and thus prevents them from creating interconnections with those tasks that fall outside the architectural design.[49] This permits other modules to treat every other module as a "black box,"[50] which greatly reduces communication costs.[51] After the computer science community criticized Parnas's proposal as "radical" and unworkable,[52] Parnas took it upon himself to validate the concept by undertaking a real-world implementation of it.[53]

Interfaces thus establish the fundamental assumptions that the various modules in a system make about one another and predetermine the manner in which the modules will interact.[54] As such, decisions about what information to include in module interfaces thus represent critical architectural choices that must be made deliberately. As one noted modularity theorist pointed out, "Managers should understand that component interfaces are not minor technical details to be left to the engineering staff. Rather, interfaces will determine the range of strategic flexibilities the company will have to configure its products and adapt its processes in the future."[55] In other words, interface design determines the functionality of the system. Moreover, the points of interactions cannot be dealt with in an ad hoc manner. Instead, they must be maintained throughout the architecture in a consistent and systematic way.[56]

Limiting the number of interdependencies that pass between modules and the number of states associated with those interdependencies of which other modules have to take account simplifies the complexity of the system. For example, interdependencies sometimes form a circuit that loops back onto itself, such as occurs when task *A* depends on the design of task *B*, which depends on the design of task *C*, which recursively depends on the design of task *A*. When this occurs, the only way to determine the effect of the interdependencies is to cycle through a series of iterations until the design converges on a solution.[57] If the design architecture does not limit the number of interdependencies, system designers must test every possible combination of interdependencies, which can lead to a "combinatorial explosion of system variants" if left uncabined.[58] Modularity reduces the

---

[49] Langlois, *supra* note 9, at 22.

[50] BALDWIN & CLARK, *supra* note 4, at 91.

[51] Langlois, *supra* note 9, at 22.

[52] BROOKS, *supra* note 12, at 78.

[53] D.L. Parnas et al., *The Modular Structure of Complex Systems*, PROC. 7TH INT'L CONF. ON SOFTWARE ENG'G 408 (1984).

[54] Parnas, *supra* note 45, at 339.

[55] Ron Sanchez, *Fitting Together a Modular Approach*, MFG. ENG'R, Oct. 2002, at 216, 217; *see also* Ethiraj & Levinthal, *supra* note 9, at 172 (noting that in addition to innovation, "modularity has other important implications for strategy, organizational coordination, incentives, and so on").

[56] BALDWIN & CLARK, *supra* note 4, at 73.

[57] *Id.* at 51–52, 68–70; *see also* Ethiraj & Levinthal, *supra* note 9, at 160 (describing the trial-and-error process in which changes in one module of the Itanium chip would cause ripple effects that disrupted the work on other modules).

[58] BALDWIN & CLARK, *supra* note 4, at 272–75.

number of cycles required by narrowing and predetermining the universe of potentially relevant interdependencies and states.[59]

D. Requisite Variety as a Demand-Side Consideration

While insightful, the interdependency-driven approach to modularity is subject to an important limitation. In focusing on the technological characteristics of production processes, the modularity literature focuses exclusively on the supply side. This approach ignores the impact of changes on the demand side, such as changes in geographic dispersion and the desire for variety.

downplays the impact of changes in the exogenous environment, such as changes in demand. The addition of demand-side considerations provides another potential impetus for architectural change.

In this regard, the literature on general purpose technologies (GPTs) is instructive. GPTs are defined as technologies that are (1) widely used, (2) capable of ongoing technical improvement, and (3) enabling innovation in applications sectors.[60] In one sense, the persistence of GPTs represents something of a puzzle. Adam Smith's famous admonition that "the division of labor is limited by the extent of the market" suggests that as the demand for a product increases, its inputs should be produced by increasingly specialized vertically disintegrated firms.[61]

Timothy Bresnahan and Alfonso Gambardella provide an answer to this puzzle.[62] When the growth of the market consists not of an increase in the output of a single good, but rather the proliferation of distinct application sectors supported by the GPT, the effect of market growth is to entrench the GPT rather than prompt its vertical disintegration.[63] At the same time, Bresnahan and Gambardella recognize that "the broad applicability of a general specialty is not free" and that as the technological needs become increasingly differentiated, the superior matching benefits from localizing technology eventually dominates the benefits from generalization.[64]

In other words, increases in heterogeneity in demand increases the benefits of relying on custom inputs tailored to the a particular application's specific needs, which in turn makes GPTs relatively less attractive. At some point, demand becomes so heterogeneous that abandonment of the GPT becomes optimal.

---

[59] Dijkstra, *supra* note 22, at 344.

[60] Timothy Bresnahan, *General Purpose Technologies*, *in* 2 HANDBOOK OF THE ECONOMICS OF INNOVATION 761, 764 (Bronwyn Hall & Nathan Rosenberg eds., 2010).

[61] ADAM SMITH, THE WEALTH OF NATIONS _ (1776); *see also* George Stigler, *The Division of Labor Is Limited by the Extent of the Market*, 59 J. POL. ECON. 185 (1951).

[62] Timothy Bresnahan & Alfonso Gambardella, *The Division of Inventive Labor and the Extent of the Market*, *in* GENERAL PURPOSE TECHNOLOGIES AND ECONOMIC GROWTH 253 (Elhanan Helpman ed., 1998).

[63] *Id.* at 255, 261–62, 272–73.

[64] *Id.* at 255, 269–70, 273.

Increases in the volatility of the external environment can also create pressures for the architecture to change, as demonstrated by W. Ross Ashby's landmark Law of Requisite Variety, which studied ways that cybernetic systems comprised of multiple independent components can restore equilibrium when confronted with disturbances.[65] Ashby's model takes the form of a two-step game. In the first step, an external disturbance occurs. In the second step, a regulator chooses from the set of responses that it has designed to respond to each disturbance. Ashby illustrated his model with the matrix depicted in Figure 1.[66]

Figure 1

		Response		
		α	β	γ
Disturbance	1	*b*	*a*	*c*
	2	*a*	*c*	*b*
	3	*c*	*b*	*a*

Ashby added the further constraint that no column contain a repeated outcome. If so, the regulator need not change its move in response to different disturbances, which made the game "too easy to be of interest."[67] In other words, any rows with duplicate outcomes can simply be collapsed into a single row and treated as the same disturbance.

Suppose that the regulator's goal is to force good outcome *a* and avoid bad outcomes *b* and *c*. The number of responses at its disposal gives it the ability to do so. If disturbance 1 occurs, the regulator chooses response β. If disturbance 2 occurs, the regulator chooses response α. If disturbance 3 occurs, the regulator chooses response γ. Indeed, this table gives the regulator complete control over all outcomes, as it can also force outcomes *b* or *c* to occur regardless of the nature of the disturbance.

Ashby next explored the relationship between the number of disturbances, the number of responses at the regulator's disposal, and the number of possible outcomes. If there are only three types of disturbances and the regulator has three responses at its disposal, the regulator can design its responses to ensure that the game always results in a single good outcome, depicted in Figure 2 by the letter *k*.

---

[65] W. ROSS ASHBY, AN INTRODUCTION TO CYBERNETICS 206 (1956).

[66] *Id.* at 202.

[67] *Id.* at 204.

Figure 2

		Response		
		α	β	γ
Disturbance	1	*k*	-	-
	2	-	*k*	-
	3	-	-	*k*
	4	*l*	-	-
	5	-	*l*	-
	6	-	-	*l*
	7	*m*	-	-
	8	-	*m*	-
	9	-	-	*m*

Suppose, however, that a fourth type of disturbance emerges. The assumption that any particular outcome cannot appear more than once in any particular column dictates that the existence of this fourth type of disturbance necessarily means that the regulator can no longer restrict the game to a single outcome. Instead, it must tolerate at least two possible outcomes, *k* and *l*, and can do so for up to six disturbances. The emergence of a seventh type of disturbance would require the regulator to permit a third outcome, *m*.[68]

Stated more generally, if the regulator has *n* possible responses, it can ensure a single outcome so long as the number of disturbances is less than or equal to *n*, after which point the number of possible outcomes must increase to two. The number of disturbances can increase up to 2*n* before the number of possible outcomes increases again to three. The number of outcomes will increase for every additional *n* disturbances. In short, the total variety of outcomes cannot be fewer than the number of disturbances divided by the number of responses.[69] This implies that as the number of potential disturbances increases, the regulator either must be able to tolerate a wider range of outcomes or must increase the number of possible responses.

---

[68] *Id.* at 205–06.

[69] *Id.* at 206.

This insight leads directly to the Law of Requisite Variety. Simply stated, increasing the variety of responses is the only way to force down the variety of outcomes arising from the variety in disturbances.[70] In other words, only variety (in responses) can destroy variety (in disturbances).[71] Or in the words of another theorist, "to be efficaciously adaptive, the internal complexity of a system must match the external complexity it confronts."[72]

The Law of Requisite Variety has direct implications for the design of a modular architecture. Two noted modularity theorists have created a model in which certain changes in the environment give rise to a certain type of modularization. Most importantly for our purposes, fragmentation of the external environment (through increased "geographic dispersion, specialized market niches, and varied demands on the system") requires that the system adopt new responses to register the additional sources of variety.[73]

Furthermore, because the number of distinct disturbances in the environment determines the number of responses that must be built into the architecture, a regulator must ascertain which disturbances are in fact distinct. While it is tempting to treat every single disturbance as unique, increasing the number of responses is expensive.[74] On the other hand, grouping multiple disturbances together risks oversimplification, which in turn would leave the system unable to respond to important contingencies.[75] The designer of a modular system must determine which disturbances must be taken into account and which can be safely ignored. Such an exercise cannot be conducted in the abstract and must reflect the motivations and expectations around which the system was built[76] as well as the context surrounding it.[77]

The demand-side literature thus augments the interdependency-focused by adding a second potential source of architectural change. Unlike interdependencies, which because they reflect the nature of the production function are necessarily supply-side considerations, factors such as geographic dispersion, market differentiation, and the desire for variety are quintessential demand-side considerations. In this manner, the addition of demand-side factors is consistent with the core economic insight of neoclassical economics that rejected value as determined solely by supply-

---

[70] *Id.*
[71] *Id.* at 207.
[72] Max Boisot & Bill McKelvey, *Complexity and Organization-Environment Relations: Revisiting Ashby's Law of Requisite Variety*, *in* SAGE HANDBOOK OF COMPLEXITY AND MANAGEMENT 279, 279 (Peter Allen et al. eds., 2011).
[73] J. Douglas Orton & Karl E. Weick, *Loosely Coupled Systems: A Reconceptualization*, 15 ACAD. MGMT. REV. 203, 207, 208, 217 fig.1 (1990).
[74] Orton & Weick*, supra* note 73, at 283.
[75] *Id.* at 279.
[76] *Id.* at 283.
[77] Jeff Goldstein, *Requisite Variety and the Difference That Makes a Difference: An Introduction to W. Ross Ashby's "Variety, Constraint and Law of Requisite Variety,"* 13 EMERGENCE: COMPLEXITY & ORG. 190, 192–97 (2011).

side factors and replaced it with a vision in which value is determined by the interaction of both supply and demand.[78]

The addition of demand-side considerations suggests that the debate over whether innovation is driven by consumer-driven "demand pull" or technologically determined "supply push" is based on something of a false dichotomy. Interestingly, these two frameworks can have very different implications for increased complexity. While an increase in interdependencies can lead to an increase in module size in order to encapsulate more of them within a single module, an increase in the complexity of the external environment implies a decrease in module size in order to allow the system to better deal with the underlying variety.

It is tempting to conclude that increases in demand heterogeneity can be met simply by increasing the number of modules and simplifying the interfaces connecting them.[79] Dividing a system into a larger number of modules increases the number of possible product configurations combinatorially.[80] Greater configurability has little value when demand is homogeneous, but increases in value as consumer preferences become more diverse.[81] Variety also provides the most benefits when technologies are changing rapidly and the level of competition is intense[82] and provides fewer advantages when the technological and business environment is relatively stable.[83]

As I shall later discuss in greater detail, this perspective ignores the fact that modularity promotes certain types of new product configurations, while inhibiting others.[84] Modularity makes feasible innovations that involve new combinations or the improvement of existing modules. At the same time, predefining the relationship between certain clusters of tasks and the information passing between them forecloses innovations that would recombine tasks in a fundamentally different way.[85] The impact of modularity on product configuration is thus ambiguous. Increasing the number of modules thus only facilitates recombinations that "are compatible

---

[78] ALFRED MARSHALL, PRINCIPLES OF ECONOMICS (London, MacMillan 1890).

[79] Orton & Weick, *supra* note 73, at 210 (noting that "registering [sources of variety] improves when elements become more numerous and the constraints among them weaken").

[80] Schilling, *supra* note 28, at 323.

[81] *Id.* at 317, 324–25.

[82] Melissa A. Schilling & H. Kevin Steensma, *The Use of Modular Organizational Forms: An Industry-Level Analysis*, 44 ACAD. MGMT. J. 1149, 1162 (2001).

[83] Langlois, *supra* note 9, at 23, 24.

[84] *See infra* Part III.C.

[85] *See* Rebecca M. Henderson & Kim B. Clark, *Architectural Innovation: The Reconfiguration of Existing Product Technologies and the Failure of Established Firms*, 35 ADMIN. SCI. Q. 9, 11–13 (1990) (distinguishing between changes in knowledge with respect to individual components and changes in architectural knowledge "about the ways in which the components are integrated and linked together into a coherent whole").

with the overall system architecture."[86] At the same time, it preempts recombinations that would reorganize the hierarchy in a different manner.[87]

E. Testing and Integration

Another essential element of any modular design is a means for testing and system integration.[88] Because any initial modular design is necessarily incomplete, the architects of any modular system must provide some way to ensure that the system is functioning properly and that each module is functioning in accordance with the overall design.[89] Indeed, such tests constitute a necessary precondition to distributed production systems in which different modules are produced by different firms. In the words of Baldwin and Clark, "The testable, verifiable dimensions of the module are the foundation that supports arm's length-contracts and market transactions."[90] Indeed, "without tests, there is no way to know what is being bought and sold."[91]

The need to test each possible state one module can occupy against every possible combination of states that other modules can occupy represents what Baldwin and Clark have called the "Achilles' heel of modular designs."[92] Without modularity, this problem would have been far worse. Dijkstra's seminal analysis concluded that under integrated, nonmodular architectures, "the number of 'relevant states' would have exploded to such a height that exhaustive testing would have been an illusion."[93] The failure to limit the number the ways different components can interact with one another leads to a "combinatorial explosion of system variants."[94]

If the only available tests operate at the system level, architects cannot escape the combinatorial explosion of states that must be tested against one another.[95] Thus, when system-level tests are the only option, designers inevitably tend towards unmodularized, integrated designs and low levels of experimentation.[96] The exponential nature of the combinatorial explosion means that this problem cannot be overcome simply by making system-level testing more rapid and less expensive.[97]

---

[86] Schilling, *supra* note 28, at 315.

[87] Kim B. Clark, *The Interaction of Design Hierarchies and Market Concepts in Technological Evolution*, 14 RES. POL'Y 235, 249 (1985).

[88] BALDWIN & CLARK, *supra* note 4, at 72, 77, 246.

[89] *Id.* at 70–72, 77.

[90] *Id.* at 380

[91] *Id*.

[92] *Id.* at 272.

[93] Dijkstra, *supra* note 22, at 344.

[94] BALDWIN & CLARK, *supra* note 4, at 273–75.

[95] *See supra* note _ and accompanying text.

[96] BALDWIN & CLARK, *supra* note 4, at 275.

[97] *Id.* at 276.

The situation is quite different when tests can be conducted on individual modules. Restricting the number of interdependencies and the number of states associated with those interdependencies that other parts of the system must take into account dramatically reduces the number of combinations that must be tested.[98] Module-level testing permits the evaluation of individual modules against the module metrics inherent in the design rather than the entire universe of possible combinations.[99] Later studies have confirmed this insight, showing that the cost of testing grow sublinearly with the number of physical elements for modular architectures, whereas the cost of testing grows exponentially for integrated architectures.[100]

The presence of robust module-level testing is thus a necessary precondition for any modular system in which different firms produce different components. Indeed, the ability to verify that each module is working as environed by the architecture is a necessary condition to third-party, distributed production of modular components.[101] Over time, new design rules are created to address the out-of-module interdependencies revealed by the testing. Eventually, the design rules may become so complete that testing and system integration can be conducted by the module designers and end users or may even disappear.[102] At this point, the architects will have modularized themselves out of a job.[103]

II. THE BENEFITS OF MODULARITY

It is widely recognized that modular architectures provide a number of benefits. Dividing the overall system into a series of subsystems makes complexity more manageable. It accelerates innovation by allowing experiments on different parts of the system to proceed in parallel. It facilitates the division of labor across different work groups and firms. It increases value by allowing the postponement of key technical decisions. Lastly, it promotes flexibility by allowing individual modules to be changed without disturbing the other parts of the complex system.

A. Making Complexity More Manageable

Breaking down complex systems into subparts helps make them more manageable. As an initial matter, partitioning a large system into smaller modules divides the larger task into subtasks “small enough to be fully

---

[98] *Id.* at 277–79.

[99] *Id.* at 277–78.

[100] Christoph H. Loch et al., *Parallel and Sequential Testing of Design Alternatives*, 45 MGMT. SCI. 663, 673 (2001).

[101] *Id.* at 380.

[102] *Id.* at 77, 250, 268.

[103] *Id.* at 268–69.

comprehend[ed] by a single individual."[104] In this manner, modularity allows architects to work on individual modules in isolation without having to grapple with the architecture as a whole.[105] Conversely, an architect can understand the overall system without having to understand the details of every module.[106]

Moreover, the fact that the design architecture predetermines the relationships between modules makes the elements and functions of complex systems easier to describe and understand.[107] One needs to know only the design rule defining the interactions with adjacent modules.[108] The details of the interactions contained within modules can be ignored without losing any relevant information.[109] Moreover, reducing the total number of interdependencies and organizing them in a systematic, predictable ways yields particularly strong benefits in reducing the cost of testing programs.

B. Enabling the Division of Labor Across Work Groups and Firms

Modularity also has major organizational implications. As Parnas noted in his seminal analysis on information hiding, modularity allows "separate groups [to] work on each module with little need for communication."[110] Richard Langlois and Paul Robertson similarly note how modularity facilitates the division of labor by enabling autonomous innovation that requires little coordination among modules."[111] Intermodule coordination is enforced not by managerial authority, but rather by the implicit relationships embedded in the module interfaces without the need for hierarchy of managerial authority.[112] It is the information structures embedded in the interfaces that bind the parts of the modular architecture together.[113]

The same independence that permits modules to be developed by separate work groups also facilitates their development by different firms.[114] Because the coordination is embedded in the information structure of the architecture,

---

[104] Blanchette, *supra* note 10, at 1046.

[105] Parnas, *supra* note 29, at 1054, 1056.

[106] *See* Parnas et al., *supra* note 53, at 410.

[107] *See* Simon, *supra* note 15, at 477.

[108] *See* BALDWIN & CLARK, *supra* note 4, at 91 ("The design rules are the visible information . . . .").

[109] *Id.*

[110] Parnas, *supra* note 29, at 1054.

[111] Langlois & Robertson, *supra* note 139, at 302.

[112] Ron Sanchez & Joseph T. Mahoney, *Modularity, Flexibility, and Knowledge Management in Product and Organization Design*, 17 STRATEGIC MGMT. J. 63, 64, 65, 66, 73 (1996).

[113] *Id.* at 66.

[114] *See* Baldwin & Clark, *supra* note 7, at 84, 85; Langlois, *supra* note 9; Sanchez & Mahoney, *supra* note 112; Schilling & Steensma, *supra* note 82; Ulrich, *supra* note 138, at 435; *cf.* Ethiraj & Levinthal, *supra* note 9, at 172 (noting that firms may not have sufficient internal variety to permit sufficient experimentation). For more cautious appraisals, see Stefano Brusoni, *The Limits to Specialization: Problem Solving and Coordination in "Modular Networks,"* 26 ORG. STUD 1885 (2005); Stefano Brusoni & Andrea Prencipe, *Unpacking the Black Box of Modularity: Technologies, Products and Organizations*, 10 INDUS. & CORP. CHANGE 179 (2001); Glenn Hoetker, *Do Modular Products Lead to Modular Organizations?*, 27 STRATEGIC MGMT. J. 501 (2006).

firms can jointly organize tasks without being subject to common managerial authority.[115] Moreover, the existence of modules separated by predefined interfaces not only allows development of different modules by different firms; it also creates more entry points for new firms, which in turn can promote greater competition.[116]

The interdependency-oriented theory of modularity is closely related to, but analytically distinct from, Ronald Coase's theory of the firm.[117] Coase argued that firms decide whether to perform certain functions internally or to contract them out by comparing the cost of negotiating a contract to outsource the function with the cost of monitoring and overseeing the performance of the function internally.[118] Modularity theory adopts a similar logic, but focuses on the costs of coordination and testing, with highly interdependent tasks being performed within a single module (and thus within a single firm) and highly independent tasks being performed by different modules (and thus potentially by different firms intermediated by a transaction).[119] The key difference is that transaction costs are potentially technologically defeasible and tend to decline over time. The complexities of testing multiple states are more fundamental and unlikely to dissipate with changes in technology.

From this perspective, the relevant primitive units of analysis are not transactions, but rather tasks and their interdependencies.[120] Indeed, it is the underlying task structure that determines the firm's boundaries, rather than considerations such as bounded rationality, uncertainty, vulnerability to opportunism, and information asymmetry that motivate transaction cost economics.[121] Indeed, particularly dense areas of interdependencies can create "transaction-free zones" that are best encapsulated within corporations.[122] This in turn creates what some have called a "modularity theory of the firm" in which "[f]irms arise as islands of nonmodularity in a sea of modularity."[123]

Modularity theory thus represents a return to a more technological vision of the theory of the firm, in which the scope of a firm is determined by the interconnections between steps in the production process.[124] This approach

---

[115] Sanchez & Mahoney, *supra* note 112, at 64, 66–68.

[116] Langlois & Robertson, *supra* note 139, at 301.

[117] *See* Christopher S. Yoo, *Coase Revisited: Property Theory and Emerging Technologies*, 160 U. PA. L. REV. 2189, 2205 (2012).

[118] R.H. Coase, *The Theory of the Firm*, 4 ECONOMICA 386 (1937).

[119] Langlois, *supra* note 9, at 34.

[120] Carliss Y. Baldwin, *Where Do Transactions Come From? Modularity, Transactions, and the Boundaries of Firms*, 17 INDUS. & CORP. CHANGE 155, 156, 162 (2007).

[121] *Id.* at 160.

[122] *Id.* at 181–83.

[123] Langlois, *supra* note 9, at 34.

[124] Perhaps the prototypical example is vertical integration in the steel industry to take advantage of thermal efficiencies, in which the integration of refining ore into ingot and then recasting ingot into forms obviated the need to reheat the molten metal. *See, e.g.*, F.M. SCHERER & DAVID ROSS, INDUSTRIAL

responds to the criticism that transaction cost economics "focuses on the conditions of exchange, to the neglect of the conditions of production."[125]

C. Promoting Flexibility

In addition to making complexity more manageable, decomposing the larger system into subsystems that interact with one another in predefined ways allows designers working on each subsystem to focus exclusively on the aspect of the problem to which they have been assigned and ignore all other aspects. As Simon pointed out, "Decomposing [a complex structure] into semi-independent components corresponding to its many functional parts" permits "[t]he design of each component [to] be carried out with some degree of independence of the design of the others, since each will affect the others largely through its function and independently of the details of the mechanisms that accomplish the function."[126]

The limitation of intermodule interdependencies also facilitates innovation by making it possible to modify individual modules without worrying unduly about the impact of those changes on the other parts of the system. Because tasks related to information hidden inside a module necessarily cannot affect other modules, those tasks can be modified without affecting other modules or having to inform them about those modifications.[127] As long as the interfaces remain stable and the number of interactions remain carefully constrained, software designers can change individual modules with a manageable amount of testing.[128] This should be true with respect both to changes to individual modules and to replacements of entire modules with new modules.[129]

The relative isolation of individual modules makes it easier to incorporate better solutions without having to make changes to the entire system.[130] At the same time, the intramodule flexibility provided by modularity accommodates uncertainty by making it easier to incorporate subsequent improvements into the system.[131] Modularity facilitates experimentation, because any disturbances that may arise from any changes can be localized and analyzed within a single module,[132] which buffers the overall system against ripple effects associated with subsequent changes.[133] The greater

---

MARKET STRUCTURE AND ECONOMIC PERFORMANCE 94 (3d ed. 1980); George J. Stigler, *The Extent and Bases of Monopoly*, AM. ECON. REV., June 1942, at 1, 22.

[125] Michael G. Jacobides & Sidney G. Winter, *The Co-Evolution of Capabilities and Transaction Costs: Explaining the Institutional Structure of Production*, 26 STRATEGIC MGMT. J. 395, 398 (2005).

[126] SIMON, *supra* note 43, at 148.

[127] BALDWIN & CLARK, *supra* note 4, at 73; Parnas, *supra* note 29, at 1054.

[128] Parnas et al., *supra* note 53, at 409.

[129] Ethiraj & Levinthal, *supra* note 9, at 164.

[130] BALDWIN & CLARK, *supra* note 4, at 91.

[131] *Id*.

[132] *Id.* at 142; Sanchez & Mahoney, *supra* note 112, at 64–65.

[133] Orton & Weick, *supra* note 73, at 214.

flexibility allows the network to evolve in response to new technological developments.[134]

D. Facilitating Parallel Experimentation

Modularity not only permits certain technological decisions to be postponed. It also allows that experiments with different technical solutions to be conducted simultaneously in parallel.[135] This not only reduces the time needed to complete the design of the complex system.[136] Isolating each module from the effects of other modules means that "its value is no longer hostage to that of distant elements of the system" and that "[a]s a result, each module becomes an independent unit of selection and a point of potential value creation within the overall system."[137]

One of the primary virtues in segmenting tasks in this manner is that the relative independence between modules allows them to be developed simultaneously in parallel.[138] Simultaneous experimentation with various alternative approaches increases the rate of learning by trial and error.[139] Although parallel testing proceeds more rapidly than serial testing, it does forego the benefits of learning between tests, which can increase the number of tests performed. Running multiple concurrent experiments through parallel testing nonetheless remains favored if learning from test to test is minimal and when testing is slow, inexpensive, or imperfect.[140]

The existence of multiple dimensions along which a design can be improved technologically unlocks the value identified by real option theory. When potential technological improvements arise for systems that consist of a single, interconnected design, the architect only has a single option: whether to adopt the improvement or not.[141]

Modular designs create many more options. At a minimum, they create as many options are there are modules.[142] In addition, the architect can combine improvements and nonimprovements to different modules in

---

[134] BALDWIN & CLARK, *supra* note 4, at 216.

[135] *Id.* at 91, 238.

[136] *See id.* at 91; *accord* Parnas, *supra* note 29, at 1054 (noting that under modularity, "development time should be shortened because separate groups would work on each module with little need for communication"); Simon, *supra* note 15, at 473 (observing how the existence of stable intermediate forms can cause complex systems to evolve more rapidly).

[137] BALDWIN & CLARK, *supra* note 4, at 237.

[138] *See id.* at 91; Karl Ulrich, *The Role of Product Architecture in the Manufacturing Firm*, 24 RESEARCH POL'Y 419, 435 (1995); *see also* Ethiraj & Levinthal, *supra* note 9, at 160 (noting that "modularity offers the advantage of parallelism").

[139] Richard N. Langlois & Paul L. Robertson, *Networks and Innovation in a Modular System: Lessons from the Microcomputer and Stereo Component Industries*, 21 RESEARCH POL'Y 297, 301 (1992).

[140] Loch et al., *supra* note 100, at 664, 674.

[141] BALDWIN & CLARK, *supra* note 4, at 236, 238, 252.

[142] *Id.* at 236.

different ways. This in turn creates a myriad of possible trajectories for the system.[143]

Consider, for example, a two-module system. Instead of a binary choice between incorporation of the new technology into the interconnected system, modularity raises the possibility of partial adoption, that is, adoption with respect to module *A*, but not *B*, or vice versa. In this sense, dividing an interconnected system into two modules increases the number of options from two to four. By extension, a three-module design yields eight options. In other words, as the number of modules (n) increases linearly, the number of options increases exponentially (2^n). This approach assumes that the modularizations are symmetric, but the analysis generalizes to more complex assumptions.

Figure 3

Adopt for system	yes	1
	no	2

		Adopt for module *B*	
		yes	no
Adopt for module *A*	yes	1	3
	no	2	4

		Adopt for Module *C*			
		yes		no	
		Adopt for module *B*		Adopt for module *B*	
		yes	no	yes	no
Adopt for module *A*	yes	1	3	5	7
	no	2	4	6	8

Modularity not only increases the number of options; it increases their value. Note first that if an architect had to accept both positive and negative values, the expected value of any experiment would be zero. But an architect can simply fail to incorporate the value of an experiment into the design. Thus, all positive experiments have a positive impact on value, while all negative experiments have no effect at all. The net result of cutting off the left-hand tail of the distribution means that the average experiment would have positive value. A simple calculation of the area under the positive half of the curve indicates that the expected value is 0.3989 times the standard deviation, σ.[144]

Baldwin and Clark's model also assumes that the variance increases as the system becomes more complex, with variance increasing linearly with the number of modules n (i.e., variance $\sigma_n^2 = \sigma^2 n$). Because the standard

---

[143] *Id.* at 238.

[144] *Id.* at _.

deviation is by definition the square root of the variance, the standard deviation equals $\sigma\sqrt{n}$.[145] This has the effect of widening the distribution of possible outcomes, increasing both the number of high value and low value possibilities. In the case of the IBM System/360, dividing the system into twenty-five modules increased the value of the system by five times.

The value is increased still further if multiple experiments are run with respect to each of the modules. In effect, each of these experiments is a random draw on the same distribution of expected value for each module, with only the one with the highest value being incorporated into the design. The existence of these multiple draws in effect increases the likelihood that the best experiment will be significantly higher than the expected result.

In essence, enabling multiple experiments taps into Robert Merton's key insight into the value of options, which is that a portfolio of options is more valuable than an option on a portfolio.[146] Indeed, experimentation increased the value of the IBM System/360 by an additional five times, which means that combining modularity with experimentation increased the value of the IBM System/360 by twenty-five times.[147] Whether this would be profitable depends on whether the increase in value exceeds the fixed costs of modularization and the cost of experimentation.

E. Permitting Permissionless Innovation

Modularity has the additional advantage of making innovation permissionless.[148] Encapsulating certain parameters within modules many that innovators can innovate with respect to those hidden parameters without disturbing the rest of the system.[149] Control is maintained by the design rules embedded in the architecture rather than by any firm or actor. Any person can connect so long as they comply with the design rules.[150] In short, the architecture is set up so that hidden-module designers, that is designers of components that are completely encapsulated within a module, "do not need to have detailed knowledge about the whole system under development. Hidden-module designers have only to master the design rules."[151]

This permits module designers to act independently of the system architects as well as one another.[152] This in turn enables decentralized decisionmaking without causing the system to lose coordination.[153] Modularity also means that those who wish to experiment at the module

---

[145] *Id.* at _.

[146] *Id.* at 259.

[147] *Id.* at _.

[148] *Id.* at 14, 223, 252, 264, 268, 336–37, 347–48.

[149] *Id.* at 348.

[150] *Id.* at 268, 306.

[151] *Id.* at 347.

[152] *Id.* at 348.

[153] *Id.* at 268.

level do not need to obtain permission from the system designers before doing so.[154] The would-be module experimenters need only make sure that their module revisions comply with the design rules of the architecture.[155] The result is to disperse and decentralize the process of innovation. It also has the effect of moving the options from the center of the system of the modules.[156] This in turn shifts the value gained from those experiments from the center of the system to the modules.[157]

III. MODULARITY'S LATENT COSTS

Although modularity yields substantial benefits, it should not be regarded as a panacea. On the contrary, an early IETF document authored by MIT computer scientist and then-chief protocol architect David Clark observed that network engineers are often unpleasantly surprised to find that "modularity is one of the chief villains in attempting to obtain good performance."[158] This statement underscores the importance of understanding the implicit tradeoffs underlying every modularity decision.

A. The Sunk Costs of Modularization

The production of a modular system takes place in three distinct phases. During the first stage, the architects formulate the design rules that determine the scope of the modules and what information to make visible in the interfaces through which each module will interact with the others. During the second stage, smaller groups engage in independent parallel activity to develop the modules. During the third stage, the various modules are tested and integrated into a single system.[159]

One of the most straightforward costs of modularity results from the need to construct and disseminate a modular design up front. Such efforts can be extremely time consuming and demanding.[160] The presence of the initial design phase makes "modular systems . . . much more difficult to design than comparable interconnected systems."[161] The architects must know a great deal about the relevant products and processes and the interdependencies in the relevant tasks and must incorporate their insights into design rules specified long in advance.[162] The cost of developing these design principles

---

[154] *Id.* at 14 ("No architect had to give permission for these changes to take place; the possibilities were inherent in the modularity of the design itself."); *accord id.* at 223, 268.

[155] *Id.* at 223, 268, 306.

[156] *Id.* at 268.

[157] *Id.* at 268.

[158] David D. Clark, *Modularity and Efficiency in Protocol Implementation* 1 (Request for Comments 817, 1982), *available at* http://tools.ietf.org/pdf/rfc817.

[159] BALDWIN & CLARK, *supra* note 4, at 72.

[160] *Id.* at 86, 247; Baldwin & Clark, *supra* note 29, at 199; Langlois, *supra* note 9, at 23.

[161] Baldwin & Clark, *supra* note 7, at 86.

[162] *Id.*

represents a sunk cost investment that must be recouped.[163] Moreover, the inevitable uncertainty about the design necessarily means that some chance always remains that the architecture is badly done, in which case it will underperform.[164]

B. Inefficiency

Beyond the sunk costs of creating the architecture, modularity embodies a more fundamental tradeoff. The generality of modular systems makes them flexible, but only at the cost of some efficiency. Consider for example, the problem of packing as many objects into a closet as possible.[165] The most efficient way to do so would be place each object in the closet one at a time, carefully taking into account the idiosyncrasies of each object's shape. The problem is that the interdependencies between the objects mean that any attempt to retrieve or shift the objects after they have been packed into will threaten to perturb the entire arrangement. A classic solution to this problem would be to pack all of the objects into storage bins before placing them in the closet. The bins would make the objects easier to shift, but would store the items less efficiently on account of the empty space in each of the bins and any gaps left between the bins when packing them into the closet

The same flexibility-efficiency tradeoff arises with respect to modularity.[166] As Baldwin and Clark note, the benefits of modularity must be traded off against the fact that "designers will lose the ability to explore some parts of the space of designs—in effect, the architects will restrict the search, declaring some parts of the design space to be out of bounds."[167] The problem, long recognized by computer scientists, is that generality exacts a cost. For example, McGee's seminal paper on generalization is based on the premise that handtailored programs, while expensive to create and modify, are usually more efficient.[168] Philip Agre notes that "all abstractions . . . entail[] a decrease in efficiency," providing one example "located at the extreme in the space of trade-offs between freedom of abstraction and efficiency of implementation" which "offer[ed] a maximum of freedom with a minimum of efficiency" as well as another example that presented "the trade-off between efficiency of implementation and freedom of abstraction in its most radical form."[169] Another commentator has similarly noted that "the

---

[163] BALDWIN & CLARK, *supra* note 4, at 247; *accord* Langlois, *supra* note 9, at 23 (noting that despite modularity's benefits, "there is no free lunch" in that "[a] well-decomposed modular system must pay a kind of fixed cost that an integrated system need not pay").

[164] BALDWIN & CLARK, *supra* note 4, at 254, 257; Ethiraj, *supra* note 9, at 172.

[165] The example is taken from Blanchette, *supra* note 10, at 1047.

[166] *See* KARL T. ULRICH & STEVEN D. EPPINGER, PRODUCT DESIGN AND DEVELOPMENT 202 (5th ed. 2011); Blanchette, *supra* note 10, at 1048.

[167] BALDWIN & CLARK, *supra* note 4, at 68.

[168] McGee, *supra* note 43, at 2.

[169] PHILIP AGRE, COMPUTATION AND HUMAN EXPERIENCE 68, 73, 76 (1997).

more a language's constructs abstract away from the underlying physical machine, the less efficient the resulting code tends to be."[170]

The tradeoff between flexibility and generality provides insights into when the benefits of modularity exceed the costs. As noted earlier, modularity yields the most benefits when demand is heterogeneous, technology is changing rapidly, and competition is intense.[171] These considerations affect not just whether to modularize, but also how to modularize. For example, increasing the number of modules increases testing expense, as tiny modules can cause a combinatorial explosion of testing permutations.[172] Sendil Ethiraj and Daniel Levinthal develop an interesting model integrating these considerations, which trades off the destabilizing effects from having too many modules against the slower rate of innovation from having too few modules. They show that creating more modules unlocks parallelism and avoids premature fixation on inferior designs, but leads to more intermodule interdependencies and requires more testing.[173] Although they tested multiple scenarios,[174] their general conclusion is that having too many modules posed greater risks than having too few.[175]

C. Restriction of Certain Types of Innovation

Decisions about how to implement modularity may affect applications differently.[176] To use an example from software design, the process for releasing memory that is no longer needed can require all programs to stop temporarily, a burden that will rest particularly heavily on real-time applications.[177]

Decisions about what information to hide and what information to make visible can also have a differential impact on applications. The set of visible information necessarily has a direct impact on the functionality of the system. Because it naturally limits the type of information that can pass between modules, any functionality that depends on any other information must be handled within one particular module. Consequently, modularization represents a precommitment to the idea that certain functions

---

[170] Blanchette, *supra* note 10, at 1047.

[171] *See supra* notes _ and accompanying text.

[172] BALDWIN & CLARK, *supra* note 4, at 272, 275; Loch et al., *supra* note 100, at 673.

[173] Ethiraj & Levinthal, *supra* note 9, at 160.

[174] Ethiraj and Levinthal considered two types of innovation: internal (which they call local search) and borrowing modules from other companies (which they call recombination). They also compared the results when one assumed that the architects have the design perfect with the results when they do not. *Id.* at 160–61, 164–65.

[175] *Id.* at 170–72.

[176] *See* Blanchette, *supra* note 10, at 1047 (noting that "the tradeoffs implied by modularity will not affect all applications equally, or even the same application under all circumstances").

[177] *Id.*

must be performed by certain modules. At the same time, it reflects a judgment of the type of functions that require coordination across modules.

The advent of USB ports to connect printers to personal computers provides a useful example. The interface between these two devices is designed to include the minimum amount of information that must pass between these two modules. Many aspects about printers are extremely dependent on whether laser printing, inkjet printing, or some other technology is used, the volume of the printer, and other key design decisions. Information hiding allows these interdependencies to be encapsulated within the printer so that the personal computers connected to those printers do not need to know anything about how any particular printer solves those functions. As long as the personal computer presents its data in accordance in the appropriate format, the printer should function properly. In addition, so long as printers are prepared to process the visible information presented by personal computers through the interface, they remain free to redesign any technical aspects that solely involve the hidden information. Note, however, that in limiting the information that can pass between personal computers and printers, the design of the interface inevitably imposes limits on the ways these two devices can interact.

Thus, although flexibility is generally regarded as one of the advantages of modular systems, closer inspection reveals that modular structures facilitate only certain types of innovation while impeding others. Specifically, modular systems are very good at promoting improvements and replacements of individual modules, which require little coordination with other modules. They are less accommodating to systemic innovations that reorganize the ways that modules interact with one another.[178]

The literature on *design hierarchies* suggests a similar conclusion.[179] Design hierarchies exist when innovations are not standalone technological developments, but rather are part of a web of interdependent technological processes. Not all of the components of a design hierarchy are equally important. Some have a higher degree of connection to other components than others.[180]

---

[178] Langlois, *supra* note 9, at 25–26 (distinguishing between modular innovation, which takes place within modules, and architectural innovation, which changes the relationship between modules, and observing how an architecture "can get stuck in an inferior modularization"); Langlois & Robertson, *supra* note 139, at 302 (citing David J. Teece, *Profiting from Technological Innovation: Implications for Integration, Collaboration, Licensing and Public Policy*, 15 RESEARCH POL'Y 285, 288 (1986)).

[179] For an earlier discussion, see Christopher S. Yoo, *Product Life Cycle Theory and the Maturation of the Internet*, 104 NW. U. L. REV. 641, 655–56 (2010).

[180] Clark, *supra* note 87, at 243; Johann Peter Murmann & Koen Frenken, *Toward a Systematic Framework for Research on Dominant Designs, Technological Innovations, and Industrial Change*, 35 RESEARCH POL'Y 925, 940–42 (2006); Michael L. Tushman & Johann Peter Murmann, *Dominant Designs, Technology Cycles, and Organizational Outcomes*, 20 RESEARCH ORG. BEHAV. 231, 249–51 (1998); Michael L. Tushman & Lori Rosenkopf, *Organizational Determinants of Technological Change: Toward a Sociology of Technological Evolution*, 14 RESEARCH ORG. BEHAV. 311, 334 (1992).

Commentators have recognized that design hierarchies represent something of a mixed blessing from the standpoint of innovation. On the one hand, they facilitate innovations that are consistent with the hierarchy. On the other hand, they discourage innovations that are inconsistent with the hierarchy.[181] Thus, any innovation that restructures the way a design hierarchy operates must overcome the challenge of coordinating the behavior of multiple autonomous actors, each of which is pursuing its own agenda.[182] This is particularly true for components of the design hierarchy that are particularly tightly connected with other components.[183] Because changes to core components necessarily require extensive changes to peripheral components, even changes that ultimately prove successful in the long run are likely to be accompanied in the short run by poorer system performance and widespread institutional resistance.[184] Empirical studies have confirmed that changes to core components are harder to displace.[185]

Characterizing innovation as residing solely on the pro-modularity side of the tradeoff, as sometimes occurs in the literature, adopts an unnecessarily narrow view.[186] In terms of innovation, modularity is more properly regarded as something of a two-edged sword that simultaneously promotes innovations that are consistent with the architecture and obstructs innovations that require a different configuration of tasks.

D. Limitations on Alternative Institutional Forms

As discussed above, some scholars believe that "products design organizations"[187] and that modular product designs necessarily lead to modular organizational designs.[188] If so, the decentralized, mix-and-match world of modularity exists in considerable tension with the insights of the New Institutional Economics.

Consider the landmark article by David Teece exploring why the entrepreneurs who come up with innovative ideas often do not end up being the person who benefits most from them.[189] Teece pointed out that most innovations are not products by themselves. Instead, they must be combined

---

[181] Clark, *supra* note 87, at 249.

[182] Henry W. Chesbrough & David J. Teece, *When Is Virtual Virtuous?: Organizing for Innovation*, HARV. BUS. REV., Jan.–Feb. 1996, at 65, 67, 68; Jacobides & Winter, *supra* note 125, at 404.

[183] Murmann & Frenken, *supra* note 180, at 942–43.

[184] *Id.* at 942.

[185] Alan MacCormack et al., *The Impact of Component Modularity on Design Evolution: Evidence from the Software Industry* (Harv. Bus. School Working Paper No. 08-038, 2007), *available at* http://www.hbs.edu/research/pdf/08-038.pdf.

[186] *See, e.g.*, JOHN PALFREY & URS GASSER, INTEROP: THE PROMISE AND PERILS OF HIGHLY INTERCONNECTED SYSTEMS 75–88 (2012) (describing interoperability as promoting innovation while having negative effects on privacy and security); ZITTRAIN, *supra* note 2, at 40–43 (characterizing modularity as promoting generativity, but being vulnerable to security risks).

[187] Sanchez & Mahoney, *supra* note 112, at 64.

[188] *See supra* note 114 and accompanying text.

[189] Teece, *supra* note 178, at 285.

with other complementary assets before they can form a marketable product.[190] When that is the case, economic success depends as much on the ability to bargain effectively with owners of complementary assets as it does on having ironclad patent protection over the innovation.[191]

If the owners of the complementary assets are in a stronger bargaining position, innovators may want to receive ex ante assurances before undertaking any irreversible investments that would be rendered valueless if it is unable to reach agreement with the owner of the complementary asset.[192] Under such circumstances, the classic solution is for the innovator and the owner to eliminate the risk that the other side may act opportunistically to claim a greater proportion of the surplus either by merging with one another or be entering into a long-term contract.[193] Although such an arrangement would deviate from the mix-and-match world associated with modularity, it would promote entry by reducing the risk faced by innovating firms.[194]

A similar insight emerges from the seminal article by Timothy Bresnahan and Manuel Trajtenberg on GPTs.[195] GPTs create positive vertical externalities for applications sectors that build products based on the GPT.[196] The GPT creator, however, will invest based on a comparison of its private cost of the investment and its private return. The positive externality causes the GPT creator's private benefit to understate the social benefit of further investments, which results in too little investment in GPTs. At the same time, allowing GPT creators to appropriate more of the surplus risks causing too little investment in the applications sector.[197] The public goods aspect of GPTs thus creates a horizontal externality, as different applications-sector players attempt to induce their peers to bear more of the cost of financing improvements in the GPT. The obvious solution to both problems is to permit GPT creators and applications sector developers to merge or enter into some type of cooperative agreement that would allow the GPT creator to realize more the benefits of its investment.[198] Preventing GPTs from coordinating with applications will result in "'too little, too late' innovation" in both sectors.[199]

Both of these lines of research suggest permitting firms in modular spaces to experiment with alternative institutional forms could yield significant benefits. Imposition of a modular structure may prevent these benefits from being realized.

---

[190] *Id.* at 288.

[191] *Id.* at 291–92.

[192] *Id.* at 294.

[193] *Id.* at 290, 293–95.

[194] *Id.* at 302.

[195] Timothy F. Brenahan & M. Trajtenberg, *General Purpose Technologies: "Engines of Growth"?*, 65 J. ECONOMETRICS 83 (1995).

[196] Bresnahan & Trajtenberg, *supra* note 195. at 94.

[197] *Id.*

[198] *Id.* at 96, 99.

[199] *Id.* at 103.

E. Lack of Coordination and Systemic Drift

The decentralized nature of decisionmaking in modular industries can lead to welfare loss. As noted earlier, for distributed production of modules to succeed, each module needs to be able to verify that the other modules are performing as expected. Indeed, Baldwin and Clark regard such testing standards to be an integral part of any modular design.[200] In the context of the Internet, the necessary verification tools simply do not exist. As discussed below, the visible information in the interfaces does not provide enough information for actors to verify whether the other modules are behaving as expected. The absence of any way to verify the conduct of the other modules effectively leaves actors in the position of having to rely on the honor system or some form of informal sanctions.[201]

Furthermore, local optimizing behavior may not maximize the performance of the overall system. Complex systems have the general property that "if each subsystem, regarded separately, is made to operate with maximum efficiency, the system as a whole will not operate with utmost efficiency."[202] Consequently, because modularity focuses each module (and the owner of each module) on optimizing its individual interests, the net result is unlikely to optimize the performance of the system as a whole.[203]

Moreover, given that all companies are drawing off a common pool of surplus that they jointly create, they have the incentive to want others to do more.[204] As was the case with GPTs, this can lead to suboptimal investment in the entire modular ecosystem, as each actor hopes that the others will bear a greater proportion of the cost.[205]

One solution would be to have some actor organize all of the activities of a modular system around the optimal outcome. The problem is that in a decentralized system, individual actors have no incentive to bear the costs of promoting what is best for the system as a whole.[206] Indeed, many would regard centralized control as improper, and any actors who would be disadvantaged by such intervention would vigorously oppose it.

---

[200] *See supra* Part I.3.

[201] Charles L. Jackson, *Wireless Efficiency vs. Net Neutrality*, 63 FED. COMM. L.J. 445, 448–51 (2011).

[202] LARS SKYTTNER, GENERAL SYSTEMS THEORY: IDEAS AND APPLICATIONS 93 (2001).

[203] *See* Randy Bush & David Meyer, *Some Internet Architectural Guidelines and Philosophy* 7–8 (Network Working Group Request for Comments 3439, Dec. 2002) [hereinafter RFC 3439], *available at* http://tools.ietf.org/pdf/rfc3439; Chesbrough & Teece, *supra* note 182, at 66; Jon Crowcroft et al., *Is Layering Harmful?*, IEEE NETWORK, Jan. 1992, at 20, 23–24; John D. Day & Hubert Zimmerman, *The OSI Reference Model*, PROC. IEEE, Dec. 1983, at 1334, 1335–36.

[204] Chesbrough & Teece, *supra* note 182, at 67–69.

[205] *See supra* note _ and accompanying text.

[206] *See* Robert Braden et al., Developing a Next-Generation Internet Architecture 16 (Univ. of S. Cal. Info. Sci. Inst. White Paper July 15, 2000) (introducing the NewArch project funded by the Defense Advanced Research Projects Agency), *available at* http://www.isi.edu/newarch/DOCUMENTS/WhitePaper.pdf ("There is no commercial provider who believes that they [*sic*] hold the responsibility for the Internet architecture.").

The result is that many modular systems will suffer from a lack of coordination. Without such leadership, modular systems often undergo long periods of systemic drift.

* * *

The foregoing analysis makes clear that modularity is not always the optimal architecture. Whether modularity is optimal is contingent on a number of factors. Moreover, any modular scheme necessarily requires the system to give up a degree of functionality. Modularity can also interfere with alternative institutional arrangements that firms can use to address unequal bargaining power or to internalize positive externalities.

The foregoing analysis also underscores that modularity is not a static construct. Indeed, modularity theorists expect the design to change as the architects gain a better understanding of the underlying interdependencies.[207] Technological and economic changes can create pressure for high-tech industries to evolve toward a fundamentally different architecture.[208] Examples include the desktop PC's absorption of functions that used to be provided by standalone peripheral devices—such as hard disks, modems, and WiFi cards[209]—and the advent of last-mile broadband networks—such as DSL and cable modem systems—both of which undercut the rationale for a standalone regional ISP.[210] Changes in technology and end-user demand for network services, however, can cause the nature and relative importance of particular interdependencies to change over time, which in turn creates pressure on the current modular architecture to change. Any such change not only involves transition costs.[211] Any change in the location of transactions inherent in any such remodularization inevitably alters the structure of industries.[212] In fact, the difficulty in analyzing how modularity changes in

---

[207] Langlois, *supra* note 9, at 25; von Hippel, *supra* note 28, at 408, 412.

[208] *See* Baldwin, *supra* note 120, at 180 (noting that "there is no process of technological determinism at work driving the task network toward ever-higher levels of modularity" and that changes in strategies, knowledge, and technologies can cause "task networks to become more integral (i.e., less modular) over time"); Jacobides & Winter, *supra* note 125, at 405 (noting how the emergence of new productive structures and new knowledge bases can cause the pattern of increasing specialization and vertical disintegration to reverse).

[209] *See* Transamerica Computer Co. v. IBM Corp., 698 F.2d 1377, 1382–83 (9th Cir. 1983) (discussing the increasing integration between the CPU and certain peripherals); Cal. Computer Prods., Inc. v. IBM Corp., 613 F.2d 727, 743–44 (9th Cir. 1979) (discussing how the IBM new computer integrated the disk drive control function with the CPU"); ILC Peripherals Leasing Corp. v. IBM Corp., 448 F. Supp. 228, 231–32 (N.D. Cal. 1978) (describing how IBM abandoned removable disk drives in favor of nonremovable drives with the head-disk assembly integrated into the computer itself that provided greater storage capacity), *aff'd sub nom.*, Memorex Corp. v. IBM Corp., 636 F.2d 1188 (9th Cir. 1980); Telex Corp. v. IBM Corp., 367 F. Supp. 258, 342 (N.D. Okla. 1973) (describing how IBM integrated its memories and control units into its CPUs), *rev'd on other grounds*, 510 F.2d 894 (10th Cir. 1975).

[210] Christopher S. Yoo, *Would Mandating Network Neutrality Help or Hurt Broadband Competition?: A Comment on the End-to-End Debate*, 3 J. ON TELECOMM. & HIGH TECH. L. 23, 33–34 (2004).

[211] *Id.* at 25–26.

[212] Baldwin, *supra* note 120, at 187.

industries in the midst of such a period of ferment is what led Baldwin and Clark to divide their work into two volumes.[213]

Attempts to evolve away from an existing modular architecture may thus represent nothing more than the natural response to changes in the economic and technological environment.[214] Indeed, experimentation in new standards and competition between standards are properly regarded as a sign of innovative health.[215]

IV.MODULARITY THEORY AND ARCHITECTURAL CHANGE

The foregoing catalog of the pluses and minuses of modularity provide a basis for assessing the value of any particular modular system at any particular time. That said, this analytical framework views modular architectures as static and fails to explain how modular architectures emerge and, once having done so, when and how they evolve. The fact that the shape of any modular architecture is determined by technological interdependencies on the supply side and by the heterogeneity of demand on the demand side indicates that the optimal architecture should change over time.

The inevitability of architectural change underscores the need for a theory of remodularization[216] as well as heuristics to help recognize when such change is necessary.[217] Fortunately, Baldwin and Clark offer a theory built around six core modular operators that represent the entire set of possible transformations and which provide significant insights into how and when modular architectures can and should change.

A. Modular Operators

Baldwin and Clark identified six operators that can transform a modular architecture: splitting, substitution, augmentation, exclusion, inversion, and porting.[218] The first two (splitting and substitution) can occur to nonmodular systems, while the other four can occur only in modular systems.[219] Moreover, five of the operators represent relatively minor changes to the

---

[213] Carliss Y. Baldwin & Kim B. Clark, Roadmap for Design Rules 8 (2006), *available at* http://www.people.hbs.edu/cbaldwin/DR2/BaldwinDRRoadmap.pdf.

[214] Langlois, *supra* note 9, at 25–26.

[215] *See* Shane Greenstein, *Glimmers and Signs of Innovative Health in the Commercial Internet*, 8 J. ON TELECOMM. & HIGH TECH. L. 25, 42–55 (2010) (discussing how economic experimentation and vigorous standards competition are signs of innovative health).

[216] Ethiraj & Levinthal, *supra* note 9, at 172 (noting that "we need to model the evolution of modularity itself").

[217] Stefano Brusoni & Andrea Prencipe, *Dynamics of Industry and Innovation: Organizations, Networks and Systems* 5 (2005) (unpublished manuscript), *available at* http://www.druid.dk/uploads/tx_picturedb/ds2005-1543.pdf.

[218] BALDWIN & CLARK, *supra* note 4, at 123, 132–42, 228.

[219] *Id.* at 123, 136.

architecture. Only inversion requires a fundamental reordering of the relevant tasks.

1. Splitting and Substitution

Consider *splitting*, which is the division of one module into two.[220] As noted earlier, splitting is not unique to modular architectures; indeed, it is how an interconnected architecture (which is essentially consists of one module) becomes modular in the first place.[221] These modules can in turn split again. *Substitution* is the replacement of one module design for another.[222] Baldwin and Clark see splitting and substitution as part of a cycle, which begins with splitting of two modules, followed by substitution of new technologies for particular modules, followed again by splitting as the benefits from further substitution begin to abate.[223]

Splitting, however, imposes only a relatively small change to a modular architecture. Because it by definition only affects the internal design structure of what had originally been one module in the system, it does not disturb the relationship with other modules or any of the information visible to other modules.

Substitution has similarly does not disrupt the basic architecture at all. Indeed, the whole point of modularity is to swap out existing modules without affecting the internal workings of the other modules. Although substitution does improve the system, by operating on only the hidden information, it does so in a way that does not disturb the basic architecture of the overall system.

2. Exclusion and Augmentation

Two other operators, exclusion and augmentation, together form part of a strategy that makes sense when users do not need all of the modules that are available or when designing the full set of possible modules is too costly at the time of launch.[224] When that is the case, the designer can reduce development costs by designing a narrow architecture (exclusion), and then expanding it after it is validated by customers (augmentation).[225] This permits the designer to conserve resources by deferring investments in additional functionality until the system has been validated by the marketplace. If the system is successful, the designers can commit additional resources to developing other parts of the system with greater impunity.[226] .

---

[220] *Id.* at 123.
[221] *Id.*
[222] *Id.* at 281.
[223] *Id.* at 281, 323.
[224] *Id.* at 136, 308, 310.
[225] *Id.* at 308.
[226] *Id.* at 136–37, 313–14.

The designers of the initial, minimal system must design it with future augmentation in mind. The key is design rules with clean interfaces that provide a large "shadow set" of potential new modules that can be created through augmentation.[227] If so, those who wish to augment the architecture need only understand how to fit the new module within the shadow design rules.[228]

Designing an architecture to support exclusion and augmentation in effect requires the virtual design of a more complete architecture that organizes the relationship between current and future models. The difference is primarily one of timing. In the case of complete architecture, all of the costs are committed up front. In the case of an architecture narrowed by exclusion in anticipation of future expansion through augmentation, many of these costs can be postponed. In this sense, the process of creating a complete architecture and a minimal architecture susceptible to later augmentation both require conceptualizing and enforcing a coherent architecture up front. If the initial minimal architecture is designed properly, subsequent augmentation does not cause any disruption to the design. In fact, the possibility of augmentation opens the door to coinvestment by other firms.[229] All these third parties need to do is look to the same set of visible information envisioned by the architects.[230]

3. Porting and Inversion

The final two modular operators effect more substantial architectural changes. Porting occurs when a module is made available to another modular system, such as when a printer made for one operating system is adapted to work with another. The process requires creating a shell around the module with a translator capable of conforming to the design rules of both systems.[231] If done properly, porting has no impact on the old architecture. Indeed, architects and designers of modules of the other system may not know that porting has taken place.[232]

The most intersecting modular operator for purposes of his Article is inversion. Inversion "tak[es] previously hidden information and mov[es] it up the design hierarchy so that it is visible to a group of modules."[233] For example, printing subroutines used to be buried inside programs, with each program including its own unique implementation tailored to its program that

---

[227] *Id.* at 321.

[228] *Id.* at 306.

[229] *Id.* at 321–22.

[230] *Id.* at 322.

[231] *Id.* at 140.

[232] *Id.* at 140. Porting can cause the shell to become a new source of visible information in its own right apart from the system. *Id.* at 141. As such, it can create a new apex within the design hierarchy. *Id.* at 339.

[233] *Id.* at 138.

was part of that program's hidden information and thus locked inside it. Over time, the fact that many programs use printing made it logical to move printing subroutines to a higher level of the design hierarchy to a place where it is visible to all other programs.[234] "[T]hrough inversion, what was hidden becomes visible."[235]

What began as a module inside another module is treated as a common solution and imposed on other programs.[236] It occurs when the value of the new architecture exceeds the sum of the cost of designing the new architecture, the cost of redesigning the hidden modules that previously relied on the new inverted module it they used to be locked inside a module, and the lost net option value that would have accrued if experiments under the old architecture had been permitted to continue.[237]

By its nature, inversion represents a significant change that that requires fundamental changes in the task structure.[238] As such, they are often highly controversial.[239] They take away designers' prerogatives over previously hidden information that was previously locked within a module, and thus subject to experimentation, and freezes it into the visible information that is part of the new design.[240] Indeed, inversion terminates all further experiments with respect to the previously hidden information that inversion makes visible.[241] Moreover, they require that adjacent modules now interact with a different set of visible information. They also require designers to adapt to a new set of rules and to develop a new set of skills.[242]

That said, inversions occupy a natural place in the life cycle of a modular regime. The creation of a new modular architecture will prompt a rash of experimentation. The returns yielded by these experiments will diminish over time. Eventually, the benefits from further experimentation no longer justify the cost, at which point pressure will mount for a new set of design rules through inversion.[243]

B. Obstacles to Significant Architectural Change

The conventional wisdom is that modern technologies are undergoing constant and rapid change, sometimes reflected in the term "Internet time."[244] Closer inspection reveals the phrase only to be half true. While modularity

---

[234] *Id.*
[235] *Id.* at 139.
[236] *Id.* at 326.
[237] *Id.* at 327–28.
[238] *Id.* at 127.
[239] *Id.* at 327.
[240] *Id.* at 327.
[241] *Id.* at 326.
[242] *Id.* at 327.
[243] *Id.* at 323.
[244] *See, e.g.*, MICHAEL A. CUSUMANO & DAVID B. YOFFIE, COMPETING ON INTERNET TIME: LESSONS FROM NETSCAPE AND ITS BATTLE WITH MICROSOFT 3 (1998).

permits rapid innovation with respect to individual modules, changes in the architecture that structures the interactions among those modules occurs glacially slowly.[245] Indeed, this stability is critical to the success of modular systems, since it provides those who are testing modular innovations with the fixed reference points around which to base their products as well as the expectation of being able to capitalize on any improvements that prove successful. And architectures that have been established are notoriously hard to change.[246]

This means that modular architectures are effective at promoting innovations that are consistent with the existing stack, such as substitution and to a lesser extent splitting, exclusion, and augmentation. At the same time, they tend to hinder innovations that require a reorganization of the tasks performed by the stack through an operator such as inversion.

One part of this slow rate of architectural change stems from coordination problems. The literature on network economic effects has long recognized how the presence of an installed base can delay the adoption of a new technology even when that technology is superior.[247] The problems are likely to be more severe in a modular architecture, since changes to the architecture would require the coordination with the modules both above and below the modules being inverted.[248] As a result, some scholars have raised the concern that modular architectures may be unduly resistant to this type of change.[249] This resistance may lead firms to abandon modular architectures in favor of more vertically integrated market structures.[250]

The literature on innovation has long recognized how each architecture creates its own technological paradigm that identifies the problems worth solving and the solutions that are the most promising.[251] As noted earlier, one key aspect of modularity is that it allows organizations to focus their attention on individual modules and to disregard the system as a whole. Ironically, the same quality that reduces complexity makes organizations less attentive to potential architectural changes. In addition, a modular architecture establishes a technical agenda for a product's development that directs research along innovation avenues consistent with the hierarchy.[252] Each modular architecture thus creates its own evolutionary trajectory.[253]

---

[245] Blanchette, *supra* note 10, at 1054.

[246] BALDWIN & CLARK, *supra* note 4, at 89.

[247] *See, e.g.*, Joseph Farrell & Garth Saloner, *Installed Base and Compatibility: Innovation, Product Preannouncements, and Predation*, 76 AM. ECON. REV. 940, 942 (1986).

[248] Blanchette, *supra* note 10, at 1054.

[249] Chesbrough & Teece, *supra* note 182, at 66, 68; Jacobides & Winter, *supra* note 125, at 404; Langlois, *supra* note 9, at 26.

[250] Langlois & Robertson, *supra* note 139, at 302.

[251] Yoo, *supra* note 179, at 656.

[252] Giovanni Dosi, *Technological Paradigms and Technological Trajectories*, 11 RESEARCH POL'Y 147, 152 (1982); *see also* Devendra Sahal, *Technological Guideposts and Innovation Avenues*, 14 RESEARCH POL'Y 61, 71, 78–79 (1985) (describing how "specific information avenues" lead to particular products).

[253] BALDWIN & CLARK, *supra* note 4, at 328.

These paradigms, moreover, become ingrained in the communication channels and information filters that organizations use to manage information, which tends to further reinforce the status quo.[254] Admittedly, these problems are worse for nonmodular systems,[255] but they exist for modular systems as well.

These considerations suggest that casting modularity as a tradeoff between long-term evolvability and short-run inefficiency paints an incomplete picture.[256] Instead, the differences between modular and architectural innovation suggest that there are evolvability considerations on both sides of the equation.

Moreover, the nature of the two types of innovation is also likely to be different. While innovation with respect to individual modules is likely to be very rapid, it is also likely to be incremental and consider only possibilities permitted by the existing architecture. The best it can achieve is to reflect the local maximum permitted by the existing architecture.[257]

Architectural innovation has the opposite characteristics: while very slow, the freedom to consider all possibilities, including those inconsistent with the current architecture, makes it more likely to make a sharply discontinuous jump to a completely different local optimum.[258] If the value of this other optimum is sufficiently large, architectural innovation can yield benefits that more than compensate for the speed advantage enjoyed by modular innovation.[259] Indeed, in contrast to the conventional wisdom, which holds that rapid technological change favors modularity, the greater potential of nonmodular architectures to take advantage of major technological developments suggests volatility actually favors the latter.[260]

The modularity literature also helps provide intuitions about whether change is likely to occur too quickly or too slowly. Several considerations suggest the latter. For example, the sunk costs associated with remodularization create the risk of path dependence that causes the system to become locked into an inferior modularization.[261] That said, as noted earlier, the value of modularity depends on having a significant degree of stability.

---

[254] Philip Anderson & Michael L. Tushman, *Technological Discontinuities and Dominant Designs: A Cyclical Model of Technological Change*, 35 ADMIN. SCI. Q. 604, 618 (1990); Stefano Brusoni & Andrea Prencipe, *Patterns of Modularization: The Dynamics of Product Architecture in Complex Systems*, 8 EUR. MGMT. REV. 67, 67, 68 (2011); Henderson & Clark, *supra* note 85, at 15–19.

[255] BALDWIN & CLARK, *supra* note 4, at 57–59.

[256] VAN SCHEWICK, *supra* note 2, at 169.

[257] Stefano Brusoni et al., *The Value and Costs of Modularity: A Problem-Solving Perspective*, 4 EUR. MGMT. REV. 121, 122, 123, 128, 130 (2007).

[258] *Id.*

[259] *Id.* at 128.

[260] *Id.*

[261] Langlois, *supra* note 9, at 26; Schaefer, *supra* note 33, at 325.

Moreover, architectural change is very expensive.[262] Circumstances thus may exist when the magnitude of the benefits do not justify the cost.[263]

To say that the architecture should change only rarely is not to say that it should never do so. The fact that the optimal modular architecture is largely a function of technical considerations and demand characteristics implies that optimal architecture will be dynamic in the long run[264] and that modular architectures thus may need to undergo remodularization from time to time.[265]

Modularity theory thus provides a theory of architectural change, along with reasons to suspect that such change may be too long in coming. The power of this analytical framework will be explored through the case studies presented in the next Part.

V. POLICY APPLICATIONS

Modularity theory provides a number of insights into many recent and current policy issues. These include unbundling under the Telecommunications Act of 1996, network neutrality, calls for opening Application Programming Interfaces (APIs) for platforms such as the iPhone and Twitter, and proposals for clean-slate redesigns of the network architecture.

A. Unbundling of Local Telephone Networks

Starting first with an example from the traditional telephone network, perhaps the key regulatory innovation of the past several decades is the requirement that incumbent provider provide unbundled access to all of the elements of their networks. The FCC articulated the seminal unbundling requirement as part of its *Computer Inquiries*,[266] which arose when local telephone companies began to move beyond offering traditional voice communications (which the FCC called *basic services*) to offer voicemail, caller ID, and Internet access, and other new advanced services that combined the transport of communications with computer processing, storage, or interaction with stored information (which the FCC called *enhanced services*).[267]

---

[262] BALDWIN & CLARK, *supra* note 4, at 89; Langlois, *supra* note 9, at 23; von Hippel, *supra* note 28, at 409.

[263] Timothy F. Bresnahan, *New Modes of Competition: Implications for the Future Structure of the Computer Industry*, *in* COMPETITION, INNOVATION AND THE MICROSOFT MONOPOLY: ANTITRUST IN THE DIGITAL MARKETPLACE 155, 161 (Jeffrey A. Eisenach & Thomas M. Lenard eds., 1999).

[264] BALDWIN & CLARK, *supra* note 4, at 156, 180.

[265] Langlois, *supra* note 9, at 25.

[266] For an overview of the *Computer Inquiries*, see Daniel F. Spulber & Christopher S. Yoo, *Access to Networks: Economic and Constitutional Connections*, 88 CORNELL L. REV. 885, 1005–09 (2003).

[267] 47 C.F.R. § 64.702(a).

The second *Computer Inquiry* attempted to prevent large local telephone companies from favoring their own enhanced service offerings by requiring them to provide enhanced services through a separate subsidiary and to offer all enhanced service providers the same access to their transport infrastructure.[268] The FCC became concerned that mandating such a high degree of vertical disintegration was preventing new services from appearing.[269] As a result, the *Third Computer Inquiry* created a regime where local telephone companies could avoid the separate subsidiary requirement if they provided unbundled access to their networks on an element-by-element basis.[270]

The Telecommunications Act of 1996 extended unbundling to every incumbent local telephone company.[271] Requiring incumbents to provide access to all of their network elements on an unbundled basis was the centerpiece of Congress's landmark effort to stimulate competition in local telephone service. The unbundling requirement imposed by the 1996 Act did include one key limitation. It required the FCC to determine whether access to proprietary network elements was "necessary" and whether the failure to provide access to such elements would "impair" the requesting carrier's ability to provide service.[272] The statute also required incumbents to provide unbundled access at any technically feasible point.[273]

The hope was that new entrants would use unbundling to combine their own facilities with facilities leased from the incumbents (those elements that still were natural monopolies) to provide service. In the words of Nicholas Economides, "[t]he basic logic behind the Telecommunications Act was to break the network into components and let everyone compete in every part."[274] In so doing, policymakers followed the same approach that they adopted when competition became possible with respect to telephone handsets (known as customer premises equipment or CPE), long distance services, and new information services that combined transmission with data

---

[268] Amendment of Section 64.702 of the Commission's Rules and Regulation's (Second Computer Inquiry), 77 F.C.C.2d 384, 217–39 ¶¶ 201–260 (1980), *aff'd sub nom.* Computer & Commc'ns Indus. Ass'n v. FCC, 693 F.2d 198 (D.C. Cir. 1982). This recapitulated the structural separation requirement imposed by the first *Computer Inquiry*, which focused on the pervious distinction between *communications services* and *data processing services*, which ultimately proved to be untenable. Regulatory and Policy Problems Presented by Interdependence of Computer and Communication Services and Facilities, Final Decision and Order, 28 F.C.C.2d 267, 270–74 ¶¶ 11–22 (1971), *aff'd sub nom.* GTE Serv. Corp. v. FCC, 474 F.2d 724 (2d Cir. 1973).

[269] Amendment of Sections 64.702 of Commission's Rules and Regulations (Third Computer Inquiry), Report and Order, 104 F.C.C.2d 958, 1002–12 ¶¶ 78–99 (1986).

[270] This FCC called this regime Open Network Architecture (ONA). *Id.* at 1064–66 ¶¶ 214–217. Until it was fully implemented, the FCC mandated a more limited, transitional regime known ad Comparably Efficient Interconnection (CEI). *Id.* at 1035–43 ¶¶ 147–166.

[271] 47 U.S.C. § 251(c)(3).

[272] *Id.* § 251(d)(2)(A)–(B).

[273] 47 U.S.C. § 251(c)(3).

[274] Nicholas Economides, *Vertical Leverage and the Sacrifice Principle: Why the Supreme Court Got* Trinko *Wrong*, 61 N.Y.U. ANN. SURV. AM. L. 379, 390 (2005).

processing: isolate those portions that continued to exhibit natural monopoly characteristics and mandate open access to those portions.[275]

Unbundling is now widely regarded as a failure.[276] As of today, meaningful facilities-based competition for wireline telephone services has yet to emerge.[277] Although no doubt other reasons exist for this outcome, modularity theory provides an answer that has not been appreciated.

Consider the statutory requirement that incumbents provide unbundled access at any technically feasible point.[278] This requirement presumes that it is possible to create a modular element at any location in the network. Modularity theory suggests why this approach was problematic. Modular interfaces cannot be established anywhere; they can be created only at thin crossing points where the number of interdependencies is relatively low.

Modularity theory also provides insights into problems with adopting an exclusively economically oriented approach for determining which elements would be modularized. After three failed attempts,[279] the FCC was finally able to implement unbundling in a manner that withstood judicial review.[280] Following the D.C. Circuit's instructions that the rules be "linked (in some degree) to natural monopoly,"[281] the agency's rules adopted a decidedly economic approach. A proprietary network element was "necessary" only if lack of access would preclude the requesting carrier from providing service,.[282] With respect to nonproprietary network elements, requesting carriers would be "impaired" if lack of access represented a barrier to entry likely to make it uneconomic for a reasonably efficient competitor to enter.[283] In both cases, regulators had to take into account whether the requesting

---

[275] Christopher S. Yoo, *Deregulation vs. Reregulation: The Clash of Conflicting Paradigms*, 36 J. CORP. L. 847, 850 (2011).

[276] *See, e.g.*, Susan P. Crawford, *Network Rules*, LAW & CONTEMP. PROBS., Spring 2007, at 51, 87 n.16 ("Mandated unbundling under the Act is widely viewed to have been a failure."); Richard A. Epstein, *Takings, Commons, and Associations: Why the Telecommunications Act of 1996 Misfired*, 22 YALE J. ON REG. 315, 315–16 (2005) ("There is widespread agreement today on all sides of the telecommunications was that something is deeply flawed with the design or implementation (or both) of the Telecommunications Act of 1996"); Mark A. Lemley & Philip J. Weiser, *Should Property or Liability Rules Govern Information?*, 85 TEX. L. REV. 783, 812 (2007) ("[T]he unbundling regime of the 1996 Act represents, on almost all accounts, a policy failure."); Kevin Werbach, *Only Connect*, 22 BERKELEY TECH. L.J. 1233, 1237 (2007) ("The current legal framework, embodied in the Telecommunications Act of 1996 (1996 Act), is widely regarded as a colossal failure.").

[277] *See* Robert W. Crandall & Leonard Waverman, *The Failure of Competitive Entry into Fixed-Line Telecommunications: Who Is at Fault?*, 2 J. COMPETITION L. & ECON. 113, 114–25 (2006). Interestingly, competition for local telephone services has arisen intermodally, with wireless telephony now serving as an effective substitute for traditional wireline services. Interestingly, regulators were initially reluctant to regard cellular telephones as substitutes for wireline service. Daniel F. Spulber & Christopher S. Yoo, *Toward a Unified Theory of Access to Local Telephone Networks*, 61 FED. COMM. L.J. 43, 86–87 (2008).

[278] 47 U.S.C. § 251(c)(3).

[279] *See* AT&T Corp. v. Iowa Utils. Bd., 525 U.S. 366, 387–92 (1999); U.S. Telecom Ass'n v. FCC (*USTA I*), 290 F.3d 415, 422–28 (D.C. Cir. 2002); U.S. Telecom Ass'n v. FCC, 359 F.3d 554, 571–73 (D.C. Cir. 2004).

[280] *See* Covad Commc'ns Co. v. FCC, 450 F.3d 528, 538–43 (D.C. Cir. 2006).

[281] *USTA 1*, 290 F.3d at 427.

[282] 47 C.F.R. § 51.317(a)(1).

[283] *Id.* § 51.317(b).

carrier could build the requested network elements itself or contract for the same services with another provider.[284] The prices for unbundled access would be based on forward-looking economic cost.[285]

The FCC's approach focuses on a network element's cost characteristics without taking into account the interdependencies between that element and the other portions of the network. If the element is tightly integrated with other components, any attempt to mandate access would be expected to fail regardless of whether it is a natural monopoly or is unavailable through other means.

Such interdependencies are more likely to arise under unbundling than under other forms of access because, as the Supreme Court noted in *Trinko,* unbundling involves network elements "deep within the bowels" of a local telephone network that can be made available only if "[n]ew systems [are] designed and implemented simply to make that access possible."[286]

Thus unlike mandated access to CPE and long distance, which involved a market boundary that was "simple, easy to monitor and require[d] little information," the higher level of interdependencies associated with unbundling made it less likely that it would prove successful.[287] In short unbundling attempted to introduce transactions into what is technologically a transaction-free zone.[288]

Modularity theory also sheds new light on some of the disputes that arose during the debates over unbundling. For example, the local telephone companies asked the FCC to interpret the unbundling requirement imposed by the third *Computer Inquiry* so as not to mandate "fundamental unbundling" on an element-by-element basis and instead only mandate access to larger combinations known as basic service elements[289] only to see that effort overturned on judicial review.[290] To the extent that those larger combinations reflected interdependencies, modularity theory suggests that requiring that bundles of network elements be leased as a unit would have been well advised.

Similarly, when implementing the 1996 Act, the FCC issued a rule that was eventually upheld by the Supreme Court preventing local telephone companies from separating network elements that had already been combined together before leasing them to competitors.[291] Again, modularity

---

[284] *Id.* § 51.317(a)(1) & (b).

[285] *Id.* § 51.505.

[286] Verizon Commc'ns Inc. v. Law Offices of Curtis V. Trinko, LLP, 540 U.S. 398, 410 (2004).

[287] Gerald R. Faulhaber, *Policy-Induced Competition: The Telecommunications Experiments*, 15 INFO. ECON. & POL'Y 73, 77, 78, 83, 86, 91 (2003).

[288] *See supra* note 122 and accompanying text.

[289] Filing and Review of Open Network Architecture Plans, Memorandum Opinion and Order, 4 FCC Rcd. 1, 13–14 ¶¶ 5–8, 41–42 ¶ 69 (1988).

[290] California v. FCC, 39 F.3d 919, 925–30 (9th Cir. 1994).

[291] Implementation of the Local Competition Provisions in the Telecommunications Act of 1996, First Report and Order, 11 FCC Rcd. 15499, 15647 ¶ 293 (2006) (codified at 47 C.F.R. § 51.315(b)), *aff'd sub nom.* AT&T Co. v. Iowa Utils. Bd., 525 U.S. 366, 393–95 (1999).

theory underscores that certain elements form integrated units that should be provided together.

B. The Internet Protocol

For the past several years, debates over Internet policy have been dominated by two big issues: network neutrality and the transition from IPv4 to IPv6. The Internet is designed around modular architecture embodied in the layered suite of protocols known as the Transmission Control Protocol/Internet Protocol (TCP/IP).[292] In many ways, the Internet represents a classic example of how to use modularity to simplify a complex system. Its central focus was to allow a heterogeneous group of hosts to interconnect through an equally heterogeneous set of transmission technologies.[293] For example, each of the first four ARPANET sites (UCLA, UC Santa Barbara, SRI, and the University of Utah) each used a different type of computer as its host (CDS Σ-7, IBM 360/75, XDS 940, DEC PDP 10), which operated fundamentally incompatible design principles.[294] In addition, the original experiment that validated the Internet as a principle successfully integrated a wireline telephone network, a satellite-based network, and a mobile wireless network to send a single transmission.[295]

The lynchpin of the TCP/IP suite is the network-layer protocol known as the Internet Protocol (IP).[296] IP is a module that mediates both the diversity of applications and computers being run by end users (known as hosts) on the one hand and the diversity of transmission technologies on the other. Because of this, the Internet's architecture is often portrayed as an hourglass, with a wide variety of host-based protocols at the top of the hourglass and a wide variety of networking and transmission technologies at the bottom.[297] IP is "[t]he hourglass's narrow waist," which "represents a minimal and carefully chosen set of local capabilities that allows both higher-level

---

[292] *See, e.g.*, Crowcroft et al., *supra* note 203, at 23 ("Layering is about modularizing the functions performed on data during its transfer from one machine to another, so that complexity of the transfer can be managed."); Douglas C. Sicker & Lisa Blumensaadt, *Misunderstanding the Layered Model(s)*, 4 J. on Telecomm. & High Tech. L. 299, 305 (2006) ("Each layer provides a well-defined set of services to the layers above it and depends on lower layers for its own operation, thus creating modularity.").

[293] Janet Abbate, Inventing the Internet 51–52, 131–32 (1999).

[294] C. Stephen Carr et al., *HOST-HOST Communication Protocol in the ARPA Network*, 36 AFIPS Conf. Proc. 589, 590 (1970).

[295] Packet Radio System Development: Quarterly Management Report No. 15, at 5–7 (Stanford Research Institute Project 2325-NS, Nov. 12, 1976).

[296] Barry M. Leiner et al., *The DARPA Internet Protocol Suite*, IEEE Comm., Mar. 1985, at 29, 31; *see also* Andrew S. Tanenbaum, Computer Networks 432 (4th ed. 2003) (calling IP ”[t]he glue that holds the whole Internet together"); Brian E. Carpenter, *Architectural Principles of the Internet* 2 (Network Working Group Request for Comments 1958, June 1996), *available at* http://tools.ietf.org/pdf/rfc1958 (noting that "The key to global connectivity is the inter-networking layer").

[297] *See, e.g.*, Computer Sci. & Telecomm. Bd., Nat'l Research Council, The Internet's Coming of Age 36–38, 126–29 (2001); Steve Deering, *Watching the Waist of the Protocol Hourglass*, Proc. 51st Internet Eng'g Task Force 2 (2001), http://www.ietf.org/proceedings/51/slides/plenary-1/index.html.

applications and lower-level communication technologies to coexist, share capabilities, and evolve rapidly."[298] In other words, IP represents a modular interface that reflects careful decisions about which interdependencies are allowed to be reflected in interactions between hosts and networks. It is the limits on the visible information built into this interface that allows applications to ignore the details of the underlying transmission technology and vice versa.

TCP/IP yields all of the advantages associated with modularity. It reduces the complexity of the system by rendering subsystems as independent as possible from one another. It promotes flexibility and allows for simultaneous parallel experiments by a diversity of organizations. It enables third-party development of new applications without permission from any other actors. However, even the architects of the TCP/IP protocol recognized that it is "a mixed blessing."[299] In fact, it is subject to a number of the drawbacks associated with all modular architectures.

1. Restrictions on Functionality in IPv4

Like all modular architectures, TCP/IP employs information hiding to limit and structure the interactions between modules. In so doing, it essentially embodies a precommitment about which types of information are important enough to pass between the modules. As a result, the limited nature of the interfaces inevitably "leads to a lack of understanding as to what one layer wishes to obtain from another.[300]

The details of why this is so is best understood by examining the information contained in the IP header, which constitutes the critical interface in this particular modularization. The IP version 4 header consists of a fixed part consisting of five 32-bit words and an optional part that can consist of up to ten additional 32-bit words.

---

[298] PETERSON & DAVIE, *supra* note 3, at 30.

[299] RFC 817, *supra* note 158, at 24.

[300] *Id.* at 12; *accord* RFC 3439, *supra* note 203, at 7–8 (noting that layering "hide[s] vital information that lower layers may need to optimize their performance" and requires "that the optimization of each layer . . . be done separately," which can "conflict with efficient implementation of data manipulation functions"); Crowcroft et al., *supra* note 203, at 23 ("[T]he flip side to modularization and data-hiding is that tuning the efficiency of the data path for transfer of data becomes difficult, as important details such as buffer sizes are hidden from each layer. Vertical partitioning emphasizes the discontinuities in the data path, which then obstruct the application from receiving the quality of service it requires.").

Figure 4: The IPv4 Header

◄——32 bits——►

<table>
<tr><td colspan="2">Version</td><td colspan="2">IHL</td><td colspan="3">Type of Service</td><td colspan="1"></td><td colspan="8">Total Length</td></tr>
<tr><td colspan="8">Identification</td><td colspan="2">DF/MF</td><td colspan="6">Fragment Offset</td></tr>
<tr><td colspan="4">Time to Live</td><td colspan="4">Protocol</td><td colspan="8">Header Checksum</td></tr>
<tr><td colspan="16">Source Address</td></tr>
<tr><td colspan="16">Destination Address</td></tr>
<tr><td colspan="16">Options (if any)</td></tr>
</table>

Source: Info. Sci. Inst., *Internet Protocol: DARPA Internet Program Protocol Specification* 11 (Request for Comments 791, 1981), *available at* http://tools.ietf.org/pdf/rfc791.

In the first word, *Version* specifies whether the packet is an IPv4 packet or some other version. Because the length of the header varies, *IHL* stands for Internet Header Length and indicates when the header ends and the data begins, which is necessary because the inclusion of options at the end of the header can cause the size of the header to vary. *Type of Service* allows senders to assign different levels of propriety to particularly packets. *Total Length* indicates the size of the entire packet, including both the data and the header. The second word consists of fields—*Identification*, *DF* (for Don't Fragment), *MF* (for More Fragments), and *Fragmentation Offset*—that are designed to allow IP to divide a large packet into smaller fragments. The third word consists of three fields. *Time to Live* is a counter that is decremented each hop the packet traverses until the counter reaches zero in order to ensure that packets without a proper destination address do not wander around the Internet forever. *Protocol* specifies whether the packet is associated with TCP, UDP, or some other transport protocol. *Header Checksum* verifies the accuracy of the header. The fourth and fifth words consist of the *Source Address* and the *Destination Address*. The *Options* fields are rarely used except for debugging purposes.[301]

The visible information contained in the IP header inevitably constrains the type of information that can pass between the hosts connected to the edge of the network and the network itself. Because the information in the IP header is basically limited to the type of service, transport protocol, source address, and destination address, it cannot support functionality that depends

---

[301] TANENBAUM, *supra* note 296, at 433–36.

on any other information. This has particular import with respect to three functions: congestion management, reliability, and security.

a. Congestion Management

Congestion management represents one of the central issues in the network neutrality debate. Indeed, the FCC listed congestion management as an example of "reasonable network management" exempted from the nondiscrimination requirements established by its Open Internet Order.[302] The problem is that the IP header does not include a field through which hosts and the network can exchange information about congestion. In other words, congestion is one of the interdependencies that is part of the hidden information of the architecture and thus is opaque to all of the modules.

Congestion occurs when multiple hosts attempt to use the network at the same time. Individual hosts, however, generally only possess information about their activities and typically lack information about the behavior of other hosts. The nodes in the core of the network are in a far better position to see the flows being generated by multiple users.[303] The fact that congestion management requires information that is available only in the core similarly favors a core-based solution. At the same time, when congestion arises, the proper response is for hosts to cut their sending rates in half.[304]

Ideally, the IP header would contain a bit that networks could use to signal hosts that the network is congested. Indeed, Raj Jain proposed just such a solution during the Internet's early days.[305] A network-based solution would have required upgrading the network's core routers, which would have taken a long time and would have been prohibitively expensive.[306] Instead, Van Jacobson and Mike Karels devised a solution based on the host-based approach to reliability embedded in TCP/IP. Under this approach, receiving hosts are supposed to acknowledge every packet they receive. If the sending host does not receive an acknowledgement within the expected amount of time, it resends the packet.

---

[302] Open Internet Order, *supra* note 1, at 17951–52 ¶¶ 81–82, 17955–56 ¶¶ 91–92.

[303] Sally Floyd & Van Jacobson, *Random Early Detection Gateways for Congestion Avoidance*, 1 IEEE/ACM TRANSACTIONS ON NETWORKING 397, 397 (1993).

[304] Van Jacobson, *Congestion Avoidance and Control*, ACM SIGCOMM COMPUTER COMM. REV., Aug. 1988, at 314, 318.

[305] Raj Jain et al., *Congestion Avoidance in Computer Networks with a Connectionless Network Layer* 1, 6–7 (Digital Equip. Corp. Technical Report DEC-TR-506, Aug. 1987), *available at* http://www1.cse.wustl.edu/~jain/papers/ftp/cr5.pdf.

[306] RICHARD BENNETT, DESIGNED FOR CHANGE: END-TO-END ARGUMENTS, INTERNET INNOVATION, AND THE NET NEUTRALITY DEBATE 16 (2009) (noting that the network adopted Jacobson's solution because changing a few lines of code was cheaper than upgrading hundreds of routers), *available at* http://itif.org/files/2009-designed-for-change.pdf.

Jacobson and Karels realized that networks typically drop packets for only two reasons: either the packet became corrupted or the packet encountered a congested buffer that was full.[307] Because wireline networks rarely corrupt packets, hosts could take the failure to receive an acknowledgement as a de facto signal that the network was congested and as an indication that they needed to reduce traffic.[308] Unlike a network-based solution, Jacobson and Karels's' solution only required changing a few lines of code in every host.[309]

Jacobson and Karels never intended their solution to be a permanent one.[310] As an initial matter, because it depended on an inference from the lack of an acknowledgement, it worked only for traffic based on transport protocols that use acknowledgments, such as TCP.[311] Modern applications such as streaming video and VoIP have placed increasing emphasis on protocols that do not use acknowledgements, such as UDP.[312]

Just as problematic is the growing importance of wireless broadband technologies. Unlike wireline networks, wireless networks often drop packets for reasons other than congestion, such as when atmospheric conditions or reflections create a dead spot that limits the amount of bandwidth available or when a bad handoff between cell sites causes a transmission to become dropped.[313] Thus, the advent of wireless broadband further undercuts the inference-based solution used to work around the absence of a field in the IP header through which networks can notify hosts directly about the presence of congestion. A later modification known as Explicit Congestion Notification (ECN) has made some information about congestion visible in the IP header,[314] and a current initiative known as Congestion Exposure (ConEx) is underway to insert additional information into the IP header.[315] Neither is as yet mandatory in Internet implementations, and neither is widely deployed.

---

[307] Jacobson, *supra* note 304, at 319.

[308] *Id.*

[309] *Id.* at 314–15, 321.

[310] *Id.* at 322.

[311] Mark Handley, *Why the Internet Only Just Works*, 24 BT TECH. J. 119, 120 (2006).

[312] JAMES F. KUROSE & KEITH W. ROSS, COMPUTER NETWORKING: A TOP-DOWN APPROACH 213 (5th ed. 2010) (calling UDP's lack of congestion control "controversial" because UDP's failure to reduce its rate in response to packet loss can cause UDP traffic to "crowd[] out . . . TCP sessions"); *id.* at 293 ("From the perspective of TCP, the multimedia applications running over UDP are not being fair—they do not cooperate with the other connections nor adjust their transmission rates appropriately. Because TCP congestion control will decrease its transmission rate in the face of increasing congestion (loss), while UDP sources need not, it is possible for UDP sources to crowd out TCP traffic.").

[313] Christopher S. Yoo, *The Changing Patterns of Internet Usage*, 63 FED. COMM. L.J. 67, 79 (2010).

[314] K.K. Ramakrishnan, The Addition of Explicit Congestion Notification (ECN) to IP (Network Working Group Request for Comments 3168, Sept. 2001), *available at* http://tools.ietf.org/pdf/rfc3168.

[315] IETF, Congestion Exposure (ConEx): Description of Working Group, http://datatracker.ietf.org/wg/conex/charter/.

b. Reliability

Changes in the ways end users are using the Internet, both in terms of technologies and in terms of applications, can cause the optimal configuration of the layered stack to change. Currently, responsibility for guaranteeing reliability rests with the hosts. While this approach made sense when networks were based around wireline telephone technologies, it makes less sense when wireless technologies are involved. Unlike wireline networks, which rarely drop packets for reasons other than congestion, wireless networks suffer from much higher loss rates caused by the sensitivity of spectrum-based transmission to local conditions.[316]

For this reason, early wireless networks like the San Francisco Bay Packet Radio Network (PRNET) employed a network-based reliability system known as forward-error correction.[317] This difference between wireline and wireless technologies also explains why modern wireless broadband networks are increasingly deploying network-based reliability systems, such as Automatic Repeat reQuest (ARQ), that detect transmission errors and retransmit the missing data without waiting for the host-based retransmission timer to expire and without consuming the additional network resources needed to retrieve the packet all the way from the host.[318] Other techniques that allow routers in the core to participate in the transport layer exist as well.[319]

These experiments with shifting responsibility for reliability from the hosts operating at the edge of the network into the network itself are part of the burgeoning literature on cross-layer design in wireless networks.[320] This shift represents a clear deviation from the allocation of functions embodied in the TCP/IP reference model. That said, the pragmatic approach of network engineering counsels against basing objections to incorporating cross-layer design into the network on rigid adherence to a fundamentalist principle.[321] Instead, it envisions that the allocation of functions will likely

---

[316] Yoo, *supra* note 313, at 77–80.

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

[318] KUROSE & ROSS, *supra* note 312, at 219–27; TANENBAUM, *supra* note 296, at 208–11.

[319] *See* TANENBAUM, *supra* note 296, at 553–55 (exploring indirect TCP and the inclusion of snooping agents as possible solutions to the problem).

[320] For surveys of this literature, *see* Fotis Foukalas et al., *Cross-Layer Design Proposals for Wireless Mobile Networks: A Survey and Taxonomy*, IEEE COMM. SURVS. & TUTORIALS, 1st Qtr. 2008, at 70; Sanjay Shakkottai et al., *Cross-Layer Design for Wireless Networks*, IEEE COMM., Oct. 2003, at 74; Vineet Srivastava & Mehul Motani, *Cross-Layer Design: A Survey and the Road Ahead*, IEEE COMM., Dec. 2005, at 112. For a more skeptical assessment, see Vikas Kawadia & P.R. Kumar, *A Cautionary Perspective on Cross-Layer Design*, IEEE WIRELESS COMM., Feb. 2005, at 3.

[321] James Kempf & Rob Austein, *The Rise of the Middle and the Future of End-to-End: Reflections on the Evolution of the Internet Architecture* 8, 10 (Network Working Group, Request for Comments 3724, 2004), *available at* http://tools.ietf.org/pdf/rfc3724.

change over time with shifts in technology and the underlying demand for network services.

c. Security

Another good example of the limits of the information contained in the IP header is network security. The current architecture does not permit verifiable information about the identity of particular end users to pass through the protocol stack.[322] Despite the growing need for security, the network has been slow to adapt to this new reality. The shifting importance of these concerns underscores the importance of not regarding any particular layered architecture as if it were a natural construct. Moreover, it underscores the potential dangers of using regulation to enshrine any particular architecture into law.

2. Increase in Variety

Another force providing impetus toward remodularization is the fundamental change in the external environment. The number of end users connecting to the Internet has exploded and become more heterogeneous in terms of geography, interests, and capabilities.[323] They are using a wider variety of applications, many of which place new and more intensive demands on the network.[324] End users are connecting through an ever broadening array of devices connected to an increasingly heterogeneous set of transmission technologies.[325] The business environment has become more complex as well, both in terms of pricing and topology.[326]

The Law of Requisite Variety dictates that the increase in the diversity of the external environment will require the architecture to evolve in response, either by increasing the number of modules or by redesigning the interfaces to reduce the number of intermodule interactions. Either would require a remodularization of the Internet stack.

3. Lack of Coordination

Network engineers have also long recognized that the decentralized nature of congestion management can lead to persistent problems.[327]

---

[322] *See supra* note 201 and accompanying text.

[323] CHRISTOPHER S. YOO, THE DYNAMIC INTERNET: HOW END USERS, TECHNOLOGY, AND BUSINESSES ARE TRANSFORMING THE NETWORK 13–18 (2012).

[324] *Id.* at 19–36.

[325] *Id.* at 37–54.

[326] *Id.* at 55–69.

[327] Bob Braden et al., *Recommendations on Queue Management and Congestion Avoidance in the Internet* 9–10 (Network Working Group Request for Comments 2309, Apr. 1998) [hereinafter RFC 2309], *available at* http://tools.ietf.org/pdf/rfc2309.

Because individual end users can maximize their private benefits by continuing to send traffic into the network while everyone else slows down, they have substantial private incentives to deviate from the social optimum.[328] The IP header does not provide any way for other actors to verify whether or not a host is responding appropriately to the presence of congestion. As noted earlier, the absence of a network-based means for ensuring that hosts reduce their transmission rates when the network is congested means that congestion management currently depends on what amounts to the honor system despite the fact that each end user has the incentive to deviate from the optimal behavior.[329] Although the Internet community may once have represented the type of close-knit community that could prevent such deviations from occurring, the rapid expansion of the Internet has undercut its ability to rely on social norms to protect against this type of behavior.[330]

The decisions surrounding what information to place in the IP header thus limits the network's ability to address certain types of problems. As in any modular architecture, the particular decisions about which interdependencies to acknowledge has a direct impact on the system's functionality. Imposing rules that forbid any deviations from the current architecture would risk foreclosing solutions to problems that continue to grow in importance.

4. Network Neutrality

The decade-long debate over network neutrality appears to be approaching a major turning point. At its core, network neutrality proponents oppose discrimination (providing some traffic more favorable treatment than others) and paid prioritization.

Providing a comprehensive survey of the network neutrality debate exceeds the scope of this article. For our purposes, it suffices to note that the ability to assign a higher priority to some packets has been a functionality contained in the visible information in the IPv4 since its inception. Prioritization would amount to nothing more than taking advantage of the

---

[328] Dmitri Dimitriou et al., *Open Research Issues in Internet Congestion Control* 31 (Internet Research Task Force Request for Comments 6077, 2011), *available at* http://tools.ietf.org/pdf/rfc6077 ("In the current Internet architecture, congestion control depends on parties acting against their own interest. 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."); RFC 2309, *supra* note 327, at 9–10, 11 (noting that TCP implementations "can grab an unfair share of the network bandwidth" by aggressively refusing to back off, which would logically lead to "a spiral of increasingly aggressive TCP implementations, leading back to the point where there is effectively no congestion avoidance and the Internet is chronically congested").

[329] *See supra* note 201 and accompanying text.

[330] YOO, *supra* note 323, at 82–99.

functionality built into the architecture. In the words of David Clark, "The Internet is not neutral and has not been for a long time."[331]

Equally importantly, many features of IP-enabled networks depend on some degree of prioritization. Proprietary voice over Internet Protocol (VoIP) services, such as Comcast Digital Voice, depend on some form of prioritization to ensure call quality, as does voice over LTE (VoLTE). Some video services, such as AT&T's U-verse network, ensure quality by prioritizing traffic associated with a single application (video) form a single provider (AT&T).[332] Finally, zero-rating systems such as T-Mobile's Music Freedom allow users to stream music without having that usage count against their data caps. Programs such as Facebook Zero, Twitter Zero, and Wikipedia Zero are using similar programs to help deploy Internet access in the developing world. They do so by making one application cheaper than others in a manner completely consistent with the existing architecture.

5. The Transition to IPv6

Perhaps the biggest change to the modular architecture of the Internet is the transition to IPv6. IPv4 addresses consist of 32 bits, which supports 2^{32} or roughly 4.3 billion addresses. Although the original Internet architects thought that was more than enough to satisfy the expected demand,[333] they did not anticipate that every consumer would have multiple devices connected to the network, let alone seek connectivity for numerous appliances in their home. As a result, the Internet began running out of new addresses in 2011.[334] Consequently, the engineering community developed

---

[331] David Clark, Written Statement to the En Banc Public Hearing on Broadband Network Management Practices (Feb. 25, 2008), *available at* http://www.fcc.gov/broadband_network_management/022508/clark.pdf; *see also* Robert W. Hahn & Robert E. Litan, *Portioning Bit by Bit: The Myth of Network Neutrality and the Threat to Internet Innovation*, MILKEN INST. REV., 1st Qtr. 2007, at 28, 31–33; Jonathan E. Nuechterlein, *Antitrust Oversight of an Antitrust Dispute: An Institutional Perspective on the Net Neutrality Debate*, 7 J. ON TELECOMM. & HIGH TECH. L. 19, 36–37 (2009); Christian Sandvig, *Net Neutrality Is the New Common Carriage*, 9 INFO 136, 139–42 (2007); Douglas A. Hass, Comment, *The Never-Was-Neutral Net and Why Informed End Users Can End the Net Neutrality Debates,* 22 BERKELEY TECH. L.J. 1565, 1576–77 (2007); Kai Zhu, Note, *Bringing Neutrality to Network Neutrality*, 22 BERKELEY TECH. L.J. 615, 634–36 (2007); Michael Grebb, *Neutral Net? Who Are You Kidding?*, WIRED, May 31, 2006, http://www.wired.com/news/technology/internet/0,71012-0.html; Andrea Renda, *I Own the Pipe, You Call the Tune: The Net Neutrality Debate and Its (Ir)relevance for Europe* 9–11 (2008), *available at* http://shop.ceps.eu/downfree.php?item_id=1755; Craig McTaggart, Was the Internet Ever Neutral? (2006) (unpublished manuscript presented at the 34th Annual Telecommunications Policy Research Conference), *available at* http://ssrn.com/abstract=2117601.

[332] Christopher S. Yoo, *The Changing Patterns of Internet Usage*, 63 FED. COMM. L.J. 6, _ (2010).

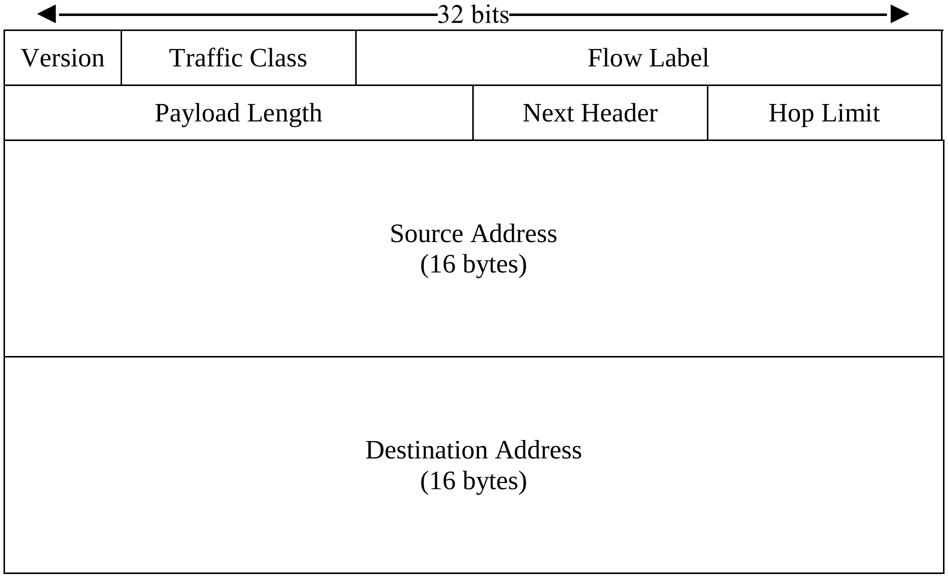
[333] Laurie Flynn, *Drumming Up More Addresses on the Internet*, N.Y. TIMES, Feb. 14, 2011, at B3; Paul McNamara, *Why IPv6? Vint Cerf Keeps Blaming Himself*, NETWORKWORLD, Oct. 22, 2010, http://www.networkworld.com/article/2227543/software/why-ipv6--vint-cerf-keeps-blaming-himself.html.

[334] The Internet Assigned Numbers Authority (IANA) allocates blocks of Internet addresses to five regional Internet registries (RIRs), which in turn allocate them to those who request them. IANA allocated its last block of addresses on February 3, 2011. The RIRs for the Asia/Pacific, Europe, and Latin American regions have already exhausted their allocations, with the RIRs for North America and

IPv6, which contained 2^{128} addresses[335] or enough to assign 47 trillion trillion addresses to every person on the planet.

In addition to increasing the size of the address space, deployment of a new network protocol provided an occasion to consider whether to modify other aspects of the network architecture. For example, goals included providing for better support for security, transmission of real-time data, and mobility.[336] The resulting IPv6 header is depicted in Figure 5.

Figure 5: The IPv6 Header



Source: Stephen E. Deering & Robert M. Hinden, *Internet Protocol, Version 6 (IPv6) Specification* 4 (Network Working Group Request for Comments 2460, 1998), *available at* http://tools.ietf.org/pdf/rfc2460.

Version specifies whether the packet is an IPv4 or IPv6 packet. *Traffic Class* permits sending nodes and routers to assign different levels of priority to particular traffic similar to the Type of Service field in the IPv4 header. *Flow Label* provides a basis for associating different packets together for quality of service or real-time service. *Payload Length* describes the length of the packet following the header (including options). *Next Header*

Africa projected to be depleted in 2015 and 2019. *IPv4 Address Report*, POTAROO, Sept. 1, 2014, http://www.potaroo.net/tools/ipv4/index.html.

[335] This is 340 undecillion or more specifically 340,282,366,920,938,463,463,374,607,431,768,211,456 addresses.

[336] TANENBAUM, *supra* note 296, at 465.

specifies whether the information immediately following the IPv6 header is one of the IPv6 extension headers that replaced the options in IPv4 or the transport protocol header that is the beginning of the payload. *Hop Limit* is a counter that is decremented by each time packet traverses a node until the counter reaches zero similar to the Time to Live field in the IPv4 header. *Source Address* and *Destination Address* specify the origin and termination point of the traffic.

As noted earlier, the information visible in the interface reflects the architecture's precommitment regarding the interdependencies that other modules are allowed and not allowed to take into account. Consequently, the visible information retained, removed, or added to the header as part of this remodularization is quite telling.

The retention of some fields is unsurprising. For example, the retention of fields specifying the source address, destination address, specifying the size of the packet, and time to live are unremarkable. The deletion of fields involving fragmentation and options reflect the shift of certain information outside the primary header in order to speed up routing by including fewer fields and standardizing the header length and shifting responsibility for fragmentation and packet integrity to the hosts operating at the edge of the network.[337]

Other decisions with respect to IPv6 header have more significant implications for the functionality of the network. These include the inclusion of fields to facilitate prioritization and the omission of fields to support security and mobility.

a. Prioritization

One of the most striking aspects of the IPv6 header is the inclusion of fields to facilitate prioritization of traffic. As noted above, the IPv4 header has always included a Type of Service field to support providing different levels of quality of service.[338] Although the original specification defined the field to focus solely on three types of quality of service: delay, reliability, and throughput,[339] it was subsequently redefined to permit more flexible use of this field through a protocol known as DiffServ.[340] The IPv6 Traffic Class field was explicitly designed to ensure that the experiments in differentiated

---

[337] *Id.* at _.

[338] Info. Sci. Inst., *Internet Protocol: DARPA Internet Program Protocol Specification* 2 (Request for Comments 791, 1981), *available at* http://tools.ietf.org/pdf/rfc791.

[339] *Id.* at 12–13.

[340] Steven Blake et al., *An Architecture for Differentiated Services* (Network Working Group Request for Comments 2475, 1998), *available at* http://tools.ietf.org/pdf/rfc2475.

services built on the IPv4 Type of Service field to continue.[341] The retention of this field thus reflects a continued commitment to support this functionality.

What is even more striking is the addition of a new field to support even more sophisticated forms of quality of service. As noted earlier, the original architecture envisioned that the Internet would route each packet through the network independently. The result is that packets associated with the same flow could follow different paths, which could cause them to arrive with inconsistent spacing or even out of order.

To solve this problem, the engineering community developed MultiProtocol Label Switching (MPLS), which assigns labels to packets associated with the same flow.[342] This permits all packets bearing the same label to be routed along the same path. In addition, labels permit the implementation of a broad range of network policies, such as route selection, prioritization, load balancing, and other forms of traffic engineering that can provide for better security and quality of service.[343] The MPLS header in which the label was embedded typically operates just below the network layer.[344]

IPv6 borrows this aspect from MPLS by including a field for a Flow Label in the header of the network layer. It did so explicitly to enable "special handling by the IPv6 routers, such as non-default quality of service or 'real-time' service."[345] The inclusion of labels in the IPv6 specification represents a renewed commitment to supporting prioritization of traffic. Even more importantly, the inclusion of the Flow Label in the spanning layer visible to all other network actors means that such traffic management represents part of the core functionality of the architectural design.

b. Security

Another area of controversy with respect to the design of IPv6 was security. Traditionally, security in IPv4 was implemented by the hosts operating at the edges of the network rather than by the network itself.[346] The IETF developed a security protocol for IPv6 that was retrofitted into IPv4 under the name IP Security (IPsec). IPsec is a fairly complex suite of

---

[341] Steven E. Deering & Robert M. Hinden, *Internet Protocol, Version 6 (IPv6) Specification* 25 (Network Working Group Request for Comments 2460, 1998), *available at* http://tools.ietf.org/pdf/rfc2460 [hereinafter RFC 2460].

[342] Eric C. Rosen et al., *Multiprotocol Label Switching Architecture* (Network Working Group Request for Comments 3031, 2001), *available at* http://tools.ietf.org/pdf/rfc3031.

[343] William Stallings, *MPLS*, INTERNET PROTOCOL J., Sept. 2001, at 2, 3.

[344] The MPLS header can either encapsulate an IP packet, in which case it operates much like a data-link layer packet, or it can encapsulate transport layer packets, in which case it operates like the network layer. This is why it is sometimes said to operate at layer 2.5.

[345] RFC 2460, *supra* note 341, at 25.

[346] TANENBAUM, *supra* note 296, at 473.

protocols.[347] Reduced to its basics, IPsec adds an extension header to the IP header with an authentication code to allow others to verify the packet's integrity and encrypts the original IP packet (along with padding needed to make the relevant encryption work) to ensure confidentiality. When it operates in transport mode, it only encrypts the payload, whereas when it operates in tunnel mode, it places the encrypted packet in another IP packet. It then uses a designated number in the *Protocol* field in IPv4 or the *Next Header* field in IPv6 to let the receiving host or router know that the communication is being made using IPsec. Thus, IPsec requires all receiving routers and hosts to recognize this designation. In other words, it requires that this designation be part of the visible information that all computers connected to the network are prepared to recognize.

The development of IPv6 involved a wide discussion of how much security-related information to place in the visible information. In the end, the change was fairly modest, consisting only of adding to the IP header the information necessary to authenticate the communication and to decrypt the payload. The actual authentication was left to the hosts. Interestingly even though the initial specifications of IPv6 made this security regime mandatory,[348] the specification was revised in 2011 to make it optional.[349] The removal of this information from the mandatory aspects of the interface represents a major change in architectural commitments. The fact that IPv6 opted not to include this much information represents a precommitment that the actual authentication be performed within one of the modules by limiting the amount of information that could pass through the interface.

c. Mobility

Another controversy surrounding IPv6 was the extent to which it would support mobility. As noted earlier, the IPv4 header includes a destination address, which identifies a physical location. The fact that mobile hosts change locations means that relying on the traditional architecture would require constantly updating the routing architecture to reflect each mobile host's current location. The problem is that updates to routing information take time to propagate through the network. That means that the network would struggle to maintain accurate information of the location of each mobile host.[350]

---

[347] For an accessible discussion, see KUROSE & ROSS, *supra* note 312, at 734–40.

[348] RFC 2460, *supra* note 341, at 7; Stephen Kent & Karen Seo, *Security Architecture for the Internet Protocol* (Network Working Group Request for Comments 4301, 2005), *available at* http://tools.ietf.org/pdf/rfc4301.

[349] Ed Jankiewicz et al, *IPv6 Node Requirements* (Internet Engineering Task Force Request for Comments 6434, 2011), *available at* http://tools.ietf.org/pdf/rfc6434.

[350] Christopher S. Yoo, *The Changing Patterns of Internet Usage*, 63 FED. COMM. L.J. 67, 81 (2010).

Instead, the mobility solution for IPv4 requires mobile hosts to designate a router on its home network as its *home agent* and asks anyone seeking to reach it send to that address. The home agent must also let other networks know that it is serving as the home agent for this particular mobile host. The mobile host then registers with a *foreign agent* on the network on which it is currently located either by contacting the foreign agent or by responding to an advertisement sent by the foreign agent. The foreign agent then sends a *care-of address* to the home agent to inform it where the packets addressed to the mobile host should be sent and to update that information if the mobile agent should transfer to a different node. The home agent then encapsulates any traffic that it receives and forwards it to the foreign agent for delivery. The mobile agent then decapuslates the traffic.[351]

This solution obviates the need for updating the routing tables, but does incur some inefficiencies. The solution is quite complex, requiring no fewer than seven protocols and large number of signaling messages between the hosts and the routers to set it up. In addition, it can lead to what is often called "triangle routing," because instead of passing directly to its destination, the packets must travel via the home agent. In extreme cases, the packets associated with a file being transferred between two people sitting next to each other in a conference room on the West Coast might have to travel to the East Coast and back. As such, it represents an apt illustration of David Wheeler's aphorism: "All problems in computer science can be solved by another level of indirection . . . except for the problem of too many layers of indirection."[352]

Some engineers suggested pursuing a more straightforward solution to this problem during the transition to IPv6. Instead of relying on the hosts to send the packets indirectly by way of the home agent, the updated location information could be included in the network layer by having packets include the address of the mobile host's current location.

The solution was never adopted because of two concerns. The first was the difficulty in updating the mobile host's location should the mobile host shift to a new network. The second was the danger that an imposter could masquerade as the mobile host and highjack all of the communications bound for it.[353]

Consequently, IPv6 declined to incorporate mobility support into the network layer and instead retained the indirect, host-based approach employed for IPv4. The IPv6 protocol did make some small refinements, such as eliminating the need for a foreign agent and routing return traffic directly instead of via the home agent (although this may be impossible if the

---

[351] Charles E. Perkins, *IP Mobility Support for IPv4* (Internet Engineering Task Force Request for Comments 5944, 2010), *available at* http://tools.ietf.org/pdf/rfc5944.

[352] *Talk:Indirection*, WIKIPEDIA (15:46, Nov. 9, 2010), http://en.wikipedia.org/wiki/Talk:Indirection.

[353] CHRISTIAN HUITEMA, IPV6 _ (2d ed. 2008); TANENBAUM, *supra* note 296, at 472.

foreign network's gateway routers check the source address of IP packets and discard those that did not originate on its subnets).[354]

The basic approach is not implemented through information in the primary interface visible to all other networks. Instead, mobility relies on the hosts to perform a number of complicated functions. Relying on hosts to handle mobility means that all of the hosts must be reconfigured in order to make any changes to the network. This can be quite cumbersome and tends to retard innovative change.

As a result, calls for embedding mobility support in the network layer have persisted.[355] One major problem is that addresses serve two distinct purposes: labeling both the identity of the host and its location. The combination of these functions was unproblematic when hosts were immobile personal computers. The growing importance of mobility has led many engineers to call for an identity-locator split, in which distinct addresses are used to specify the identity of the host and its location.[356] New proposals along these lines continue to appear,[357] although other engineers have questioned that approach.[358] Still others place the responsibility for exchanging signaling messages with the home agent on proxy agents instead of the mobile host.[359] Empirical studies suggest that network-based approaches may improve network performance.[360]

For the purposes of this article, whether the architecture ever embraces network-based mobility support is immaterial. The ongoing debate underscores how shifts in the underlying technology and demand for wireless devices are changing the demands that people are placing on the network. It also provides an apt illustration of how influential decisions about what information to include in interfaces can affect the functionality of the network.

C. Open Application Programming Interfaces (APIs)

Modularity theory also sheds new light on calls for platform providers to open up their *application programming interfaces* (APIs). APIs are the

---

[354] Charles E. Perkins, *Mobility Support in IPv6* (Internet Engineering Task Force Request for Comments 6275, 2011), *available at* http://tools.ietf.org/pdf/rfc6275.

[355] For a survey, see Chakchai So-In et al., *Future Wireless Networks: Key Issues and a Survey (ID/Locator Split Perspective)*, 8 INT'L J. COMM. NETWORKS & DISTRIB. SYS. 24, 30–31 (2012).

[356] For a survey, see W. Ramirez et al., *A Survey and Taxonomy of ID/Locator Split Architectures*, 60 COMPUTER NETWORKS 13 (2014).

[357] *See, e.g.*, Arun Venkataramani et al., *MobilityFirst: A Mobility-Centric and Trustworthy Internet Architecture*, ACM SIGCOMM COMPUTER COMM. REV., July 2014, at 74.

[358] Dave Thaler, *Why Do We Really Want an ID/Locator Split Anyway?*, Keynote Address at the 3rd ACM Workshop on Mobility in the Evolving Internet Architecture (MobiArch '08) (Aug. 22, 2008), http://research.microsoft.com/en-us/um/people/dthaler/dthaler-mobiarch-keynote.pdf.

[359] Sri Gundavelli et al., *Proxy Mobile IPv6* (Network Working Group Request for Comments 5213, 2008), *available at* http://tools.ietf.org/pdf/rfc5213.

[360] Nitul Dutta et al., *Survey on Mobility Management Protocols for IPv6 Based Network*, ADVANCES IN NETWORK & COMM., July 2013, at 15, http://www.humanpub.org/ANC/ppl/ANC6PPL.pdf.

interfaces that application providers use to give instructions to the underlying platform.[361] APIs increase the modularity of the platform by providing a uniform set of commands that applications can use to communicate with the platform. For example, Microsoft maintains a number of APIs that application providers can use to produce software for the Windows operating system. Apple's mobile operating system has a rich set of APIs that has unleashed a torrent of fascinating applications for the iPhone.[362] So many people are producing applications for Facebook that some observers have speculated Facebook's APIs will displace the Internet suite protocols as the dominant communications platform.[363]

There have been increasing calls to require platform owners to make their APIs open. Strategic denial of access to APIs to certain outside developers was the heart of the antitrust case against Microsoft.[364] Calls for open access to smartphones such as the iPhone[365] and social networking sites such as Twitter and Facebook[366] have also emerged. Opening APIs would yield many potential benefits of faster innovation and greater flexibility.

At the same time, mandating open interfaces would have a number of drawbacks. For example, Apple has traditionally insisted on closed APIs in order to guarantee that the end-users' experience remained positive. Although it has loosened this policy somewhat with the iPhone, it still reviews all software before making it available through its app store and insists on a share of the revenue that app developers generate through its product.[367] Moreover, following the same approach as one's competitors is generally bad business strategy; the mere fact that one's competitor had

---

[361] PETERSON & DAVIE, *supra* note 3, at 31.

[362] *See* Kevin J. Boudreau & Karim R. Lakhani, *How to Manage Outside Innovation*, MIT SLOAN MGMT. REV., Summer 2009, at 69, 69.

[363] *See* Jason Chew, *Is Facebook the New Internet?*, PROACTIVEINVESTORS, Oct. 1, 2011, http://www.proactiveinvestors.com/companies/news/19122/is-facebook-the-new-internet--19122.html; Mark Cuban, *Is Facebook the New Internet and How Long Before Microsoft Tries to Buy It?*, BLOG MAVERICK, Apr. 22, 2010, http://blogmaverick.com/2010/04/22/is-facebook-the-new-internet-and-how-soon-before-microsoft-tries-to-buy-it/; Andrew Yee, Why Facebook Is the New Internet, CLOUD TALK, Mar. 6, 2011, http://www.ebizq.net/blogs/cloudtalk/2011/03/why_facebook_is_the_new_intern.php.

[364] *See* Massachusetts v. Microsoft Corp. 373 F.3d 1199, 1216–22, 1240–41 (D.C. Cir. 2004). Note that the district court rejected the states' request that Microsoft be ordered to modularize Windows so that outside developers could replace portions of the actual Windows operating system itself. New York v. Microsoft Corp., 224 F. Supp. 2d 76, 162, 251–52 (D.D.C. 2002), *aff'd sub nom.* Massachusetts v. Microsoft Corp. 373 F.3d 1199 (D.C. Cir. 2004).

[365] ZITTRAIN, *supra* note 2, at 1–5.

[366] The Federal Trade Commission has investigated Twitter's recent moves to limit access to its APIs. *See* Amir Efrati, *Antitrust Regulator Makes Twitter Inquiries*, WALL ST. J., July 1, 2011, at B3. Facebook previously restricted access to its APIs, but provide more access to counter Google's Open Social Initiative. *See* Brad Stone, *To Counter Google, Facebook Opens Its Code*, N.Y. TIMES BITS BLOG, June 2, 2008, http://bits.blogs.nytimes.com/2008/06/02/to-counter-google-facebook-sets-code-free.

[367] *See* Jonathan M. Barnett, *The Host's Dilemma: Strategic Forfeiture in Platform Markets for Informational Goods*, 124 HARV. L. REV. 1861, 1920 (2011). Apple reportedly restricted the programming techniques that iPhone application developers could use, only to back off in the face of inquiries by European antitrust regulators. *See* Sascha D. Meinrath et al., *Digital Feudalism: Enclosures and Erasures from Digital Rights Management to the Digital Divide*, 19 COMMLAW CONSPECTUS 423, 461–62 (2011).

adopted an open platform provides a substantial reason to adopt a closed one.[368] The presence of a popular open platform, such as smartphones based on Google's Android operating system, helps ensure that Apple's iPhone business practices will not harm competition or innovation.

The advent of smartphones has also unleashed a dramatic increase in the heterogeneity of applications running on mobile operating systems. To meet this demand, mobile operating system providers are experimenting with a wide range of new functionalities, many of which are implemented in widely different ways. For example, instead of providing video chat as a separate application, Apple's FaceTime builds that feature into the underlying operating system. Google Wallet goes a step farther, taking a payment-system functionality traditionally provided as an application and building it into the chip. The growing heterogeneity of demand and dynamic nature of this platform make any attempt to mandate open access to any fixed set of APIs inherently problematic.

Even more interesting from the perspective of modularity theory is the lack of coordination. Modularity focuses actors on optimizing their individual interests, which may or may not in the aggregate optimize the performance of the system as a whole.[369] Given that all companies are drawing off a common pool of surplus that they jointly create, they have the incentive to want others to do more.[370] As was the case with GPTs,[371] this can lead to suboptimal investment in the entire modular ecosystem.

The rise and fall of the IBM PC provides a useful example of the importance in the shift. Unlike Apple, who was the early market leader, IBM adopted an open modular architecture for its PC. Initially, this strategy proved spectacularly successful, as the open architecture reduced the magnitude of IBM's upfront investment, allowed it to market is PC in only fifteen months, and unleashed a wide range of innovation within its platform.[372] Fairly quickly, however, IBM's strategy began to evince a number of downsides. The same modularity that unleashed parallel implementation permitted IBM's former partners to capture an increasing amount of the value and eventually to ally with other computer manufacturers that would ultimately displace IBM.[373] The result was a lack of innovation in the platform. Similar complaints have been levied at Android. This contrasts starkly with Apple, which continues to both manufacture and innovate in computers and smartphones.

---

[368] CARL SHAPIRO & HAL R. VARIAN, INFORMATION RULES 23–27, 298(1999).

[369] *See* Chesbrough & Teece, *supra* note 182, at 66.

[370] *Id.* at 67–69.

[371] *See supra* notes 196–199 and accompanying text.

[372] *Id.* at 267–68; Chesbrough & Teece, *supra* note 182, at 68–69.

[373] BALDWIN & CLARK, *supra* note 4, at 268; Chesbrough & Teece, *supra* note 182, at 69–70.

Companies must think carefully about which parts of the value chain to continue to develop internally.[374] Most notably, modularity allows imitation as well as innovation.[375] The lack of coordination can lead to significant economic harms.

D. Future Internet Architecture Proposal

Policymakers and commentators frequently assert that the Internet's architecture has played an instrumental role in its success and argue that that architecture must be preserved.[376] Interestingly, this assertion is contradicted by a steady drumbeat of articles in the engineering literature noting the functions that the Internet does not perform well.[377] These include security, mobility, quality of service, mass media distribution (multicasting), and support for multiple connections to the same location (multihoming).[378] Although these were not important when the Internet first emerged as a mass-market phenomenon during the mid-1990s, they have now become mission critical.[379]

These analyses suggest that TCP/IP is a forty-year-old technology that was designed for a very different context and is being asked to do more than its designers ever imagined. Many engineering scholars complain that the Internet has become ossified and impervious to significant architectural change.[380] Others are pursuing a variety of "clean slate" initiatives examining how the Internet's architecture might better meet these needs if it were designed from scratch today.[381] Government agencies in the U.S.,

---

[374] Chesbrough & Teece, *supra* note 182, at 71.

[375] Ethiraj et al., *supra* note 28, at 941.

[376] *See supra* note _ and accompanying text.

[377] *See, e.g.*, Jon Crowcroft, *Net Neutrality: The Technical Side of the Debate*, COMPUTER COMM. REV., Jan. 2007, at 49, 51; Handley, *supra* note 311; Raj Jain, *Internet 3.0: Ten Problems with Current Internet Architecture and Solutions for the Next Generation*, PROC. MIL. COMM. CONF. (MILCOM 2006) (2007), http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4086425; Thrasyvoulos Spyropoulos et al., *Future Internet: Fundamentals and Measurement*, 37 COMPUTER COMM. REV., Apr. 2007, at 101; Sixto Ortiz, Jr., *Internet Researchers Look to Wipe the Slate Clean*, COMPUTER, Jan. 2008, at 12.

[378] *Id.*

[379] YOO, *supra* note 323, at 3–4.

[380] For citations to eight representative examples, see Paul Laskowski & John Chuang, *A Leap of Faith? From Large-Scale Testbed to the Global Internet* 2 (Sept. 27, 2009) (paper presented at 37th Annual Telecommunications Policy Research Conference); *see also* Olivier Martin, State of the Internet & Challenges Ahead 1, 29 (2007), http://www.ictconsulting.ch/reports/NEC2007-OHMartin.doc (noting that "there appears to be a wide consensus about the fact that the Internet has stalled or ossified").

[381] *See, e.g.*, Steven Bellovin et al., *A Clean Slate Design for the Next-Generation Secure Internet*, Report for NSF Global Environment for Network Innovations (GENI) Workshop (July 12–14, 2005), *available at* http://sparrow.ece.cmu.edu/group/pub/bellovin_clark_perrig_song_nextGenInternet.pdf; Jon Crowcroft & Peter Key, *Report from the Clean Slate Network Research Post-SIGCOMM 2006 Workshop*, COMPUTER COMM. REV., Jan. 2007, at 75; Anja Feldmann, *Internet Clean-Slate Design: What and Why?*, COMPUTER COMM. REV., July 2007, at 59; Ortiz, *supra* note 377; Thomas-Rolf Banniza et al., *A European Approach to a Clean Slate Design for the Future Internet*, 14 BELL LABS TECHNICAL J. 5 (2009); International Center for Advanced Internet Research (ICAIR), *Grand Challenges in Advanced Networking Research*, (May 10, 2010), http://www.icair.org/mission/grand-challenges.html; Stanford University Clean Slate, http://cleanslate.stanford.edu/ (last visited Mar. 14, 2011). For a more cautionary

Europe, and Asia have all launched programs to explore alternative architectures better suited to supporting the new demands that end users are placing on the network.[382] For example, the National Science Foundation's Future Internet Architecture program asked all participating projects to "mak[e] security an integral part of the architecture" and "build security into the design."[383] For example, both the NEBULA and Named Data Networking projects required that each packet be cryptographically signed in the network layer, effectively making each packet self-authenticating.[384]

Modularity theory emphasizes that all designs are the products of their times rather than the result of some preconceived vision of the ideal architecture.[385] In addition, changes in needs and technological interdependencies can lead to remodularization. Although they are rare, remodularizations do occur.[386] For example, although Windows once represented a form of middleware running on a DOS operating system, it has now displaced DOS as the primary operating system platform. Similarly, although packet transmission began as an application riding on a voice network, packet transmission has now become the relevant platform, and voice has become an application riding on a data network rather than vice versa.[387]

The costs of remodularization dictate that any such proposals should be approached with considerable caution. Modularity theory does offer some guidance as to the types of circumstances under which remodularization may be justified, including changes in the nature of technological interdependencies, increases in the degree of technological opportunity, and the growing heterogeneity of demand. Moreover, the inherent costs of remodularization raise the real possibility that an architecture may become locked into a modularization that is suboptimal. As the literature exploring how the presence of a large installed base can dampen innovation demonstrates, the real risk may be too little change rather than too much.[388]

---

assessment, see Constantine Dovrolis, *What Would Darwin Think About Clean-Slate Architecture?*, COMPUTER COMM. REV., Jan. 2008, at 29.

[382] For an overview, see Jianli Pan et al., *A Survey of the Research on Future Internet Architectures*, IEEE COMM., July 21, 2011, at 26, 28–34. *See also* David D. Clark et al., *New Arch: Future Generation Internet Architecture*, Final Report (2003), *available at* http://www.isi.edu/newarch/iDOCS/final.finalreport.pdf.

[383] Darleen Fisher, *A Look Behind the Future Internet Architectures Efforts*, ACM SIGCOMM COMPUTER COMM. REV., July 2014, at 46, 46, 47.

[384] Lixia Zhang et al., *Named Data Networking*, ACM SIGCOMM COMPUTER COMM. REV., July 2014, at 66, 68; Tom Anderson et al., *A Brief Overview of the NEBULA Future Internet Architecture*, ACM SIGCOMM COMPUTER COMM. REV., July 2014, at 81, 82.

[385] ABBATE, *supra* note 293, at 51 ("The ARPANET's builders did not start out with a specific plan for how functions would be divided up among layers or how the interfaces and protocols would work. Rather a layered model evolved as the ARPANET developed.").

[386] RFC 817, *supra* note 158, at 25.

[387] Dave Clark et al., *Overlay Networks and the Future of the Internet*, 63 COMM. & STRATEGIES 1, 1–2 (2006).

[388] *See, e.g.*, Joseph Farrell & Garth Saloner, *Installed Base and Compatibility: Innovation, Product Preannouncements, and Predation*, 76 AM. ECON. REV. 940, 941 (1986).

CONCLUSION

The benefits of modularity are frequently invoked as a justification for the imposition of regulations designed to preserve the status quo. There is no question that existing modular architecture can reduce complexity, speed innovation, facilitate the division of labor, and promote flexibility. At the same time, modularity carries with it certain drawbacks, a fact that is rarely acknowledged or understood. The balance of countervailing forces determines whether a particular modular architecture is beneficial as well as how it should be implemented in terms of the number of modules, the division of tasks, and the construction of the necessary interfaces. Modularity theory underscores that this tradeoff cannot be determined a priori. Indeed, some scholars argue against a "presumption of 'promodularity' bias."[389]

Policymakers should therefore avoid treating the existing network architecture as if it were a natural construct that must be preserved at all costs. Instead, modularity's contingent quality underscores the importance of developing heuristics for determining how and when a modular architecture should evolve.

Even determining that modularity is the preferred policy is not sufficient by itself to justify regulatory intervention. Network effects already provide strong incentives toward compatibility. Governmental action is necessary only if systems refuse to adopt modularity when it would be beneficial to do so. Moreover, the policy question should not be posited as a choice between modularity and nonmodularity. The optimal solution may be a hybrid, such as the one created by the coexistence of the open Google Android platform and the closed platform of the Apple iPhone. Purely as a matter of business strategy, the fact that one's competitor had adopted a modular strategy makes it beneficial to pursue the opposite course. Moreover, consumers benefit from having more choices, at the same time as the presence of a major open platform protects against anticompetitive harms.

On a broader level, the increasing frequency with which architectural concepts such as modularity are invoked during policy debates makes understanding of their conceptual underpinnings all the more important. Otherwise, the opacity of the engineering concepts will obscure debates rather than providing insights while at the same time permitting advocates to introduce normative assumptions without appearing to do so.[390]

---

[389] Ethiraj & Levinthal, *supra* note 9, at 172.

[390] *See* Tarleton Gillespie, *Engineering a Principle: "End-to-End" in the Design of the Internet*, 36 SOC. STUD. SCI. 427, 450 (2006) (noting that "[b]y highlighting certain features of the technology and obscuring others," framing policy debates in terms of architectural concepts has "the power to frame the entire technology in terms of an assumed set of priorities" in a way that is "often cloaked in a discourse that performs its political neutrality," but actually "embody political standpoints, even as they obscure them"); Marjory S. Blumenthal, *End-to-End and Subsequent Paradigms,* 2002 L. REV. MICH. ST. U. DET. C.L. 709, 710 (2002) (noting that "[a]lthough the embrace of engineering principles . . . appears to impart

---

a legitimacy to certain kinds of advocacy, that advocacy reaches beyond the engineering to the ideology long associated with the Internet”).

Abstract

This preliminary paper is a summary of an analysis of net neutrality rules in some twenty countries. A database of the individual components of each country's rules and corresponding country and market statistics was built to compare the similarities and differences across countries. The analysis shows that only the US and Canada have net neutrality rules carved from existing telecom regulations. Parliamentary legislation and guidelines have been used equally in the remainder of the countries, with soft law having a slightly better outcome to deter net neutrality violations. Where Parliamentary legislation is used, it generally includes updating existing communications laws to include net neutrality and/or to conduct a formal political process that legitimizes the rulemaking.

With seven lawsuits already filed against it since voting on rules on February 26, the FCC faces a third court case to defend its net neutrality rules. Canada's telecom regulator faces a court case as well. A lawsuit against Chile's telecom regulator precipitated the need for the country to resolve the issue with congressional legislation. In Europe regulators contend that their telecom laws already empower them to address net neutrality concerns, but most nations have agreed to support some kind of net neutrality law to clarify the government's responsibilities and limitations.

Dynamic Coalition of Net Neutrality (the association of the world's net neutrality advocacies) founder Luca Belli observed that the FCC's approach is not "sustainable" and expressed his support for a legislative solution provided that the agency's funding is not "gutted" by Congress.[1] The findings of this paper suggest that a legislative solution to net neutrality is preferable to continued FCC rulemaking and resulting litigation.

Global overview of the legal instruments of net neutrality

A general theory[2] of the framework for telecom regulation is that the legislative body, a Parliament or Congress, sets the strategic direction and makes the laws. It defines the goals and authority of the regulator. The regulatory body implements the laws as defined by the legislative body. The operator provides service within the confines of the rules and the economic and physical realities of the service area.

Under this theory, whether there is a need for net neutrality rules under telecom regulation would first be a question of whether the marketplace is competitive and can obviate the problem. If not, then the telecom regulator would implement regulations within its charter and authority. Where such direction is not clear, the legislative body would empower the telecom regulator to do by making a law so that the regulator can deliver the required regulation.

---

[1] Personal interview with Luca Belli 22 April 2014 and elaborated in "A Discourse-Principle Approach to Network Neutrality: A Model Framework and Its Application," *MediaLaws - Law and Policy of the Media in a Comparative Perspective* -, accessed April 24, 2015, http://www.medialaws.eu/model-framework-on-network-neutrality/.

[2] Castaneda Araceli, Mark A. Jamison, and Michelle Phillips, *Considerations for the Design and Transformation of Regulatory Systems*, 2014, Warrington College of Business Administration, PURC Working Paper, University of Florida

While each country in the study generally conforms to this theory of regulation, the experience with net neutrality, the choice of instrument, and the components of the instrument may differ. However rules fall into three basic categories defined for this paper, legislation, telecom regulation, and guidelines. Following is a review of the types of instruments for net neutrality, their similarities, and their differences.

Legislation

It is significant to note that at least eight countries and the European Union have chosen the legislative approach. In Latin America this includes not only Chile in 2010, but Colombia in 2011 which copied the Chilean law nearly word for word, including misspellings and punctuation mistakes. In 2012 Peru passed its net neutrality law as part of the National Broadband Plan and Construction of a Fiber Backbone, though net neutrality like rules were codified since 2005. In 2014 Brazil created net neutrality laws as part of its Civil Framework for the Internet, and Mexico wrote net neutrality law last year, with a section of the new Mexican constitution dedicated to telecommunications.

In Europe net neutrality rules were promulgated in Netherlands in 2011, Slovenia in 2012, and Israel in 2013. In 2014 the European Parliament voted in favor of net neutrality rules, but passage awaits approval by the Council of Ministers, the heads of state of the EU countries.

Among some of the challenges in the EU are that governments are shareholders in telecom networks, and many perceive that net neutrality rules to be detrimental to earnings and to the efficient operations of networks. Moreover the EU still suffers the lingering effects of the 2008 financial crisis, and there is a not evidence that net neutrality improves the global competitiveness of the EU or creates jobs, two of the key goals of the Digital Single Market legislation under which the net neutrality rules have been offered. The current European Commission wants to foment European platforms for search engines, social networks, ecommerce and so on. The proliferation of American platforms is seen as the outgrowth of suboptimal tax policy and failed antitrust enforcement. Thus for European leaders to shift to support home-grown solutions is not a policy of neutrality.

The Dutch net neutrality law is used as a litmus test for the continent. The Dutch net neutrality rule that was supposed to support a flowering of local content and services has instead created a “Netflix effect”[3] in which the rules allow a single American company to consume twenty percent of the country’s bandwidth with a small subset of users. The law that was supposed to be a “silver bullet” has created new problems.[4]

Regulation

The US and Canada are the two countries remaining which attempt to carve net neutrality rules out of existing telecom regulation. In the case of Canada, the telecom regulator began a proceeding in 2008 specifically on the issue of throttling of peer2peer traffic. In the following year it published its internet traffic management principles (IMTPs). The IMTPs are based upon normative considerations of transparency, innovation, clarity,

---

[3] Van Eijk and Nico, *The Proof of the Pudding Is in the Eating: Net Neutrality in Practice, the Dutch Example*, SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, August 2, 2014), http://papers.ssrn.com/abstract=2417933.

[4] Ibid

and competitive neutrality, but not technical standards. Provisions against blocking and throttling are included. The Canadian rules have not been challenged until the CRTC's recent ruling against the exemption of an operator's mobile video service from a data cap. Bell Mobility is appealing the ruling, claiming that the CRTC is incorrectly regulating a broadcast service with telecom law. Moreover Bell asserts that it complies with Canadian broadcaster obligations which include a series of taxes and levies to support Canadian content.

In the US, the FCC has bolstered its decision to impose net neutrality through its interpretation of the Verizon v. FCC court case, that net neutrality rules can only be enforced common carriers.[5] The FCC's vote of February 26 simultaneously classified broadband as a Title II service and applied net neutrality rules.

Guidelines

The provision of soft law[6] in the form of guidelines, co-regulatory approaches, and voluntary codes of conduct is an interesting and important as a form of net neutrality rule making. These approaches have been practiced, if not formalized, in at least eight countries. They include regulation between regulators and operators, codes of conduct for operators, and multi-stakeholder dialogues in which a range of actors participate. The analysis for this paper shows that the incidence of net neutrality violations is less in these countries. It suggests that soft law works as an effective deterrent to violation, for operators fear the threat of harder rulemaking.

Japan, South Korea and France use which may be termed as a co-regulatory approach. With this instrument the regulator plays a central role where rules are understood but not necessarily codified in law. This approach may rely on the regulator to define and enforce a quality of service. This approach may reflect countries where the telecom regulator may be protective of national telecom industries.

Having strong telecommunication companies has long been the goal of industrial policy in Japan and South Korea, and building world-class broadband networks has been a national project in both countries. Therefore there has been an understanding that the government would provide certain conditions to ensure economic viability for operators, whether through subsidies for build-out or adoption, the ability to quickly amortize losses and capital costs, requirements that landlords pay for coverage, or a stream of revenue in the future. Telecommunications and media industries were vertically integrated on purpose, also by necessity as both countries have distinct languages for which content development is almost non-existent outside of the country.

Net neutrality emerged as an issue on mobile. Over-the-top (OTT) messaging services such as the Korean Kakao Talk and the Japanese Line compete with operators' native SMS solutions. The conflict puts the

---

[5] *Comments of Public Knowledge and Common Cause*, March 21, 2014, https://www.publicknowledge.org/assets/uploads/documents/Public_Knowledge_Common_Cause_Open_Internet_706_Public_Notice_Comments.pdf.

[6] To be clear, these designations are not absolutely distinct as some countries may practice some aspects of more than one model Norway calls it approach co-regulatory, but it also conforms to the pan-Nordic coordination of telecom regulators on net neutrality, a voluntary effort in the region to coordinate telecom regulation, which has eschewed hard rulemaking in lieu of an inclusive dialog. Moreover the Norwegian includes an element of outreach to civil society stakeholders as part of its process. The classifications are offered as a shorthand.

government in a difficult position; it doesn't want to restrict citizens' access to applications, nor does it want to punish operators earning a return on their investment.

Japan has chosen to use a "co-regulatory" approach to focus on congestion management with minimum standards for quality of service. Unfortunately "QoS literacy" is considered too technical for most consumers understand.[7] South Korea has also struggled with creating standards for reasonable network management, the terms and measurement of which are not practical for users.[8] Thus net neutrality is left to the regulatory elite, reinforcing the government's role to regulate the Internet.

The French approach to net neutrality includes the regulator's role to oversee interconnection. Moreover the country presented the notion of platform neutrality as well as the neutrality of devices and content.[9]

The United Kingdom and Switzerland have similar codes of conduct. Both sets of rules recognize the rights of users and agree to a no blocking rule. The UK code emphasizes that managed services are allowed, provided that they does not degrade the best efforts internet. A Key Facts Indicators is issued by the regulator to offer as an objective measure of the success of operators adhering to the code.

In late 2014 cable and telecom operators in Switzerland adopted their code of conduct after a regulatory-facilitated multi-stakeholder dialogue with a variety of actors from industry, civil society and government. The effort also created an Arbitration Board to oversee disputes.

Sweden, Norway and Denmark have formalized their process to conduct a multi-stakeholder dialogue and have a pan-Nordic cooperation on net neutrality. The Swedish approach to transparency is so strong that the regulator believes that the EU laws would be a step backward to their existing regime.[10] The Norwegian dialogue was initiated by the regulator, but in Denmark, operators and their industrial association took the lead.

That these countries chose a multi-stakeholder dialogue may reflect a culture that favors consensus and dislikes hierarchy. Moreover high labor cost forces government agencies to find solutions that do not require hiring more workers. Indeed Denmark's center left government dismantled its telecom authority in 2011,[11] and redeployed employees in other agencies. This overnight decision happened without rancor and reflects a belief that telecom experts add little value micromanaging networks but instead should work to enable broadband in other sectors. Delivering the rule of law does not require a traditional telecom authority,

---

[7] Toshiya Jitsuzumi, "An Analysis of Prerequisites for Japan's Approach to Network Neutrality," 2012, doi:10.2139/ssrn.2030029.

[8] Dong-Hee Shin, "A Case Study of Network Neutrality in Korea," 2012, http://www.econstor.eu/handle/10419/60383.

[9] CNNum, *Platform Neutrality : Building an Open and Sustainable Digital Environment*, May 2014, http://www.cnnumerique.fr/wp-content/uploads/2014/06/PlatformNeutrality_VA.pdf.

[10] "Ola Bergström, Director at Swedish Post and Telecom Authority - PTS, Gives an Interview at ETNO-MLex Summit 2014," *viEUws*, accessed April 24, 2015, http://www.vieuws.eu/etno/etno-etno-mlex-summit-2014-interview-with-ola-bergstrom-director-for-international-affairs-swedish-post-and-telecom-authority-pts/.

reasoned the officials that made the decision.[12] While telecom investment in Denmark is nearly 100 percent private, the Norwegian and Swedish governments are still owner in the incumbent telecom company as well as in municipal fiber networks.

There is no doubt that multi-stakeholder models take the regulators time, but so does monitoring the network the network and enforcing punishment with ex ante rules. The advantage of the model is that it allows the participation of many stakeholders, and users have greater involvement. Moreover it avoids regulatory errors and lowers costs to society, especially when violations are not present. It might also be observed that multi-stakeholder models may be the appropriate tool where an industry is emerging such as broadband and OTT services.

Competition/Antitrust

A number of countries have no specific net neutrality rules of any kind, and there is not necessarily greater incidence of violation. This is one reason why many believe that problems of net neutrality can be adjudicated on a case by case basis with competition law. Proponents of competition law assert that net neutrality is just a recapitulation of issues related to market access, bundling and tying, predatory pricing, squeezing, discrimination, and refusal to supply. They assert that net neutrality concerns could be fully investigated and discharged with competition law. Net neutrality advocates counter that competition law often fails to define the relevant market, and therefore that the neutrality concept to must be elevated with specific legislation.

Competition law is focused on abuse of dominant position when there is individual or collective dominance, abuse of horizontal restraints, and abuse of vertical restraints when such vertical restraints exceed the market share thresholds of 20–30 percent. Smaller companies are not subject to the prohibitions on the grounds that they have no impact on competition and efficiency conditions. A disadvantage of the competition rules is that the rules do require administrative resources not only by the authorities, but also for businesses. This is particularly true in the context of defining the relevant market and evaluating the effects of restraints and abuse.

As such, net neutrality supporters have advocated the need for proactive rules to prevent misbehavior rather than wait for abuses to occur. They argue that it is less costly to prevent a problem rather than to address a problem after it starts. However net neutrality proposals have not accurately indicated the costs that such ex ante rules impose, costs of which could be greater than the supposed harm they propose to avert. Some of the costs include enforcement, foregone opportunities of firms that decline new business models because of restrictive rules, and errors of misdiagnosis.

Indeed, estimates and judgments would be a feature of net neutrality, just as it is in competition law. There would be margin for error in uncertainty. At least the advantage of competition law is that it brings evidence of consumer harm, rather than net neutrality which relies on presumption. In recent years, the practice of

---

Michael K. Powell, *Preserving Internet Freedom: Guiding Principles for the Industry*, February 8, 2004, https://apps.fcc.gov/edocs_public/attachmatch/DOC-243556A1.pdf.

[12] Ibid

competition law has led to greater legal certainty about the rules of interpretation. That is to say that competition law is a highly evolved platform which could investigate the charges brought by net neutrality supporters.

There have been a number of competition cases on national and international levels in relation to the Internet. Examples of such cases are: market access (Microsoft), bundling and tying (Microsoft and Viasat), predatory pricing (Wanadoo), squeeze (Deutsche Telekom), discrimination (Intel), and refusal to supply (Microsoft). The cases show that the competition rules are effective and can be enforced. It also suggests that network neutrality concerns could be broadly enforceable through general competition rules, except for restrictions and abuses that have no noticeable impact on competition and efficiency.

Along with general data protection rules, consumer legislation, standardization rules, universal human rights, and a host of other general rules, it is therefore sufficient to ensure a fair enforcement of network neutrality rules without the introduction of specific regulation. Specific regulation on network neutrality will likely lead to regulatory failure, reduced competition, reduced innovation, extra administrative hassle, and welfare loss.

Market access is the main function of competition. In telecommunications, there has been a tradition of sector-specific rules to ensure access to networks with significant market power and to ensure an equitable distribution of spectrum when there is scarcity. However as recognized in other capital-intensive, highly regulated industries, as industries become more competitive, sectoral regulation can be removed in favor of general competition law. This is understood to be the socially beneficial outcome to make efficient uses of regulatory resources.

Key Elements of Rules

As part of this research, the net neutrality rules of the countries were analyzed and coded for their key elements. Perhaps the key difference between guidelines on the one hand and legislation and regulation on the other is that the former emphasizes the user's rights first and foremost, the right of the user to access the content, application, services of his choice and to connect the device of his choice.[13] Norway, Denmark, France, and Switzerland also note the user's right to a connection with a declared quality and capacity.

Legislation and regulation have a punitive orientation and emphasize what an operator cannot do (no blocking, no throttling) and what it must do (be transparent and provide information about its offers and practices). A no-blocking rule is present in all net neutrality legislation and in the code of conduct for the UK and Switzerland. Throttling is prohibited in six countries and in the proposed EU legislation solution. Exceptions to blocking and throttling vary but generally include situations of congestion, network security, court order, and emergency.

---

[13] "Remarks of Michael K. Powell Chairman, Federal Communications Commission At the Silicon Flatirons Symposium on 'The Digital Broadband Migration: Toward a Regulatory Regime for the Internet Age,'" *PRESERVING INTERNET FREEDOM: GUIDING PRINCIPLES FOR THE INDUSTRY*, February 8, 2004, https://apps.fcc.gov/edocs_public/attachmatch/DOC-243556A1.pdf.

Traffic shaping is allowed and regulated in Chile, Peru, and Colombia. In Brazil and Chile, operators must notify users before taking action. Operators must deliver blocking at the user's request in Netherlands, Chile, and Colombia.

Chile and Colombia require that operators protect users from viruses, ensure network security, and offer parental controls. Both laws are criticized because they express that "arbitrary" discrimination is not allowed. However it is not clear whether non-arbitrary discrimination is, in fact, allowed.

Zero rating is a current issue in net neutrality, but was not historically part of net neutrality rulemaking. A subsequent legislative pronouncement made zero rating in Chile in 2014. Regulators in Netherlands and Slovenia have updated their rules to include zero rating, but this is being challenged in court by Slovenian operators and may be by Dutch as well. Interestingly Colombia specifically allows operators to make offers tailored to the market needs and segments.

Protecting users' privacy is codified in rules in Chile, Colombia, USA, EU, and Brazil. Privacy advocates criticize the Brazilian approach because it requires operators to retain customer data, creating a set of risks for consumers.[14]

Transparency is codified in different ways from disclosing speeds and qualities of connections (Chile and Colombia), to reduce costs to switch service providers (Sweden), to requiring disclosures on packet loss (USA).

Managed services are allowed in the US and Europe, but not discussed in rules of Latin America. Brazil makes a point to protect operators from intermediary liability. Interconnection is discussed in the rules of France and USA. Paid prioritization, which is vaguely defined, is only prohibited in the USA.

Conclusion

The paper shows how countries use either legislation or guidelines to make net neutrality rules. Legislation allows the authority of telecom regulator to be clarified. Formal guidelines allow the parties a dialogue, which frequently resolves disputes. Moreover it provides low-cost regulation with a credible deterrent for bad behavior. The American approach may reflect a predilection for litigation, but if the goal is to create lasting net neutrality rules, then the approach taken by the FCC is increasingly untenable.

The Dynamic Coalition on Net Neutrality supports legislative solutions in other countries, and there is no reason why this should not be acceptable in the US, especially when the Communications Act was written by Congress in the first place, the very authority that the FCC invokes with its proposed rules. With some seven lawsuits against the FCC, a legislative solution seems the obvious, if not inevitable, way to realize net neutrality rules in the US.

---

[14] 2nd Report for Dynamic Coalition of Net Neutrality, Patricia Vargas-Leon, Chapter 2, accessed April 24, 2015, https://docs.google.com/file/d/0B4CMvT0NORh9RHhKa2IybThhR0U/edit?pli&usp=embed_facebook.

NET NEUTRALITY
AROUND THE WORLD

SOFT LAW/GUIDELINES

	YEAR	RULES	LINK TO RULES IN LOCAL LANGUAGE	LINK TO RULES IN ENG
SWEDEN	2009	Rules	link >	link >
JAPAN	2010	New Competition Promotion Program 2010	link >	link >
NORWAY	2009	Guidelines	link >	link >
FRANCE	2010	10 Proposals	link >	link >
DENMARK	2011	Principles	link >	-
UNITED KINGDOM	2011	Code of Conduct	link >	
SOUTH KOREA	2011	Net neutrality and guidelines for Internet traffic management	link >	-
SWITZERLAND	2014	Code of Conduct	link >	link >

PARLIAMENTARY LEGISLATION

CHILE	2010	“Law that establishes the net neutrality for consumers and Internet users”	link >	-
NETHERLANDS	2011	Update to Communications Act	link >	link >
COLOMBIA	2011	National Development Plan 2010-2014	link >	-
PERU	2012	Law to Promote Broadband and the Construction of the National Fiber Optic Backbone	link >	-
SLOVENIA	2012	Update to Electronic Communications Act	link >	link >
ISRAEL	2013	Neutral Networks	link >	-
BRAZIL	2014	Civil Framework for the Internet	link >	link >
MEXICO	2014	Federal Telecommunications and Broadcasting Law	link >	link >
EUROPEAN UNION	Not yet approved	European single market for electronic communications	link >	

TELECOM REGULATION

CANADA	2009	Review of the Internet traffic management practices of Internet service providers	link >	
UNITED STATES	2015	Open Internet Order	link >	

Roslyn Layton

Current Trends and Controversies in Internet Peering and Transit: Implications for the Future Evolution of the Internet

By
David Reed,[1] Donny Warbritton[2], and Douglas Sicker[3]

Abstract

The hierarchy of Internet peering and transit interconnection agreements between Internet Service Providers (ISPs) defines the core economic structure of the Internet. Recent interconnection agreements (e.g., between Netflix and AT&T, Comcast or Verizon) have generated considerable comment and concern, much of which is misleading as to the root causes for new peering agreements and even the future of the Internet itself. Given this critical setting, this paper examines how and why Internet peering and transit is changing, and discusses the implications of these changes on the future evolution of the Internet. What makes this research novel is the analysis of new Internet traffic data to identify future trends relevant to Internet interconnection. These data provide insight into the congestion level of different interconnects in a few places on the Internet. We next develop a framework to describe past evolution in Internet Interconnection in order to better understand the changing characteristics in Internet traffic, and other new developments impacting the economics of Internet interconnection. This framework demonstrates how most network providers have been progressively expanding their networks over a larger proportion of the end-to-end connection (i.e., increasing their network investment), providing an explanation of the changes in Internet interconnection, and resulting controversies, to keep pace with this network evolution. The strategic implications of our analysis: 1) a more complete understanding of the full network investment undertaken by network providers may help avoid, or at least better explain, the future controversies, 2) rapidly changing nature of Internet content will be constant source of change in interconnection, and 3) the changing role of Tier 1 providers will be further impacted by increasing peering among Tier 2 and Tier 3 providers through the implementation of remote peering based upon open standards for Internet Exchange Providers. Taken as a whole, the increasingly overlapping roles of network providers call into question the utility of the traditional *Tiered* framework as a useful description of the current Internet peering ecosystem,

---

[1] Director, Interdisciplinary Telecommunications Program, University of Colorado at Boulder, and whose support for this research was provided in part through a research gift from Comcast Corporation.

[2] Ph.D. Graduate Student, Computer Science Department, University of Colorado at Boulder.

[3] Department Head, Engineering and Public Policy, Carnegie Mellon University with a joint appointment in the School of Computer Science.

and suggests the need for a new, updated framework to predict and describe future peering relationships.

Section I: Introduction

The models of interconnection and traffic exchange have continued to evolve since the initial creation of the Internet. While it is not the purpose of this paper to provide a detailed description of this entire evolution, it is helpful to understand the history and important trends that are associated with the changes in Internet interconnection over time. We could begin this discussion by describing the ARPANET or discussing the advent of CSNET, which was a networking effort starting in the early 1980s with the intention of bringing the benefits of networked systems to computer science departments that could not connect directly with the ARPANET. However, it might be more useful to begin with NSFNET, as it set in motion the interconnection models that followed.

A few decades ago, traffic exchange on the Internet was across one backbone, the NSFNET. At this point, routing and addressing were fairly straightforward matters and there was little concern for the underlying economics in that NSFNET had an open traffic exchange policy. This backbone was paid for by the U.S. Government and organizations (mostly educational) made use of this resource to interconnect services such as access to supercomputing resources. In 1991, after much debate, NSF removed the restrictions on commercial use of the network, and interesting discussions ensued on the settlement processes that should drive traffic exchange. One early development in this space was the creation of CIX (Commercial Internet Exchange) among a small set of early commercial network providers. This was an alternative model to interconnecting and exchanging traffic across NSFNET. As the Internet (and then the Web) continued to grow in the early 1990s, NSFNET laid plans for discontinuing support of NSFNET. This required other networks to carry the traffic across the network (backbone) and points for interconnecting these networks, which gave rise to the Network Access Points (NAPs) in the mid 1990s. Because routing among a set of networks was more complex than simply pointing to NSFNET, a new routing protocol was required, which resulted in the creation of the Border Gateway Protocol (BGP). The plug was pulled on NSFNET in 1995 and a number of NAPS across the country replaced this infrastructure as a set of interconnection points. At that time, the big players in the backbone space included PSINET, Sprint, MCI, CERFnet, BBN, ANS and AGIS, and the exchange of traffic between these players was based on settlement free peering, which meant traffic was exchanged without any payment between each other.

As the Internet and the Web grow in popularity, the traffic increase became significant at the NAPs and various models of interconnection and traffic exchange developed. One such change was the move toward paid transit service, where a regional ISP would pay a larger national ISP for access to the rest of the Internet (through advertising routes and providing network connectivity). While a pyramid or hierarchy of interconnection did actually exist during the NSFNET days (in that hanging off of the NSFNET there were regional networks that then connected with small institutions), this structure became the norm in the mid to late 1990s (See Claffy and Clark's Physical Representation of Internet Interconnection in Figure 1).

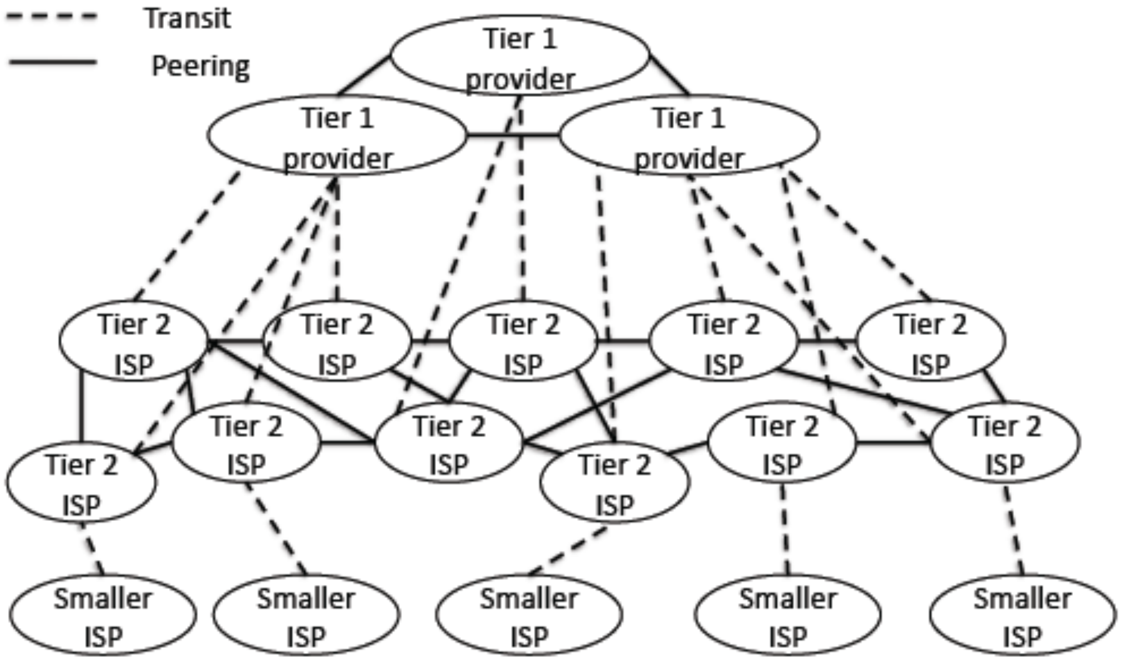


Figure 1: Physical Representation of Internet Interconnection (c. late 1990s)[4]

This pyramid model, where the bottom edge would pay up toward the top, became the norm of the 1990s. At the top of the pyramid were the small set of Tier 1 backbone providers who could reach any network on the Internet without paying transit or paid peering arrangements, below this were the Tier 2 regional networks, and below them the Tier 3 local ISPs and large institutions. Tier 1 providers would provide (for a fee) the lower tiers with transit service to the rest of the Internet and were all interconnected and exchanged traffic as peers in a settlement free manner. These networks had peering agreements with each other based on a set of peering policies. Over time, these agreements and policies became more formal and structured and were generally under NDAs. These peering policies would require (among other things) that the provider: 1) had a presence at all of the major interconnection points; 2) that they carried a sufficient amount of traffic; and 3) that they conformed to common networking best practices.

These policies and hierarchical agreement structure did eventually present problems. For example, Tier 2 ISPs found it very difficult to become a Tier 1 (Level 3 Communications, an infrastructure ISP started in 1997, spent years fighting for this status, even after they could establish that they met the criteria based on the peering policies). Furthermore, transit prices (again, the fee that the lower tiers would pay an upper tier) were high, which some blamed on the market power that the Tier 1s could exert due to the small number of backbone alternatives. In the late 1990s and early 2000, a series of peering disputes arose and some called for regulatory intervention. While this did not occur, one result was that the Tier 1s voluntarily agreed to publish their peering policies (although individual peering agreements would remain under strict NDAs).

---

[4] Source: Claffy, KC and Clark, David D., Platform Models for Sustainable Internet Regulation (August 15, 2013). TPRC 41: The 41st Research Conference on Communication, Information and Internet Policy.

One response to the high cost of transit service was for Tier 2 and Tier 3 ISPs to interconnect or peer directly. This was particularly beneficial to access ISPs and content providers, where they could build infrastructure to one another and avoid the recurring transit costs. This model of interconnection became know as “donut peering”, in that it made a circle around the Tier 1 ISPs and therefore avoided the need to exchange a large fraction of traffic over a transit agreement. This, together with technical innovation and competition from emerging backbone ISPs resulted in decreasing transit prices from the Tier 1. As we shall see later in the paper, this cost avoidance is still playing out among the interconnection disagreements as we see with the recent series of disputes between Netflix and ISPs with access networks. The consequence of this donut peering was that the Internet was becoming more densely interconnected.

Another important change that occurred in the early to mid 2000s was the move toward broadband access. Prior to broadband access, consumers (particularly at home) would access the Internet through dial-up modems running over the telephone network. This model allowed the consumer to pick among a myriad of ISPs and even switch ISPs without much impact. As a more convenient, “always-on” broadband emerged, the provider of the broadband service (the underlying high data rate network connection) increasingly also became the ISP (providing the connection to the Internet and services such as DNS, email, and so on). This meant that the amount of choice between wireline broadband ISPs has been reduced down to the number of physical networks connecting to the consumers’ households, which typically is two in most areas of North America.

As the Internet grew in popularity, the traffic running across the backbones continued to increase at amazing rates. This presented problems in terms of the load on the network and the ability to increase investment to keep up with the demand. A clever solution that emerged was to cache commonly accessed content near the consumer’s access network. This meant that content would not need to travel across the core network each time that it was subsequently requested. An outgrowth of this strategy was for third-party companies, referred to as Content Distribution Networks (CDNs), to host popular content on the behalf of the content provider at points close to the consumer. Akamai and Limelight networks were among the early players in this space but over time, ISPs and content providers started to build out their own infrastructure to provide CDNs. Amazon, Apple, Google, Microsoft, and Netflix are all content providers of sufficient size to have developed their own CDN at this time. This replication of content drastically reduced the load on the backbone and improved the latency associated with the delivery of the content.

This prior history of the evolution of the Internet interconnection sets the stage and context for the development of a modified framework of the physical representation of Internet interconnection in order to help predict future trends in interconnection. To do this, the remaining sections of this paper are organized as follows. Section II examines Internet traffic data to illustrate and identify current issues in interconnection. Section III introduces a modified physical representation of Internet interconnection as originally developed by Claffy and Clark (Figure 1) to gain additional understanding of past and current trends as discussed in our Introduction. Section IV provides a discussion of future interconnection trends based upon knowledge gained from this modified Internet interconnection model.

Section II: Current Traffic Trends on the Internet

There are databases of real-world Internet traffic data available to researchers. This section describes current traffic patterns over the Internet based upon publically available data, primarily as collected and made available by the Cooperative Association for Internet Data Analysis (CAIDA) Archipelago project.[5] We also briefly review and discuss other publicly available data with relevance to Internet interconnection from Netflix, the Measuring Broadband America program by the FCC, and RIPE Atlas.

Analysis of CAIDA Archipelago Project Data

Although tools like Internet speed tester sites and data provided by Netflix and Google about ISP speeds are helpful for evaluating the quality of an end-to-end connection, they obscure many details and may mislead consumers about the reasons for poor performance during a particular use. The complexity of interconnection and the challenge of explaining to consumers the reason they may be experiencing poor streaming video quality despite paying for a fast broadband connection encourages a much simplified story of bandwidth where blame is frequently assigned to a single party.

At this point it behooves us to consider what is meant when content providers and ISPs talk about congestion at some point in the network. Congestion on a data network occurs when too many data packets are being sent over a link resulting in contention among the packets. Such congestion results in delay as the packets are held in a queue, or packet loss if the packet cannot be held until there is room on the link for transmission. From an end-to-end perspective, there is really two components to delay: 1) fixed delays related to propagation, transmission, and processing; and 2) variable delays due to queuing and protocols (such as TCP retransmission). Extended overutilization of some portion(s) of the network fall into the second category, and is the focus of our data analysis. Perhaps more important than the reason for the congestion is its result. The traditional networking view is that congestion occurs when service quality breaks down. This does not mean that the service is blocked completely, but could be an inability to stream music without buffering, or view Internet video at a high quality. Recent speculation about reasons for Internet congestion place much of the blame on the interconnection points throughout the Internet where traffic is exchanged between network operators. By analyzing data made available by the Archipelago project, we hope to provide a clearer understanding of the characteristics of interconnections that have previously been opaque.

The data produced by the Archipelago project is collected by 80 monitoring nodes spread across the globe, 24 of which are located in the US (as of July 2014), performing traceroute[6] measurements to all routed /24s on a continual basis. What this means is that each monitor is randomly assigned a block of addresses to which it is responsible for running a traceroute measurement. A particular /24 IP address block is defined by the first 3 parts of the traditional "dotted quad", the collection of numbers that uniquely define a destination on the Internet. For example, one of Google's IP addresses is 74.125.207.102, which is contained in the

---

[5] The IPv4 Routed /24 AS Links Dataset – 1/3/2014-7/25/2014 http://www.caida.org/data/active/ipv4_routed_topology_aslinks_dataset.xml.

[6] An application used for mapping the path to a given destination that reports all intermediate steps and the round trip time to each step.

74.125.207.0/24 block. By working together, the entire Internet can be measured in a short period of time, and the repeated measurements are stored for later analysis.

Although the ultimate destination of the majority of these measurements is of little interest to us at this time (due to their random nature), the path traversed between monitor and destination along with the latency at each intermediate step can be very instructive. Because of the way that networks are connected together, some paths are part of a large number of measurements. This is analogous to driving to a series of random locations in your town where each journey starts in your garage. Regardless of the ultimate destination, the first step will always be the driveway and after that the street running in front of your house. Each successive turn is an opportunity to deviate from previously driven routes, but the initial steps will be repeated many times over. By looking for frequently used paths of traffic through the Internet, we can discern repeating patterns where unexpected delays or high latency between steps may indicate times of congestion in intermediate links.

Due to the large number of monitors and the continual data collection work they perform, the corpus of information that could be parsed is dauntingly large. For this reason we focused our efforts on studying the types of interconnection between ISPs that have become newsworthy of late – that is, connections in the U.S. between ISPs with access networks like Comcast, and backbone ISPs like Level 3. Focusing our efforts on these interconnects solves two important issues. First, because destinations are random and only scanned once per cycle (roughly once per day), there is not enough end-to-end data between a monitor and any given destination for useful analysis. Second, we want to avoid lumping several factors together and attempting to draw conclusions based on coarse inputs.

By using publically available *whois*[7] data, it is possible to identify the network operators at each step of the traceroute path. Each network, or Autonomous Systems (AS) is logically grouped together by the entity by which it is operated. In many cases there will be several steps with potentially very different IP addresses, and discussing portions of the route according to AS lends clarity to the discussion. As it turns out, it is not necessary to explicitly work to identify transitions between ASes as a precursor to finding interesting data. While the majority of distinct IP addresses show up a relatively small number of times, a much smaller set dominate the data in terms of raw occurrence. The size of this set can be further reduced by only considering intermediate destinations, hops, whose incremental round trip time (RTT) value is above some threshold.[8] In this case, by ranking the most frequently occurring hops that had an incremental RTT of 80 or more milliseconds, we were able to find a frequently occurring link going between Comcast and Level 3. The existence of this link is unsurprising; Level 3 carries data for a large number of entities, and at some point they have to send that data onto Comcast so it can be delivered to customers and *vice versa*.

This link is of particular interest because the latency between the two networks typically dominates the total RTT for the trace. Figure 2 shows the percentage of the total RTT contributed by the link between Comcast and Level 3, along with the total contribution of all the links before and after. We generate the graph by taking the rolling average of percent

---

[7] A tool used for querying databases containing registered assignees of an Internet resource.

[8] RTT, or round trip time, is the time it takes for a packet to make a round trip between source and destination. It is a measurement of propagation delay.

contribution over a roughly 1-hour window over the course of about two weeks. This demonstrates not only the large proportion of total latency this link contributes, but also a strong diurnal trend where the peak lasts for a large portion of the day. A tempting conclusion from this information is that these peaks represent congestion between the two parties. While this may be the case, we caution against assuming it must be the case – traceroute RTT data is not the same as directly measuring IP traffic or streaming video, and it is possible traceroute packets are deprioritized at some interfaces when utilization reaches a certain threshold, or that the path taken by packets in one direction is different then the return. Nevertheless, we believe the data is not without insight.

The dashed lines are at 45% and 55% of the total RTT and indicate different potential thresholds for considering an interconnection to be experiencing some level of congestion. The percentage of time the interconnection line spends above those thresholds is 93.3% and 49.2% respectively. This means that more than half of the total delay for sending a packet across the Internet from this particular node can be ascribed to this one connection for nearly half of the day. For reference, Figure 3 provides a graph over the same time period showing the absolute RTT values on the y-axis.

As noted above, it is difficult to say if this truly represents a state of network congestion. Whether or not a user would experience a lower service quality will depend upon the service in use. Figure 3 suggests the average RTT is roughly 200 ms, a latency figure that might result in service problems for VoIP, but not for video streaming which has the benefit of buffers to mitigate against latency in the network.

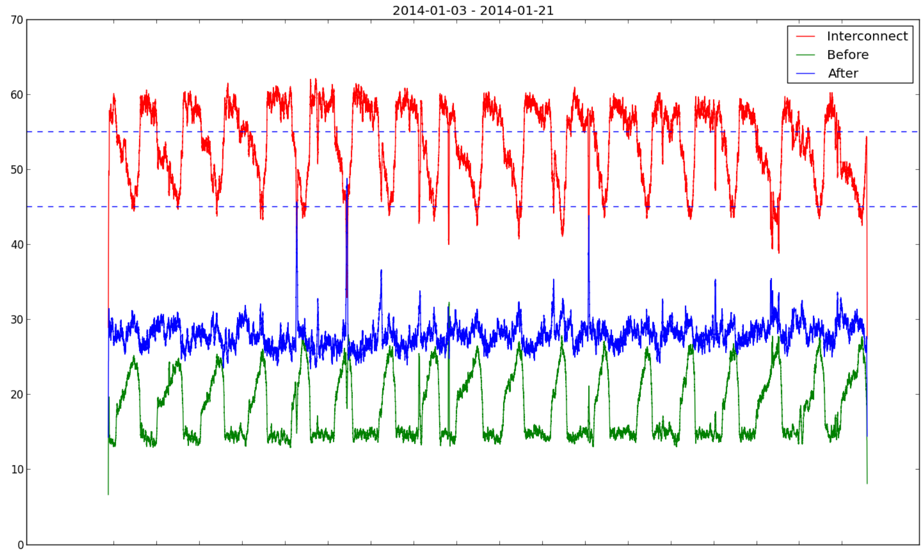


Figure 2: Percentage of total RTT contributed by interconnection link, and sum of the links before and after, using CAIDA data from early January. Time is in 24-hour increments along the X-axis.

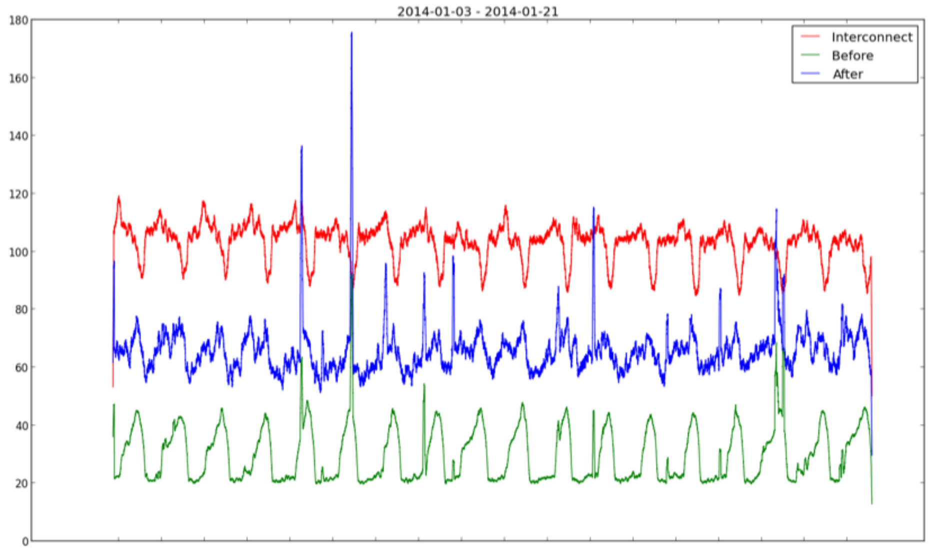


Figure 3: RTT in milliseconds for the interconnection link along with the sum of the RTT for links before and after. Time is in 24-hour increments along the X-axis.

Moreover, the potential congestion in this link appears to vary over time as well. Figure 4 shows the percentage of total RTT from the Comcast/Level3 link from mid July rather than early January. In this case the percentage of total RTT that the link contributes has dropped and the diurnal effects are less pronounced. It is possible that any congestion existing at this link has been reduced, although it is unclear the magnitude to which this may be. In this case the same candidate congestion threshold lines show the interconnect RTT contributing to 45% and 55% of the total RTT at 99.64% and 13.24% respectively. While the January 2014 data did appear to show congestion between Level 3 and Comcast, it is important to note that Comcast and Level 3 had "resolved their prior interconnect dispute on mutually satisfactory terms" in July 2013.[9] Comcast engineers inform us that Level 3 capacity issues in January were due "to unexpected major changes in traffic with a customer of Level 3". While adaptive content delivery can cause, or avoid congestion in real-time, it is more difficult for ISPs to react to large, unexpected capacity changes with content.[10] Given this picture, it would be useful to better understand how these agreements evolve over time and all that is involved in upgrading connections (and all that is associated with changing interconnections). One interesting question has to do with the time it takes to resolve congestion and how action is tied to business arrangements.

We also analyzed a similar interconnection between Verizon and Level 3 as shown in Figure 5. As with the connection between Comcast and Level 3, a strong diurnal pattern of somewhat

---

[9] See Drew Fitzgerald, "Level 3, Comcast Reach Accord on Internet Traffic Costs", *The Wall Street Journal*, July 16, 2013.

[10] Adaptive or dynamic routing technology allows system operators to dynamically alter the path of traffic based upon changing circumstances, thus allowing content providers like Netflix substantial flexibility in changing the different traffic routes of the carried traffic over the network mesh.

flattened peaks is present indicating that packets are potentially being queued due to limited capacity.

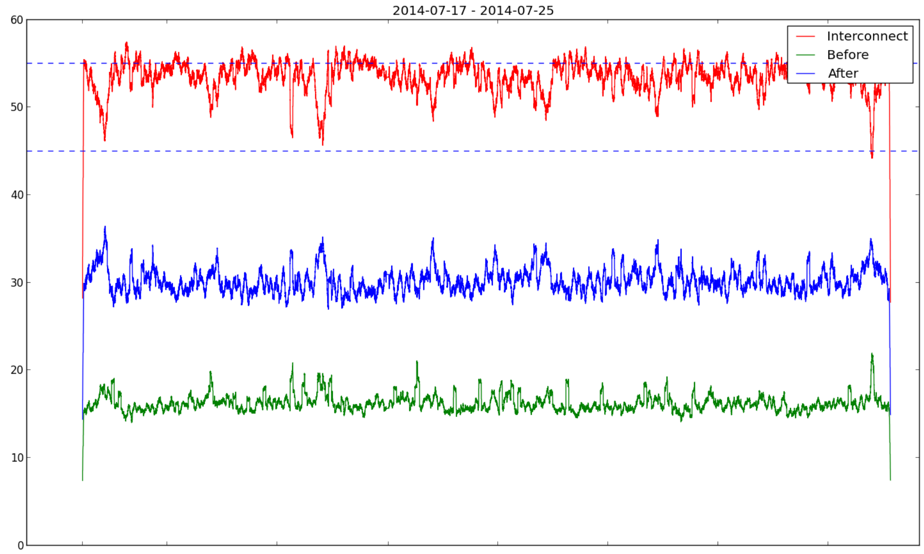


Figure 4 –Percentage contribution to total RTT, data from mid-July. Time is in 24-hour increments along the X-axis.

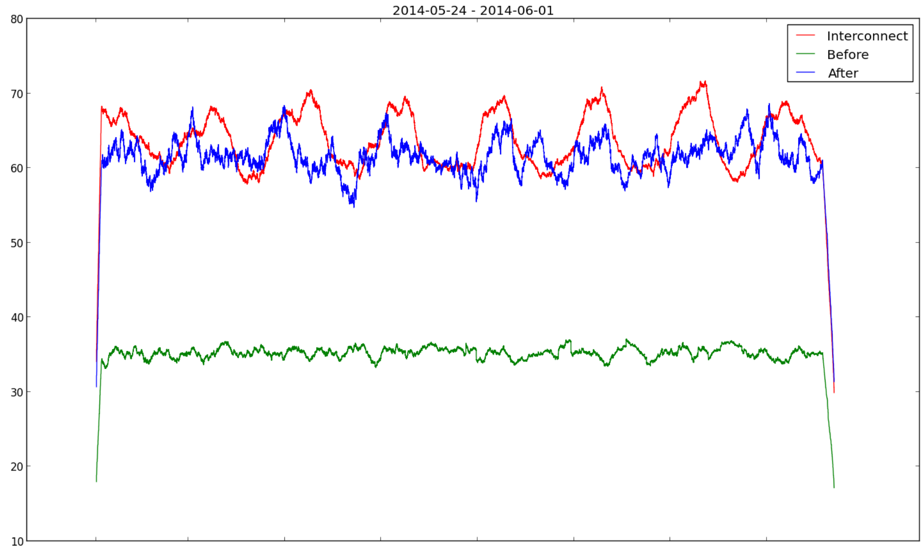


Figure 5: RTT in milliseconds for the interconnection link between Verizon and Level 3 along with the sum of the RTT for links before and after. Time is in 24-hour increments along the X-axis.

In order to understand if this pattern is representative, we have also analyzed a similar interconnection that occurs between Iowa State University and Level 3. Figure 6 shows RTT values for the sum from each hop before this interconnect, after the interconnect link, and the RTT for the interconnect itself. Not only is the interconnect portion of the trip a much smaller percentage of the overall, but the RTT values themselves are significantly lower than the Comcast/Level3 link as well. What is going on here? We do not have direct knowledge. Many universities act as small ISPs, however, and it is reasonable to believe that Iowa State is purchasing transit from Level 3 in order to reach the rest of the Internet. As such, the link looks well dimensioned for its load (i.e., there are no “flat tops” indicating periods of congestion where the link is fully saturated).

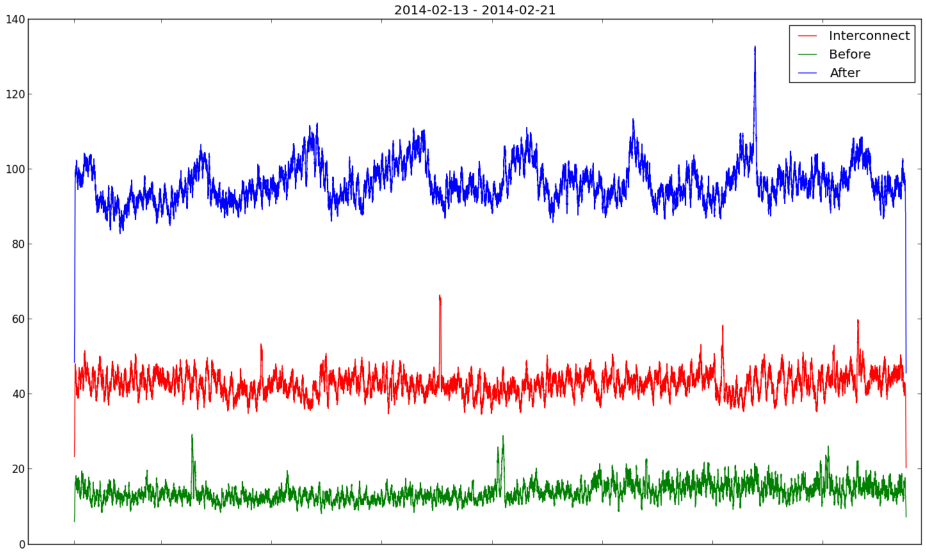


Figure 6: RTT in milliseconds for interconnection link along with the sum of the RTTs before and after. Time is in 24-hour increments along the X-axis. Interconnect is between Iowa State University and Level 3

We have two other reactions to these data worth noting. First, we cannot judge the overall level of congestion on the Internet given the random nature of the IP addresses to which traceroutes are available. Traffic on the Internet is concentrated to a relatively small number of IP addresses of the most popular content providers and is not uniformly distributed across all IP addresses. So, for example, the data can tell us that for one day in July, the percentage of all hops over 100, 200, and 300 ms was 4.11%, 0.71%, and 0.22%, respectively, and these percentages were relatively constant throughout the entire day. This would appear to be a small number and hence indicate little congestion on the Internet. There is no way to know, however, whether the random distribution of destinations may avoid the congested links associated with the most popular websites, those that cause the most significant amount of consumer dissatisfaction.[11]

---

[11] Consider the conservative proposition that 80% of the traffic is generated by 10 or 20% of websites, then a random distribution of traceroutes across all IP addresses will result in traceroute measurements from popular sites generating 80% of Internet traffic only 10 to 20% of the time. Instead, 80% of the traceroute data will come from websites generating only 20% of the data. This bias could cause the data to underestimate the presence of Internet congestion.

The fact that the overall proportion of high-latency hops remains relatively constant throughout the day, including the peak usage period, may indicate little congestion on the Internet during this time since we would anticipate this proportion to increase in the presence of a substantial percentage of saturated links. Figure 7 plots the average end-to-end link latency for the same day in July for all the U.S. data. Now the peak usage period can be seen in the data, corresponding to roughly 18 hours of the day.

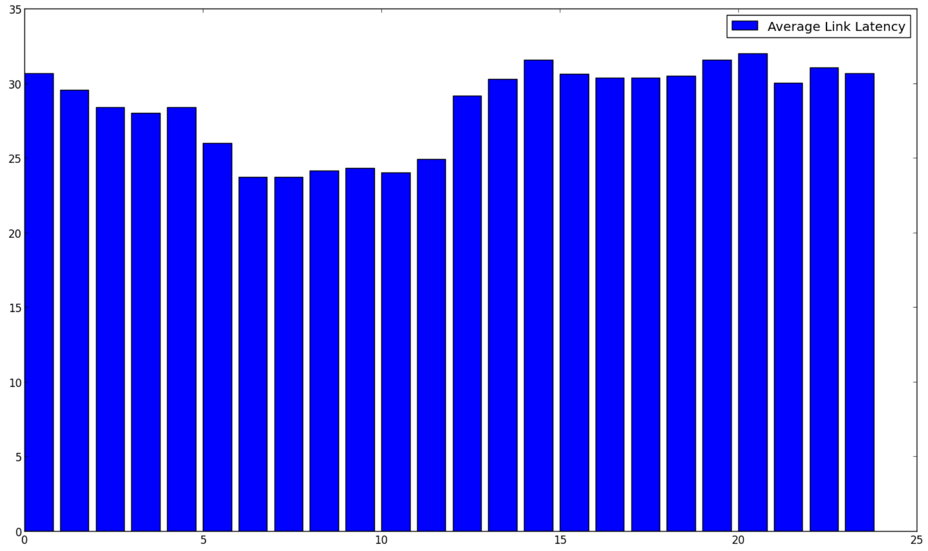


Figure 7: Average End-to-End Link Latency in Milliseconds for One Day in July for All U.S. Data

Second, typically each side of a peering interconnect is responsible for maintaining their equipment to keep peak utilization below a certain threshold. Based upon the CAIDA data itself, we cannot know whether this link is congested because Comcast, Level3, or both have chosen to not upgrade the capacity of the link. On its company blog site, Level 3 complains that some "broadband consumer networks" peers – what we call ISPs with access networks in this paper – are not attempting to upgrade peering capacity.[12] In support of this position, Level 3 provides on their blog a picture of some traffic data from a 100Gbps interconnect in Dallas with a broadband consumer network showing severe congestion, though the other network provider is not identified. The conventional explanation for the "permanent" congestion on these peering links is that the ISP with access networks is seeking to change the peering agreement from settlement free to paid peering. Once a paid peering arrangement is established, as was the case with the

---

[12] In a company blog post this year, Level 3 reports having 51 peers in 45 cities connecting over 1,360 10 Gigabit Ethernet ports (plus a few smaller ports). Individual peers range from a single port to 148 ports. The average number of interconnection cities per peer is five, but ranges from one to 20. The average utilization across all those interconnected ports is 36 percent. Level 3 says its interconnection with six broadband consumer network peers is permanently congested at 90 percent utilization because "our peer refuses to augment capacity".
See Mark Taylor, Observations of an Internet Middleman, May 5, 2014, at
http://blog.level3.com/global-connectivity/observations-internet-middleman/

Comcast/Netflix agreement in February, then additional capacity is added and the network congestion is resolved, as we will see below.

Overall, our analysis indicates is a rather liquid market and changes are occurring in terms of traffic flow, costs and implications on congestion. To understand this interconnection ecosystem, we need to observe the changes and the stress points that are arising. We hope to do this across the Internet in further work, as there are interesting questions that are arising in the connection between content and access networks.

Other Public Data Sets

There are other public sources of data for Internet traffic that have implications for peering arrangements.

Netflix ISP Index (http://ispspeedindex.netflix.com/).
This index lists speeds that reflect "the average performance during prime time of all Netflix streams on each ISP's network."[13] The index moved into the spotlight when it showed an improvement in Comcast's speed from 1.68 Mbps to 2.5 Mbps between February and March of 2014 after Netflix and Comcast agreed to a direct connection agreement in February. Netflix has since expressed public regret for entering into this business arrangement,[14] though it has also agreed to subsequent paid peering arrangements with AT&T and Verizon. The performance speed of Verizon has yet to improve post agreement like it did for Comcast, and it is too soon after the late-July agreement with AT&T to see any change in performance speed as well. The Federal Communications Commission (FCC) announced in June 2014 that it is reviewing these agreements to address consumer concerns regarding sources of Internet congestions impacting online video service quality.[15]

With regard to Comcast, the agreement and subsequent improvement in performance over the Netflix ISP Index implies there was congestion in interconnection links prior to the new peering agreement. Likewise, Verizon announced to its customers that the source of Netflix service quality issues was due to congestion on Internet connection links.[16]

M-Lab (http://www.measurementlab.net/)

---

[13] Average performance is below peak performance due to factors including the variety of encodes Netflix uses to deliver content, the variety of devices customers use, and home network conditions. Netflix contends that these factors cancel out when comparing across ISPs.

[14] See Netflix company blog, Reed Hastings, "Internet Tolls And The Case For Strong Net Neutrality" March 20, 2014 at http://blog.netflix.com/2014/03/internet-tolls-and-case-for-strong-net.html

[15] See Chairman Statement on Broadband Consumers and Internet Congestion, June 13, 2014, at http://www.fcc.gov/document/chairman-statement-broadband-consumers-and-internet-congestion

[16] See http://publicpolicy.verizon.com/blog/entry/why-is-etflix-buffering-dispelling-the-congestion-myth stating: "There was, however, congestion at the interconnection link to the edge of our network (the border router) used by the transit providers chosen by Netflix to deliver video traffic to Verizon's network." See also: "Netflix is responsible for either using connections that can carry the volume of traffic it is sending, or working out arrangements with its suppliers so they can handle the volumes."

M-Lab is a consortium of research, industry, and public interest partners developing a new ecosystem for the "open, verifiable measurement of global network performance". M-Lab provides researchers access to network servers to operate measurement tools to assess network performance throughout the public Internet. A recent M-Lab report and blog posting identified the use of quality of service by transit provider Cogent as a possible explanation for poor performance where Cogent interconnected with Comcast, with Time Warner Cable, and Verizon as well.[17] M-Lab measurements detected that Cogent was giving some traffic flows higher priority over others, resulting in the significant slowing of Netflix traffic at each these interconnection points.[18]

FCC Measuring Broadband America Program (http://www.fcc.gov/reports/measuring-broadband-america-2014)

The FCC publishes a Measuring Broadband America Report on Fixed Broadband on an annual basis. Each report contains data collected from fixed (wireline) ISPs regarding how close their sustained download speeds in the access network are to the advertised speeds. The 2014 report found that ISPs provided on average 101 percent of advertised speeds based upon measurements conducted during a peak usage period defined as weeknights 7 – 11 pm. The upshot of these findings with regard to Internet interconnection is that the sustained speed offered in the access network by ISPs is not likely to be a bottleneck or source of congestion assuming the advertised speeds of the service are sufficiently high to deliver the services requested by the consumer. More precisely, the end-to-end data rate of an Internet connection is a complex series of interconnections between different network segments (which we will introduce in the next Section). The access network is but one network component in the end-to-end connection. The FCC data indicates that the access network is not a source of congestion (at least in terms of providing a data rate during the peak usage period that is below the advertised data rate).[19]

RIPE Atlas (https://atlas.ripe.net/),

Réseaux IP Européens Network Coordination Centre (RIPE NCC) is an independent, not-for-profit membership organization that operates the RIPE Atlas project. This effort employs a global network of over 6000 probes that measure Internet connectivity and reachability. As the version of this paper is a working draft, we plan to analyze these data in the near future to see what information it may hold for the discussion of the evolution of Internet interconnection.

---

[17] See Measurement Lab Consortium Technical Report, "ISP Interconnection and its Impact on Consumer Internet Performance," Release 1.0, October 28, 2014. Accessed at http://www.measurementlab.net/static/observatory/M-Lab_Interconnection_Study_US.pdf. See also ," M-Lab blog posting by Colin Anderson, "Research Updates: Beginning to Observe Network Management Practices as a Third Party, October 31, 2014. Accessed at http://www.measurementlab.net/blog?date=october%202014

[18] See Dan Rayburn, "Cogent Now Admits They Slowed Down Netflix's Traffic, Creating A Fast Lane & Slow Lane," November 5, 2014. Accessed at http://blog.streamingmedia.com/2014/11/cogent-now-admits-slowed-netflixs-traffic-creating-fast-lane-slow-lane.html.

[19] The FCC indicates on its website that its data did identify congested interconnection points. The FCC chose, however, to "rely upon data from unaffected servers" in reporting performance results in its report. The FCC did collect "test results from impacted servers and are releasing this data as part of our reporting process for use by academics and others in examination of this issue". We have yet to request to examine the FCC data in the preparation of this report. See http://www.fcc.gov/reports/measuring-broadband-america-2014.

Section III: Internet Interconnection Framework

In this section of the paper we develop a qualitative framework of Internet peering and transit in order to gain a better understanding of the interdisciplinary tradeoffs inherent to Internet interconnection. There has been a substantial amount of prior high-quality research and analysis published on Internet interconnection by numerous academic and industry authors.[20] Our framework builds upon the physical representation of Internet interconnection presented by Claffy and Clark as shown in Figure 1. What is novel about our framework is the addition of the different network segments of an end-to-end Internet connection into the representation.

Elements of a Internet Interconnection Framework

Fundamentally, peering and transit agreements define the business relationships between the two networks providers that are transferring Internet traffic between one another. The technology defines the actual costs of the transfer, while the larger Internet transport market defines the business position of each player.

Throughout this paper, we assume the fundamental interoperability of the Internet remains intact. That is, the technical architecture of end-to-end Internet transport supports multiple, standalone networks operating together as a federation of networks without significant interoperability concerns in the interconnection of these networks. This is the case today, and we know of no pending developments that threaten Internet interconnection from this technical point of view. Figure 8 illustrates a simplified view of the network segments that collectively define the traditional view of the end-to-end Internet network. This picture illustrates the typical network boundaries between network segments common on the Internet.[21] With the exception of the interface between the end user and the local access ISP, all of the remaining interfaces between network segments are the network interconnection points where peering and transit agreements are required – or business relationships amongst transport providers need to be defined.

Figure 8: End-to-End Internet Transport System

---

[20] Notable recent contributions in this area include Claffy, KC and Clark, David D., Platform Models for Sustainable Internet Regulation (August 15, 2013). TPRC 41: The 41st Research Conference on Communication, Information and Internet Policy; see also William B. Norton, The Internet Peering Playbook: Connecting to the Core of the Internet, 2014 Edition.

[21] We are not aware of precise, accurate definitions of each network segment. This physical representation of the network, however, is useful for describing general trends impacting the Internet interconnection model.

Building upon this network model and Claffy and Clark's physical representation, Figure 9 shows the location of the typical peering and transit agreements between network operators for the Internet network around the year 2000. At this point in time, Tier 2 providers were early in building their networks, and generally purchased transit from Tier 1 backbone providers. Content providers purchased transit from either backbone ISPs or third-party CDN providers (using transport leased from national backbone ISPs) to efficiently distribute and deliver their content to the Internet. On the other end of the connection, campus or local ISPs purchase transit from access network ISPs for Internet connection. This fairly uniform interconnection model was widely employed by the majority of Internet transport providers during this time.

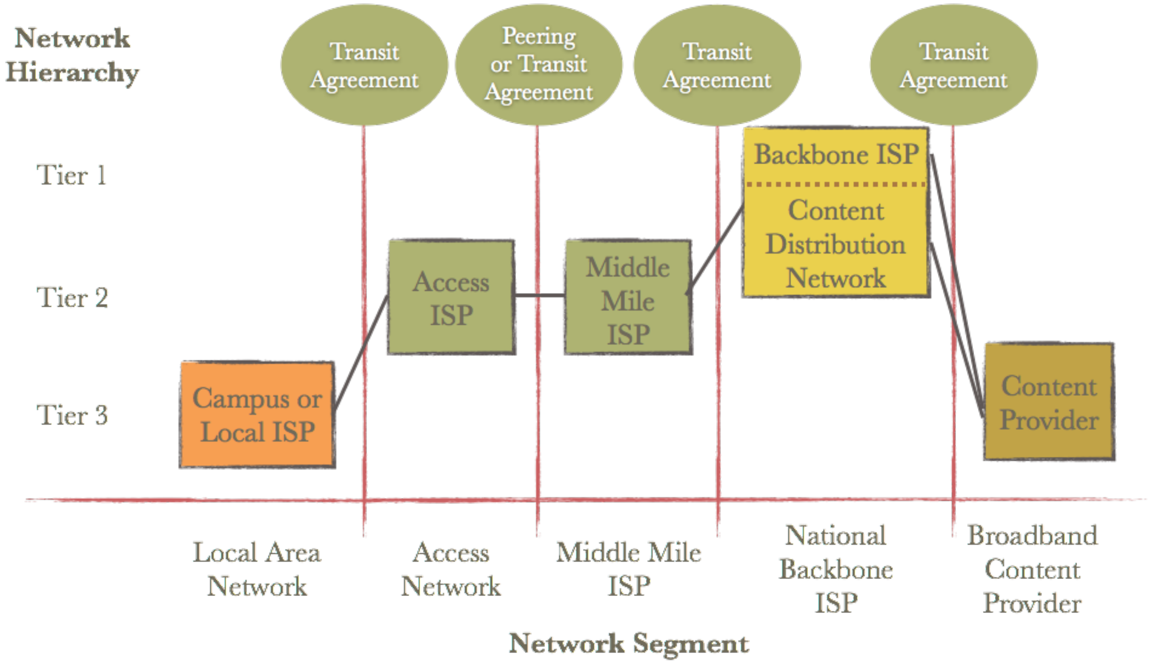


Figure 9: Internet Interconnection Model (c. 2000)

Figure 10 shows the evolution of the interconnection model in year 2010 by illustrating how the different network providers have grown as the Internet has matured over the next 10 years. In general, most of the network providers have expanded to provide transport over a wider range of network segments in the core network. Backbone ISPs have expanded into regional areas and provide CDN services to content providers as well as leasing facilities to third-party CDN providers. ISPs with access networks have built their own regional and national backbone networks. More campus and local ISPs are connecting to access network ISPs or backbone ISPs using transit agreements in areas where such direct interconnection is possible.

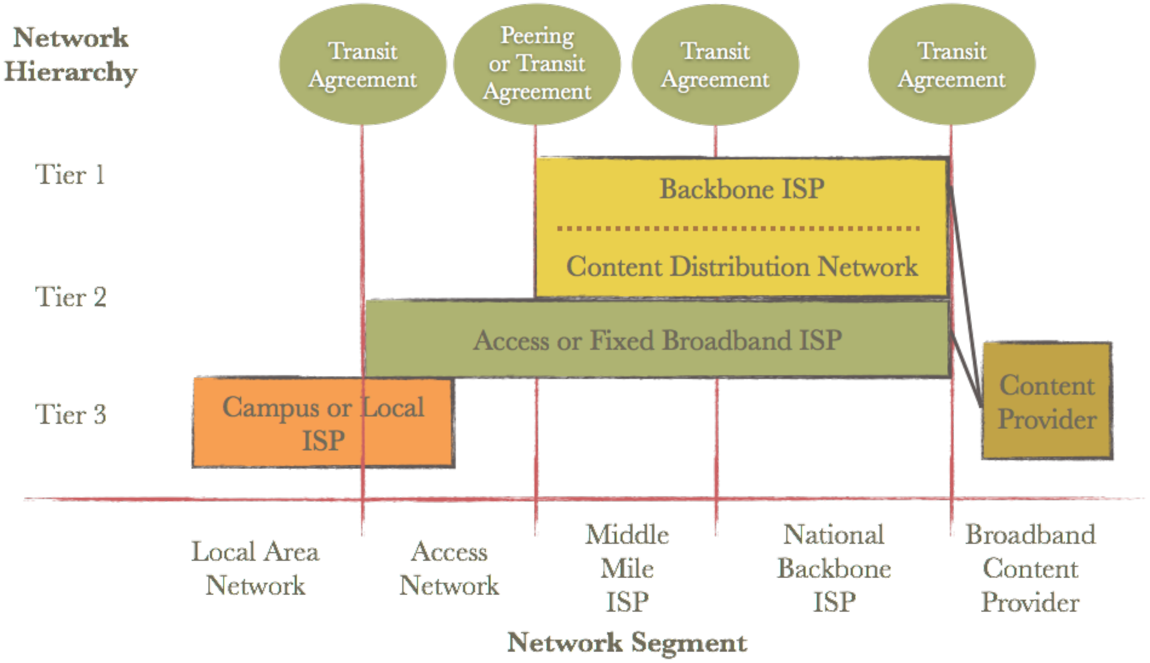


Figure 10: Internet Interconnection Model c. 2010

Figure 11 shows the continuing evolution since 2010 until today. Now, large content providers have built and operate their own CDNs which span the entire Internet core network segments extend deeper into the access network, and paid peering arrangements to ISPs with access networks from backbone ISPs and content providers are becoming the norm. To be sure, the content provider CDNs are not the same as ISP networks, as they are customized to only fit the traffic needs of the content provider and not a host of transport customers which the ISP networks are designed to accommodate.[22] This simplifies the interconnection picture and allows the large content providers to better control their user experience by reducing the number of interconnects to only the hand-off between the content provider and the access network. Looking at the progression of Internet interconnection from 2000 until today, the role of Internet interconnection can be seen as holding together an emerging lamination of network providers, or uniting the superimposed set of networks running across similar and diverse network segments. The modified physical implementation of this evolution clearly shows this trend.

---

[22] The design, operation and business case of the various CDN providers (be that a third-party provider, a content provider or an ISP) thus may differ. The nature of these differences and how they drive interconnection is another interesting question to consider.

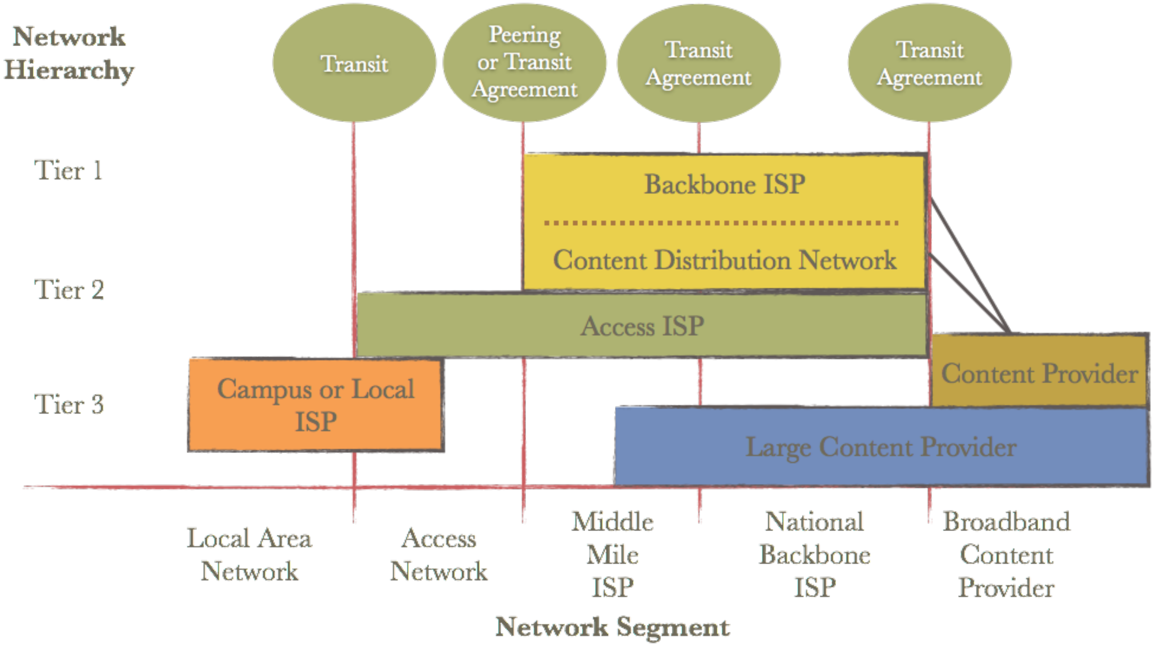


Figure 11: Internet Interconnection Model c. 2014

What are the key drivers leading to this dynamic environment and significant changes in Internet interconnection over the past 15 years? We create two categories to describe the new interconnection conditions and hence different business relationships for interconnection among Internet transport providers. These two major categories driving and shaping current and future Internet interconnection are: 1) Changing Characteristics of Traffic, and 2) New Developments Changing the Economics of Peering. We explain each of these categories in more depth in the remainder of this section.

Changing Characteristics of Internet Traffic

As with any network interconnection agreement, a primary consideration is obviously the nature of the network traffic to be exchanged between the two network providers. The important characteristics of Internet traffic include its destination, volume, symmetry, peak usage duration, and emergence of adaptive streaming. Below we discuss each of these topics and highlight the key change occurring at the current time on the Internet.

Destination

In terms of determining the other networks to which network interconnection is needed, the most important characteristic of the Internet traffic is the distribution of possible destinations across the Internet. ISP customers generate and request traffic to/from a number of different Internet addresses associated with web sites or servers. The location of these websites establishes the different networks to which the traffic is directed or received. An ISP can break down the percentage of the traffic it carries by the destination and source networks. An important driver impacting the distribution of traffic across different network destinations is the network segment of the ISPs –ISPs with access networks have large numbers of subscribers spread over large geographic areas, for example, and are frequently a popular source or destination for Internet traffic as opposed to local or regional ISPs which have a much more localized impact on traffic distribution.

> *Key Change*: The ISP with access network segments is increasingly becoming an important destination and source of Internet traffic due to the growing popularity and necessity of broadband service to commercial and residential consumers.

Volume

Once the distribution of traffic across other networks is known, the next most important traffic characteristic is the volume of traffic being carried to and from the other networks. When the traffic exchanged between networks exceeds some threshold, it merits further analysis as to whether a peering relationship can deliver enough value to merit an agreement. The volume of Internet traffic is highly sensitive to the applications being carried over the Internet. Broadband applications that use lots of bandwidth and are popular, such as video streaming, can cause significant and rapid changes in the total amount of traffic flowing over the network. Internet traffic is forecast to increase at a compound annual growth rate of 21% for the next five years, over 30% in the peak usage period.[23] With Netflix leading the charge, video streaming now ranks as the top traffic application on the Internet.[24] The adaptive protocol employed by video streaming services further compounds the amount of traffic generated by these services through automated adaption of the streaming speeds to consume unused bandwidth on the network. Such rapid growth due to the dynamic nature of broadband applications places a priority on frequent monitoring of the interconnection point to detect and address prolonged periods of congestion.

> *Key Change*: Internet traffic growth is significant and highly dynamic due to the rapid adoption of new broadband applications. The emergence of online video services to residential broadband subscribers is the major broadband application driving significant increases in traffic and new traffic patterns throughout the core Internet.

Symmetry

Symmetry represents the ratio between the traffic delivered and received between each network provider. This is an important traffic characteristic as it has historically been used to determine whether a peering agreement can suffice (in the case of symmetry ratios below 1.5 to 2.0 or lower) or a transit agreement is necessary. Over time, symmetry may have an even larger impact on the overall framework for Internet interconnection. Residential broadband traffic is inherently asymmetric with average usage over access networks usually delivering an average more traffic to the customer than from the household. Growth in the popularity of video streaming services will serve to further increase the asymmetry of traffic over local access networks. Consequently, if the traditional rules for peering regarding symmetry are applied in the future, then ISPs with access networks can be expected to charge for interconnection to their networks by other Tier 1 and Tier 2 ISPs since the symmetry of this traffic will be inherently more traffic in the downstream as compared to the upstream direction.

---

[23] See Cisco Visual Networking Index Report on Global IP Traffic, 2013-2018 at http://www.cisco.com/c/en/us/solutions/collateral/service-provider/ip-ngn-ip-next-generation-network/white_paper_c11-481360.html

[24] Netflix estimates that it now generates 33 percent of all traffic on the Internet, a staggering multiple Terabits of video streams during peak periods. See David Fullagar, Netflix Open Connect, presentation at Content Delivery Summit, May 2014, at http://conferences.infotoday.com/documents/197/2014CDNSummit-Netflix.pdf

> *Key Change*: The asymmetry of traffic flowing over access networks will continue and potentially become even higher with the growing popularity of streaming video. This may lead to more paid peering interconnection agreements to ISPs with access networks.

Peak Usage
Internet networks are built to deliver a specific capacity of data traffic. With accurate understanding of the volume of traffic intended to and from each destination, it is possible to design your network. When the capacity of the network is over-utilized, network congestion occurs. The peak usage period represents the amount of time that network traffic consumes a significant percentage of the existing network capacity (e.g., 70 percent). Our analysis of Internet traffic discussed earlier showed that the peak usage period of the Internet is now a significant percentage of each day. The implication of such a long peak usage period is that links that are saturated with traffic will experience congestion for long periods of time each day, resulting in higher latency in the delivery of traffic.

> *Key Change*: The peak usage period for the Internet now occurs for most of the day and late in the evening of most days. This broad peak usage period means that over-utilized links will result in congestion, with its associated negative performance implications, for long periods of time each day.

Impact of Adaptive Streaming Protocols on Network Congestion
Online video is rapidly growing, and most streaming video services use adaptive streaming protocols. Adaptive streaming senses the available capacity in the connection between the client and server, and constantly adjusts the streaming speed upward until either the ceiling for speed on the link has been achieved, or the highest encoding version of the content is being played. Consequently, adaptive streaming can lead to aggressive consumption of network capacity.

Adaptive streaming is not the only protocol on the Internet that senses the presence of additional capacity available in the connection. TCP typically follows a "slow start" that ramps up the connection speed to a target speed. If packet loss occurs, then connection speed slows down. Short-lived TCP connections therefore operate at a disadvantage on busy links relative to longer-lived TCP connections typically associated with video. Additional consumption of capacity by adaptive streaming services like Netflix comes at the expense of other services on the busy link, slowing web surfing, gaming and VoIP traffic.[25]

In short, the surge in popularity of online video, the use of adaptive streaming to distribute the content, and the asymmetric nature of the traffic present a "perfect storm" of sorts for the prior Internet interconnection regime. All the prior peering agreement have to adjust to reflect the new realities presented by the changed traffic flows on the Internet, all in a relative short period of time.

> *Key Change*: Adaptive streaming can consume additional network capacity at the expense of other network services during periods of congestion. Capacity planning for interconnection is complicated by this situation.

---

[25] See Global Internet Phenomena Report, "Exposing the Technical and Commercial Factors Underlying Internet Quality of Experience," Sandvine, September 2013.

New Developments Changing the Economics of Peering

In earlier sections we have already covered the significant historical trends and current events in Internet interconnection. We now discuss new developments that are changing the economics of peering, both now and for the near future. Some of these developments have been touched upon earlier; some are new. All have the potential in our view of leading to substantially different network economics with the likely outcome of changing the business relationships among Internet transport providers and hence the economics of peering.

The general economic principle most often applied to describe peering agreements is that they occur wherever they deliver "value" to both parties. Presumably this means that the cost benefit analysis of the peering relationship delivers more benefits than costs for both parties. At its simplest, this means the costs of both the interconnection equipment and monthly fees for housing this equipment at an Internet Exchange Point (IXP) are less than the cost of transit for the traffic that would be carried over the interconnection point. In some circumstances, however, the cost benefit calculus can be considerably more complex as some providers will trade-off the loss of transit fees, for example, for a new settlement free peering arrangement with the addition of other transport services that might already exist in the business relationship, or assign additional value to a strong bargaining position in a particular transport market (such as the access network segment or international segment).

Given that such a value equation has served to establish peering relationships in the past, we now turn to discuss some new developments with the potential to change the economics of peering going forward in the future.

Interconnection Costs

The role of IXPs in Internet interconnection is critical as their presence defines the geographic location where peering and transit agreements can occur, as well as helping to reduce the cost of interconnection by reducing the number of locations that networks need to reach in order to connect together. Over the past decade the number of IXPs has increased significantly to 2000 in North America.[26] But some of the IXPs have more ISPs connecting together than others, and "neutrality" at some of the IXPs has emerged recently as a concern as well. Some IXPs not only offer routing and cross-connect infrastructure but also data center and colocation services, often on an exclusive basis in the IXP and therefore with higher profit margins. ISPs that utilize an IXP can face very high switching costs to move to another IXP location, something which IXPs take advantage of by charging high prices for these other services (sometimes even go so far as to offer the routing services for no charge). Technical standards to support "Open" IXPs are being developed in North America in order to address the neutrality concern and generate more competitive services environment.[27]

With the emergence of more IXPs there has been interest in developing an "IXP grid" or "remote peering" that would permit an ISP to connect to one IXP, but peer with another ISP in a different

---

[26] The European Internet Exchange Association (Euro-IX) lists 2159 IXP participants in their database. See https://www.euro-ix.net/tools/asn_search. Europe has over 6000 IXPs in comparison.

[27] Open-IX is a membership organization establishing open standards for data centers and IXPs. See http://www.open-ix.org/.

IXP without having to install interconnection facilities at this second ISP. In this fashion, IXPs are evolving to look like networks rather than just single points of interconnection.[28] Remote peering would require the IXPs to interconnect with each other in a network grid, or for an IXP to expand to multiple locations under the same brand. This would be a benefit to Tier 2 and Tier 3 ISPs who have limited geographic scope since it would allow them to peer with other Tier 3 ISPs at lower cost by avoiding having to install equipment at every ISP.

> *Key Change*: Expect more IXPs to connect among themselves in a grid and to support "open" standards. This will allow more Tier 3 ISPs to peer with each other on a settlement free basis.

Network Investment

What is changing are peering arrangements whose terms are shifting to also reflect the investment in the network capacity required to carry the traffic to its eventual destination rather than solely the amount of traffic passing through the peering point. The full "network cost burden" for carrying the traffic to its destination changes across the different network segments of the end-to-end Internet transit path. The cost per mile for installing a backbone fiber optic network will usually be more expensive than the installation cost per mile of a fiber cable in the local access network. But the backbone fiber cable might hold 96 fiber lines, while the middle mile only carries 12 to a specific region. The resulting total cost per Gbps of transport therefore might be much lower in the backbone than in the local access network, though the number of miles over which the traffic is carried is much larger in the backbone than the local access network.

While still trying to establish positive value on both sides, recent peering disputes appear to reflect negotiations between ISPs as to the total network cost burden of carrying broadband traffic. Level3 understandably advocates a "bit-mile" approach that takes into account both the amount of traffic and the number of miles it has to be carried on its nationwide network. Comcast and Verizon's push for paid peering reflects their point of view that content provider companies like Netflix need to contribute to the cost of carrying large amounts of their traffic over the local access network. On the other side, Netflix may be trying to flex it position as the top traffic source on the Internet today into much lower transit or paid peering costs that the company has been paying in the past. The change here is that these payments are migrating from transit charges or payments to third-party CDN providers to paid peering arrangements with ISPs with access networks, and Netflix is trying to further reduce its cost, adding difficulty to the negotiations.

As the Internet evolves and different players take on different roles in terms of traffic generation and interconnection, it is useful to understand how the money flows are changing. In the latest and most newsworthy interconnection controversies, we see the disintermediation of some players and changes in the expectations of who pays whom. This is not as simple as some may first believe given the complexity of the deals that underpin interconnection. It is also worth noting that at this point, the amounts of money that appear to be involved are minor in terms of the overall operation of these various players. For example, it has been noted that the

---

[28] For a useful discussion of IXPs trends, see Steve Wilcox, "The Role IXPs and Peering Play in the Evolution of the Internet," PTC'14, 'New World, New Strategies', 19-22 January 2014. At http://www.ptc.org/ptc14/images/papers/upload/Presentation_IB2_WilcoxStephen.pdf.

cost of content for Netflix far outweighs the cost of interconnection. We cannot confirm these assertions, but the impact of these costs now, and their potential increase in the future, are interesting points to consider.

> *Key Change*: ISPs with access networks will ask for paid peering with Tier 1 ISPs and content providers to recover more of the cost of carrying increasing traffic volumes to their customers and the asymmetric nature of this traffic in favor of the downstream towards the customer.

Emergence of Online Video
The emergence of online video is a well-known trend; its implication for Internet interconnection is less understood. The rapidly growing popularity of video on the web has generated a surge in Internet traffic, with new traffic flows that undermine or require modifications to existing peering and transit agreements.[29] A pattern for content providers generating large amounts of online video content for their users appears to be forming. In their formative years, video-based content providers employ third-party CDNs to efficiently handle the distribution of their traffic to Tier 2 ISPs and avoiding large transit payments to Tier 1 ISPs (event though transit prices have been falling significantly, the huge amount of capacity consumed by streaming video can be very expensive if purchased through transit). At some point as their traffic requirements grow, these content providers shift to building and operating their own private CDNs to better control the use experience and to reduce costs by no longer outsourcing this function. [30]

The popularity in online video will also create a surge in Internet traffic. This was demonstrated last year when Netflix launched its new SuperHD format, which requires a 50% increase in transport capacity per video over 1080p content. During the first week of availability it caused a 10-15% increase in Netflix traffic and a 5% increase in peak traffic to one North American ISP.[31] Going forward, all the network segments will have to reinforce their network capacity to accommodate the high bandwidth requirements and growing demand for online video. As noted above, the network cost varies across the different network segments. The network segment incurring the most significant magnitude of cost increase will likely be the local access networks that have the lowest amount of cost sharing across Internet users. The increasing cost burden driven by online video will likely drive the justification of ISPs with access networks for paid peering arrangements.

---

[29] Sandvine reported in the first half of 2014 that real-time entertainment accounted for 63% of all downstream Internet traffic. See Global Internet Phenomena Report, Sandvine, 1H 2014, at https://www.sandvine.com/downloads/general/global-internet-phenomena/2014/1h-2014-global-internet-phenomena-report.pdf

[30] Netflix, for example, has said it is in the final stages of migrating all of its traffic off of three third-party CDNs (Akamai, Level 3 and Limelight) to its own private CDN called Open Connect by the summer of 2014. Notable, the "control plane" of the CDN continues to use Amazon Web Services. See David Fullagar, Netflix Open Connect, presentation at Content Delivery Summit, May 2014, at http://conferences.infotoday.com/documents/197/2014CDNSummit-Netflix.pdf.

[31] See Global Internet Phenomena Report, Sandvine, 2H 2013, at https://www.sandvine.com/trends/global-internet-phenomena/.

Key Change: A larger number of parallel nationwide networks will emerge as more video-based content providers develop private CDNs as the scale of the operations grow. The growing demand for online video by broadband customers will drive further capacity upgrades throughout the Internet, with the highest cost incurred in the local access network which, in turn, will provide further justification for paid peering to ISPs with access networks.

Strategic Business Considerations
At the outset of our paper we defined peering as a business relationship, and we have described a number of the important technical and economic elements that are relevant to the definition of this peering relationship. Beyond these tangible elements exist other strategic business considerations that might also influence the interest or disinterest in establishing a peering agreement between providers. Such considerations might include:

- Public Policy Goals. The peering relationship might impact the public policy goals of an organization by establishing a public record of congestion that might be beneficial to one of the parties in a public policy forum.
- Competitive Issues. A peering arrangement might enable one of the parties to become a more effective competitor to the other in existing or emerging markets.
- Market Protectionism. A peering arrangement could open a particular market segment to more competition and therefore cause incumbent providers to refuse such agreements.
- Market Leverage. Peering arrangements are likely to reflect the market strength of the different providers. ISPs in a position of market strength are more likely to be able to cast agreements to their advantage given their strong position or leverage in the relevant market.

Key Change: As long as peering arrangements reflect a business relationship, we are confident that strategic business considerations are likely to be reflected in their positions. Such a diversity of positions is to be expected as part of the normal outcomes of the peering market. However, to the extent that the peering arrangements run counter to public policy objectives, additional narrow review of the peering model may be necessary to preserve and protect the policy objectives.

Section IV: Strategic Implications

Our analysis of recent Internet traffic data and other developments in Internet interconnection lead to the following set of strategic implications of future trends impacting Internet interconnection.

1. *Defining Network Investment.* Interconnection agreements will evolve to increasingly reflect the true burden of total network cost or investment. This migration is likely to be difficult as prior business relationships established through peering have to adapt to new circumstances. The definition of network cost will be difficult to resolve. Past proxies for network investment have been traffic symmetry and range of destinations of Internet access provided by a network. As new traffic patterns have emerged due to online video, the view of network cost is changing. Level3 has proposed the concept of bit-miles, but access networks have issues regarding the degree of shared costs that can be realized as compared to middle-mile and backhaul networks. Future research into a useful framework for comparing the costs of carrying Internet traffic on different network segments would be a useful tool for network providers as well as policy makers in order to establish a more

complete understanding of the full network investment undertaken by network providers to better explain and predict future controversies.

2. *Allowing Dynamic Nature of Content*. Rapidly innovating broadband applications will continue to change the business relationships being defined through Internet interconnection agreements. As we noted earlier, when Netflix added a new high definition format, it increased total traffic of one ISP with an access network by 5% in one week. Peering agreements will need to adjust to such rapid change, along with the evolving roles of different network segments based upon the particular attributes of the new content. For example, consider the impact that a popular new application that created a huge amount of upstream content from consumers was to take hold. Then, ISPs with access networks would be in the position of having to negotiate new agreements with content providers to accept all this new traffic, perhaps in a very short time span. Peering agreements that preclude this flexibility, or public policy prescriptions that prevent it, would have significant impact on the ability of the Internet to innovate and adapt new applications that are popular to consumers.

3. *Standalone Content Networks in the Core*. Content providers with scale will continue to build their own end-to-end networks in parallel to existing Tier 1 and large Tier 2 networks. Large content providers will operate their own network infrastructure throughout the middle mile to the core backbone in order to better control the quality of their distribution network and deliver their traffic directly to the access networks.

4. *Less Reliance on Tier 1 ISPs*. Despite the significant decrease in transit prices over the past decade, the share of IP transit in the overall market for IP interconnection market has been contracting. This reflects the efforts of Tier 2 providers to peer amongst themselves instead of paying for transit from Tier 1 providers. A recent market report estimated the size of the global IP interconnection market to be $4.6B in 2013, consisting of $2.1B in revenues for IP transit and $2.5B for the circuits connecting to IXPs.[32] This same source forecasts that the share of global Internet traffic connected by transit agreements will decline from 47 percent in 2010 to 41 percent in 2014, and continued contraction of transit at historical rates will shrink the global IP interconnection market to $4.1B by 2020. Thus, transit prices will continue to fall due to improved technology and less demand. Access network, small ISPs, and content providers will continue to shift traffic away from Tier 1 transit agreements in favor of more direct peering between each other (the direct interconnection between smaller ISPs and content providers is also know as "donut peering"). Smaller ISPs will also peer more between each other as remote peering opportunities increase through multi-location and Open-IX IXPs.

5. *Need for New Framework.* Taken as a whole, the preceding analysis has documented the increasingly overlapping roles of network providers in the Internet ecosystem. The lack of differentiation among the different network providers calls into question the utility of the traditional *Tiered* framework as a useful description of the current Internet peering ecosystem, and suggests the need for a new, updated framework to predict and describe future peering relationships might be more useful.

---

[32] See http://www.telegeography.com/press/press-releases/2014/07/08/ip-transit-revenues-volumes-dependent-on-peering-trends/index.html

Losing la Video Local – or –
The Role of Local Video in the Modern Video Marketplace

Cable Academic Workshop

Abstract

Justin (Gus) Hurwitz

Cable Show Academic Workshop participants: For our workshop discussion, I expect to focus on part three of the paper, described in the abstract below. In particular, I will focus on whether intervention is needed to promote local video in the modern video marketplace and, assuming it is, different forms that such intervention may take. This of course assumes affirmative answers to sections one and two of the paper – whether local video is being left behind and whether this is lamentable – both of which are open to discussion and I am happy to discuss at the workshop

Local content has long been an important part of the video marketplace; promoting the development of local content has been a central regulatory and statutory goal. Promoting localism – both as a good in itself and as a means to promoting competition and diversity – has long been a cornerstone of the FCC's approach to video regulation. And protecting local television against competition from imported distant signals gave rise to policy, judicial, and legislative decisions that define much of the current regulatory framework governing the video industry. Indeed, one could characterize local broadcaster channels as the basic "unit" of the video marketplace – arguably both the relevant statutes and FCC regulations do just this.

But despite its historical importance, localism seems a fleeting concern today. Much of the modern video industry is focusing on Internet-based modes of production, distribution, and consumption – and it is doing so largely outside the reach of regulators and with little concern for localism values. And participants in the traditional video industry – both on the consumer and production/distribution side – are following along this path.

This paper considers the role of localism is the modern video marketplace: whether it is really being lost, whether this loss is lamentable, whether we should do anything to preserve or promote localism in the modern video marketplace, and, if so, whether and how we can do so. The paper's primary two contributions are its consideration of the ways in which localism may be important in the modern video setting and development of potential ways to promote localism in this setting.

It proceeds in three parts: the role of localism over time, from the FCC's early regulatory efforts to the modern setting; localism's virtues, considering ways in which localism may be important in the modern setting; and then turning to localism's future, to ask whether localism needs to be promoted in the modern video marketplace and to consider ways of doing so. Each of these sections is discussed in slightly more – but still general – detail below.

The discussion of *localism over time* starts in the 1950s with the FCC's mantra of "localism, diversity, competition," its efforts to promote these virtues with UHF broadcast, and the subsequent tensions this created between FCC policy and the nascent cable industry. It then turns to the regulatory and statutory framework that emerged in FCC rules and the Copyright and Cable Acts – a framework that, generally, ensured favorable treatment for local broadcasters on non-broadcast media, in exchange for a relatively modest oversight of those media. It then turns to the modern setting to consider what role local content plays in the modern – and future, presumptively Internet-oriented – video marketplace.

The next section considers *localism's virtues*. Again, it starts by returning to the FCC's longtime mantra of "localism, diversity, competition," reviewing subsequent critiques of the premise that localism supports values of diversity and competition – indeed, subsequent evidence demonstrates that promoting local ownership can come at the expense of diversity and competition. It then considers localism in the modern setting, identifying unique values that localism supports, or values that localism supports uniquely well. These are placed into two general categories: ways in which localism promotes important democratic values, and in which it produces important content. "Democratic values" include, for instance, encouraging interest in and facilitating access to the media industry, serving as a laboratory for media formats and as a "farm team" for new media participants, and providing an important platform for minority-, foreign-language, and otherwise underrepresented programming. And important local content includes, for instance, news and information programming, emergency information, and local political and commercial information (especially through advertising). Critical among consideration of these content concerns is the "search" and "discovery" functions of local programming.

The final section looks to *localism's future*, asking whether the place of local video content in the emerging video marketplace should be protected and, if so, how. It begins with skepticism, asking whether content with localisms' virtues is really being lost – or rather whether the market is merely adapting to new technologies – and, relatedly, whether, to the extent such content is being lost, the market will develop new mechanisms to fill the void. Despite this skeptical start, there are some forms of local content that are both valuable and unlikely to be produced in the new, Internet-based, video marketplace. For instance, emergency information services, local political programming, and "search" and "discovery" of new information, are unlikely to translate to or be replicated by emerging platforms.

In response to these concerns, a number of approaches to preserving and promoting the development of local content are considered – along with challenges to them. Chief among these approaches, this paper explores incentive- and quid-pro-quo models to encourage the development of local programming and the favorable carriage of that programming. For instance, it considers the use of incentives for carriage local programming and beneficial spectrum swaps for producers of local programming. In considering such approaches, this paper pays special attention to the myriad ways in which the existing regulatory regime – often as the result of efforts to promote localism – is ossified or may slow the development of the video marketplace. This leads to consideration of ways to promote local video content with the secondary goal – as part of a quid-pro-quo model – of addressing such ossification and other regulatory impediments to the video marketplace.

The Welfare Effects of Vertical Integration in Multichannel Television Markets*

Gregory S. Crawford† Robin S. Lee‡ Michael D. Whinston§ Ali Yurukoglu¶

[PRELIMINARY AND INCOMPLETE; PLEASE DO NOT CIRCULATE]

February 14, 2015

Abstract

We investigate the welfare effects of vertical integration of regional sports networks (RSNs) in U.S. multichannel television markets. Vertical integration can enhance efficiency by reducing double marginalization and increasing carriage of these channels, but can also harm welfare due to foreclosure and raising rivals' costs incentives. We estimate a structural model of viewership, subscription, distributor pricing, and affiliate fee bargaining using a rich dataset on the U.S. cable and satellite television industry (2000-2010). We use these estimates to analyze the impact of simulated vertical mergers and de-mergers of RSNs by distributors on competition and welfare, and examine the efficacy of regulatory policy introduced in the 1992 Cable Act by the U.S. Federal Communications Commission.

1 Introduction

The welfare effects of vertical integration is an important, but controversial, issue. The theoretical literature on the pro- and anti-competitive impacts of vertical integration is vast (c.f. Perry, 1990; Whinston, 2006; Riordan, 2008), and typically contrasts potential efficiencies related to the elimination of double marginalization (Spengler, 1950) and the alignment of investment incentives (Willamson, 1985; Grossman and Hart, 1986) with the potential for losses arising from incentives to foreclose rivals and raise their costs (Salop and Scheffman, 1983; Krattenmaker and Salop, 1986; Hart and Tirole, 1990; Ordover et al., 1990). However, despite a growing literature, empirical evidence on the quantitative magnitudes of these potential effects (and the overall net welfare impact) is still limited.

This paper attempts to quantify the welfare effects of vertical integration in cable and satellite television in the context of high value regional sports programming in the U.S. Whether or not

---

*We would like to thank ESRC Grant RES-062-23-2586 (Crawford), the NYU Stern Center for Global Economy and Business (Lee), and the Toulouse Network for Information Technology and the NSF (Whinston) for support. All errors are our own.

†Department of Economics, University of Zurich; gregory.crawford@econ.uzh.ch

‡Department of Economics, Harvard University; robinlee@fas.harvard.edu

§Department of Economics and Sloan School of Management, MIT; whinston@mit.edu

¶Graduate School of Business, Stanford University; yurukoglu_ali@gsb.stanford.edu

the ownership of content by distributors harms welfare is at the heart of several recently proposed (e.g., Comcast and Time Warner, AT&T and DirecTV) and consummated (e.g., Comcast and NBC, approved in 2011) mergers in the television industry. The attention these mergers attracted is due to the industry's overwhelming reach and size: nearly 90% of the 115 million television households in the U.S. subscribe to multichannel television, and the mean individual consumes about three hours of television per day. Regional sports programming is a large part of this industry, receiving \$4.1 billion out of the over \$30 billion per year in negotiated affiliate fees paid by distributors to all content providers, and an additional \$700 million per year in advertising dollars.

Our focus on the multichannel television industry, and in particular sports programming, is also driven by several factors that create empirical leverage to address this question. First, there is significant variation across the industry in terms of ownership of content by multichannel video programming distributors (MVPDs). Second, although this variation is primarily at the national level for most channels, regional sports networks (RSNs) are present in smaller geographic areas, and thus provide useful variation in ownership patterns both across regions and over time. Third, the industry is the subject of significant regulatory and antitrust attention in addition to merger review, including the application of "program access rules" and exceptions to this rule, such as the "terrestrial loophole" which exempted certain distributors from supplying integrated content to rivals.

The heart of our paper is the specification and estimation of a structural model of the multichannel television industry that captures consumer viewership and subscription decisions, MVPD pricing and carriage decisions, and bargaining between MVPDs and content providers. Our paper builds off and significantly extends the model in Crawford and Yurukoglu (2012) into an empirical framework for the analysis of vertical integration and mergers, and importantly incorporates: (i) incentives to foreclose rivals' access to inputs, (ii) the potential for double marginalization, and (iii) the possibility of imperfect coordination and internalization within an integrated firm. We use data on both aggregate and individual level consumer viewership and subscription patterns along with price, quantity, and channel carriage for cable and satellite at the local market level over the years 2000 to 2010.

We leverage our structural model to highlight the mechanisms through which pro-competitive and anti-competitive effects of vertical integration might occur. In particular, we estimate: (i) the degree to which firms internalize the profits of integrated units when making pricing and channel carriage decisions; and (ii) the degree to which integrated channels act on incentives to deny access to rival distributors. Central to identifying both of these effects are our estimates of the changes in firm profits from the addition or removal of an RSN from its bundles, which in turn relies on using variation in distributor market shares as channel bundle offerings change to inform how much consumers value content when subscribing to an MVPD, and variation in observed viewership patterns and negotiated affiliate fees across channels to infer the relative values consumers place on different channels. Given these estimated profit effects, the pro-competitive effect of vertical integration is largely identified from the degree to which carriage is higher for integrated distributors, while the anti-competitive foreclosure effect is identified by lower supply to downstream rivals of integrated channels.

Using these sources of identification, we find that integrated distributors do internalize the effects of their pricing and carriage decisions on their upstream channels' profits, although not to the full extent typically assumed in theoretical models of integration. We also find that integrated RSNs take into account to a substantial degree the benefits their downstream divisions reap when a rival distributor is denied access to the RSN's programming.

We use our estimated model to examine the effects of vertical integration, both when program access rules ensuring non-integrated rival distributed access to integrated content are effectively enforced, and when they are not. In the former case, foreclosure arising from vertical integration is prevented and our counterfactual captures the effects due to improved internalization of pricing and carriage decisions between the integrated units. In the latter case, the integrated firm (typically a cable distributor) may engage in foreclosure, denying non-integrated rival distributors (typically a satellite distributor) access to its RSN.

We find that vertical integration in the presence of effective program access rules leads to significant gains in both consumer and aggregate welfare. These benefits arise both due to lower retail prices and greater carriage of the RSN. Averaging across results for 27 RSNs that were active in 2007, we find that integration of a *single* RSN would reduce average cable prices by approximately \$0.50 per subscriber per month in that RSN's market, and increase carriage of the RSN by its integrated owner by approximately 8%. Combined, these effects would yield, on average, approximately a 1.2% increase per month per capita increase in consumer surplus.

However, when vertical integration occurs in the absence of effective program access rules, consumers are often harmed. This occurs in cases in which the integrated firm denies access to its RSN to rival distributors: for 8 of our RSNs, we predict that exclusion of satellite by integrated cable providers would be preferred, and in the majority of these cases, consumer welfare harmed. Such foreclosure tends to occur when the RSN is owned by a cable operator whose overall market share in the region served by the RSN is large: i.e., we do not find an integrated cable provider wishing to exclude satellite unless its percentage of households that it could serve exceeds 40%. Furthermore, when we do not predict foreclosure, we document that integration would encourage an RSN to increase the prices charged to rivals of its integrated distributor by over 30%.

Overall, we find the net effect of vertical integration—allowing for both efficiency and foreclosure incentives—to be heterogeneous. For those RSNs that are integrated in our setting, we predict that the impact on total and consumer welfare to be negative; for those RSNs that are not, the impact is slightly positive.[1]

The effects we document are only partial, as our analysis does not incorporate the potentially important effect of vertical integration on investments by RSNs and MVPDs, both those that integrate and their rivals. In principle, these investment effects on consumer and aggregate surplus

---

[1] Our analysis considers integration of each RSN individually, and therefore assumes that (national) satellite prices are unchanged. It may therefore understate consumer losses from widespread integration absent effective enforcement of program access rules.

could go either way, as emphasized in the literature on investment effects of vertical integration (Bolton and Whinston (1991), Hart (1995)). For example, in all cases, the impact on satellite distributors is negative, raising the possibility that widespread integration by cable distributors of RSNs might impact satellite distributors' effectiveness as a competitor to cable to a greater extent than admitted in our analysis.

Related Literature. Previous work in the cable industry, including Waterman and Weiss (1996), Chipty (2001), and Chen and Waterman (2007), have primarily relied on reduced form cross-sectional analyses for a limited subset of channels and found that integrated cable systems are more likely to carry their own as opposed to rival content. One exception that uses variation over time is Suzuki (2009), which studies the 1996 merger between Time Warner and Turner broadcasting and finds that integrated channels were more likely to be carried by Time Warner systems following the merger, and that non-integrated rival channels were less likely to be carried in Time Warner markets after the merger. In this regards, both this and our companion paper (Crawford et al., 2014) complement previous work in the cable industry with a richer panel dataset, and with a structural model we are able to both provide welfare measurements and shed light on the mechanisms through which vertical integration has an effect on welfare.

This paper also adds to the growing empirical literature on the effects of vertical integration and arrangements (e.g. Shepard, 1993; Asker, 2004; Hastings, 2004; Hastings and Gilbert, 2005; Hortacsu and Syverson, 2007; Villas-Boas, 2007; Mortimer, 2008; Houde, 2012; Lee, 2013). We build on existing approaches by estimating a model that explicitly incorporates avenues for vertical integration to both improve the efficiency of pricing and channel carriage decisions and to generate foreclosure of rival distributors, and by providing estimates of the degree to which integrated firms, in practice, act on each of these incentives. Using these estimates, we provide estimates of the welfare impacts of vertical integration that weigh these pro- and anti-competitive effects.[2]

2 Institutional Detail and Data

Our study analyzes the U.S. cable and satellite industry for the years 2000 to 2010 and focuses on the ownership of "Regional Sports Networks" (RSNs) by cable and satellite distributors. In this section, we describe the industry structure, RSNs, and regulatory policy during this period. We then discuss the data that we use to estimate the model. The tables referenced in this section are contained in Appendix B.

In the time period we study, the vast majority of households in the US were able to subscribe to a multichannel television bundle from one of three downstream multichannel video programming distributors (MVPDs): a local cable company (e.g., Comcast, Time Warner Cable, or Cablevision) or one of two nationwide satellite companies (DirecTV and Dish Network).[3] Cable companies

---

[2] See also Conlon and Mortimer (2013).

[3] In our analysis, we focus only on markets where there is a single cable provider. Telecommunications MVPD providers AT&T and Verizon did not enter a significant number of markets until 2007; by the end of 2010, AT&T

transmit their video signals through a physical wire whereas satellite companies distribute video wirelessly through a south-facing satellite dish attached to a household's dwelling. The majority of distributors' revenue comes from subscription to three different bundles of programming: a limited basic bundle which retransmits over-the-air broadcast stations, an expanded basic bundle containing 40-60 of the most popular channels available on cable (e.g., AMC, CNN, Comedy Central, ESPN, MTV, etc.), and a digital bundle containing between 10 to 50 more, smaller, niche channels.

Downstream distributors negotiate with content producers over the terms at which the distributors can offer the content producers' channels to consumers. These negotiations usually center on a monthly per-subscriber "affiliate" fee that the downstream distributor pays the channel for every subscriber who has access to the channel, whether the subscriber watches it or not. According to industry estimates, RSNs command the second-highest per-subscriber affiliate fees after ESPN. For example, NESN is reported to have per-subscriber monthly fees that averaged \$2.72 per month in 2010 whereas highly-rated national channels such as Fox News, TNT, and USA hover around \$1 per subscriber per month.

2.1 Vertical Affiliation of RSNs in Multichannel Television Markets

RSNs carry professional and college sports programming in a particular geographic region. For example, the New England Sports Network (NESN) carries televised games of the Boston Red Sox and the Boston Bruins that aren't concurrently being televised nationally. Metropolitan areas can have multiple RSNs: e.g., in the New York City metropolitan area, there are four different RSNs (Madison Square Garden (MSG), MSG Plus, SportsNet NY, and Yankees Entertainment and Sports (YES)). Some RSNs also serve multiple metropolitan areas: e.g., the Sun Sports network holds the rights to the Miami Heat and the Tampa Bay Rays, amongst others. Table 6 provides a variety of information about the largest RSNs in the US, including their availability, their average (across systems and years) affiliate fee, and average (across DMAs and years) viewership.

Figure 1 shows each RSN's years of operation between 2000 and 2010 and ownership affiliation with a downstream distributor. Many RSNs are owned, to some degree, by a downstream distributor. For example, in 2007, downstream distributors had ownership interests in 16 out of the 30 active RSNs. The cable MVPDs that owned RSNs are Comcast, Cablevision, Cox, and Time Warner; DirecTV, the largest satellite operator (and second-largest US MVPD), indirectly had stakes in numerous RSNs through its partial owners News Corporation and Liberty Media Corporation.[4]

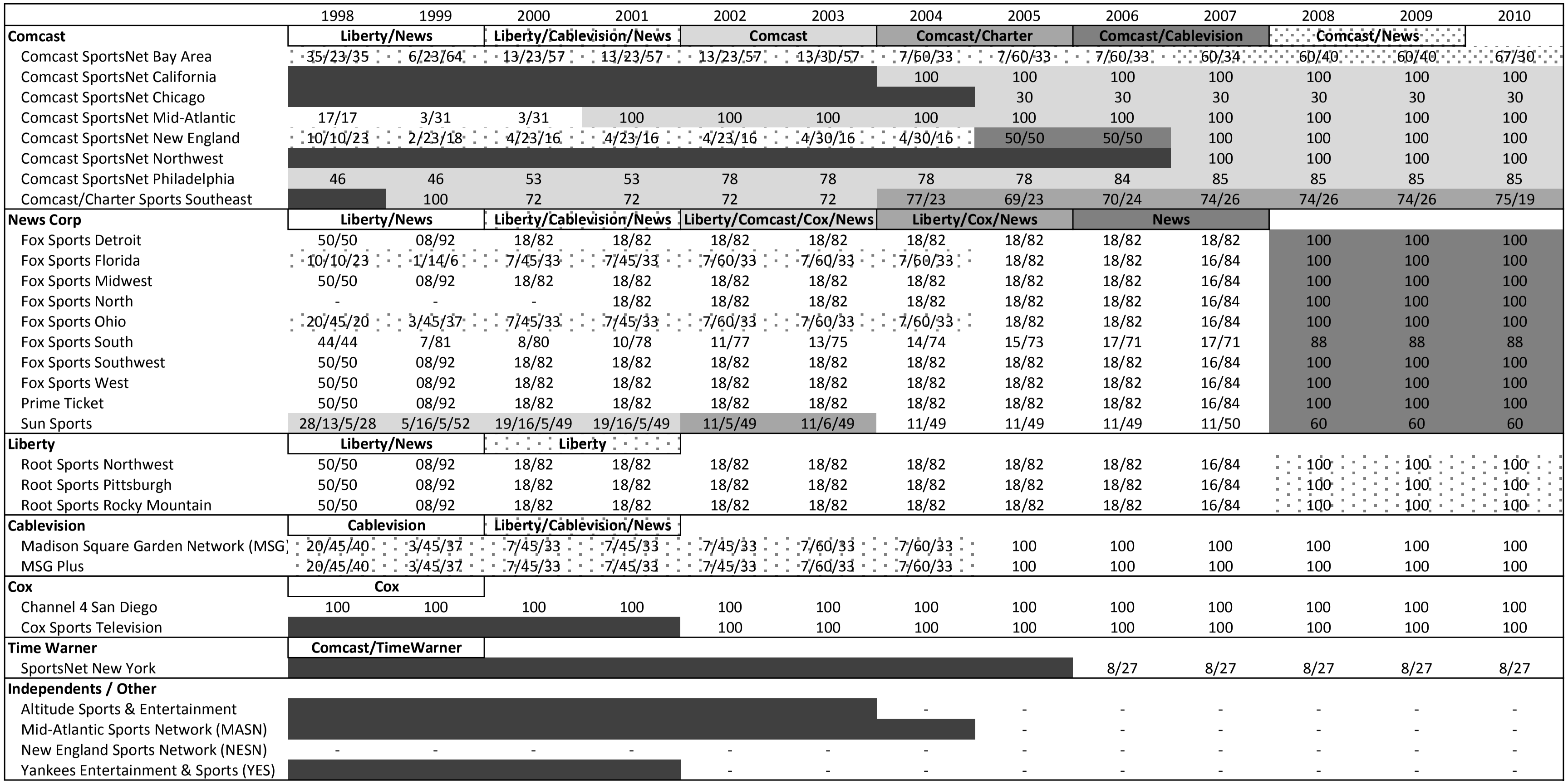
Regulatory Policy. There are several key features of the regulatory environment for RSNs, and vertically integrated content more generally, that are relevant during our sample period. First, during our sample period, vertically integrated firms were subject to the "Program Access Rules"

---

and Verizon had a total of 3.0M and 3.5M subscribers each.

[4] News Corporation and Liberty Media both had a partial ownership stake in DirecTV from 2003 onwards; News Corporation sold its stake in 2006.

Figure 1: RSN Ownership



	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Comcast	**Liberty/News**		**Liberty/Cablevision/News**		**Comcast**		**Comcast/Charter**		**Comcast/Cablevision**		**Comcast/News**		
Comcast SportsNet Bay Area	35/23/35	6/23/64	13/23/57	13/23/57	13/23/57	13/30/57	7/60/33	7/60/33	7/60/33	60/34	60/40	60/40	67/30
Comcast SportsNet California	■	■	■	■	■	■	100	100	100	100	100	100	100
Comcast SportsNet Chicago	■	■	■	■	■	■	■	30	30	30	30	30	30
Comcast SportsNet Mid-Atlantic	17/17	3/31	3/31	100	100	100	100	100	100	100	100	100	100
Comcast SportsNet New England	10/10/23	2/23/18	4/23/16	4/23/16	4/23/16	4/30/16	4/30/16	50/50	50/50	100	100	100	100
Comcast SportsNet Northwest	■	■	■	■	■	■	■	■	■	100	100	100	100
Comcast SportsNet Philadelphia	46	46	53	53	78	78	78	78	84	85	85	85	85
Comcast/Charter Sports Southeast	■	100	72	72	72	72	77/23	69/23	70/24	74/26	74/26	74/26	75/19
News Corp	**Liberty/News**		**Liberty/Cablevision/News**		**Liberty/Comcast/Cox/News**		**Liberty/Cox/News**		**News**				
Fox Sports Detroit	50/50	08/92	18/82	18/82	18/82	18/82	18/82	18/82	18/82	18/82	100	100	100
Fox Sports Florida	10/10/23	1/14/6	7/45/33	7/45/33	7/60/33	7/60/33	7/60/33	18/82	18/82	16/84	100	100	100
Fox Sports Midwest	50/50	08/92	18/82	18/82	18/82	18/82	18/82	18/82	18/82	16/84	100	100	100
Fox Sports North	-	-	-	18/82	18/82	18/82	18/82	18/82	18/82	16/84	100	100	100
Fox Sports Ohio	20/45/20	3/45/37	7/45/33	7/45/33	7/60/33	7/60/33	7/60/33	18/82	18/82	16/84	100	100	100
Fox Sports South	44/44	7/81	8/80	10/78	11/77	13/75	14/74	15/73	17/71	17/71	88	88	88
Fox Sports Southwest	50/50	08/92	18/82	18/82	18/82	18/82	18/82	18/82	18/82	16/84	100	100	100
Fox Sports West	50/50	08/92	18/82	18/82	18/82	18/82	18/82	18/82	18/82	16/84	100	100	100
Prime Ticket	50/50	08/92	18/82	18/82	18/82	18/82	18/82	18/82	18/82	16/84	100	100	100
Sun Sports	28/13/5/28	5/16/5/52	19/16/5/49	19/16/5/49	11/5/49	11/6/49	11/49	11/49	11/49	11/50	60	60	60
Liberty	**Liberty/News**		**Liberty**										
Root Sports Northwest	50/50	08/92	18/82	18/82	18/82	18/82	18/82	18/82	18/82	16/84	100	100	100
Root Sports Pittsburgh	50/50	08/92	18/82	18/82	18/82	18/82	18/82	18/82	18/82	16/84	100	100	100
Root Sports Rocky Mountain	50/50	08/92	18/82	18/82	18/82	18/82	18/82	18/82	18/82	16/84	100	100	100
Cablevision	**Cablevision**		**Liberty/Cablevision/News**										
Madison Square Garden Network (MSG	20/45/40	3/45/37	7/45/33	7/45/33	7/45/33	7/60/33	7/60/33	100	100	100	100	100	100
MSG Plus	20/45/40	3/45/37	7/45/33	7/45/33	7/45/33	7/60/33	7/60/33	100	100	100	100	100	100
Cox	**Cox**												
Channel 4 San Diego	100	100	100	100	100	100	100	100	100	100	100	100	100
Cox Sports Television	■	■	■	■	100	100	100	100	100	100	100	100	100
Time Warner	**Comcast/TimeWarner**												
SportsNet New York	■	■	■	■	■	■	■	■	8/27	8/27	8/27	8/27	8/27
Independents / Other													
Altitude Sports & Entertainment	■	■	■	■	■	■	-	-	-	-	-	-	-
Mid-Atlantic Sports Network (MASN)	■	■	■	■	■	■	■	-	-	-	-	-	-
New England Sports Network (NESN)	-	-	-	-	-	-	-	-	-	-	-	-	-
Yankees Entertainment & Sports (YES)	■	■	■	■	-	-	-	-	-	-	-	-	-

Notes: Reported are the vertical ownership stakes held by major distributors of cable and satellite television service for the 30 Regional Sports Networks (RSNs) in our data that are active in 2007. Ownership data was collected by hand from company stock filings and industry sources. The ownership share for each distributor is reported and individual owners (or combinations of owners) are shaded according to the channel-specific legend located in the first line of each potential owner. Dark grey shading corresponds to a year in which the given RSN is not active (i.e., has not yet entered or has exited the market). Hyphens correspond to years of active operation for an RSN without a vertical ownership affiliation.

(PARs). These required that vertically integrated content be made available to rival distributors at non-discriminatory prices, subject to final-offer arbitration if necessary.

The PARs only applied, however, to content that was transmitted to the MVPD via satellite. This covered all national cable channels (who need satellite transmission to cost-effectively reach cable systems around the country) and most RSNs. A handful of RSNs, however, transmitted their signal terrestrially (usually via microwave), thereby avoiding the jurisdiction of the PARs. This was called the "terrestrial loophole" in the Program Access regulation. In 2007, only two cable-integrated RSNs were able to leverage the terrestrial loophole: Comcast SportsNet in Philadelphia and SD4 in San Diego (owned by Cox Cable); in both cases, the channel was not provided to satellite providers.[5] As a result, Major League Baseball (MLB), National Basketball Association (NBA), and National Hockey League (NHL) games in Philadelphia were only available through cable and not through DirecTV or Dish Network. Similarly in San Diego, MLB games were available only through cable. This accident of regulatory history will be an important source of identifying variation in our econometric estimation.

The PARs were introduced in 1992 and required renewal by the FCC every five years. They were allowed to lapse in 2012 and replaced by rules giving the Commission the right to review *any* programming agreement for anti-competitive effects on a case-by-case basis under the "unfair acts" rules the Commission established in 2010 (FCC (2012)). The new case-by-case rules explicitly include a (rebuttable) presumption that exclusive deals between RSNs and their affiliated distributors are unfair.[6]

During our sample period (2000-2010), most integrated RSNs outside of loophole markets had agreements to be carried by all MVPDs; however, even though the PARs were in effect, there were cases in which a cable-owned RSN was not carried by satellite providers: in 2007, these channels were Comcast Sports Northwest, Comcast/Charter Sports Southeast, and Cox Sports Television. Furthermore, independent channels were not necessarily always provided to all MVPDs. For example, this happened once in 2007: YES was not carried by Dish Network.

2.2 Data

We collect a wide variety of data to analyze the effects of vertical integration. We have three categories of data: (1) downstream prices, quantities, and characteristics of cable and satellite bundles, (2) channel viewership data, and (3) channel affiliate fees and advertising revenues. We briefly describe each in turn.

---

[5] Time Warner Cable also employed the terrestrial loophole from 2006 to 2008 for the (then relatively new) Charlotte Bobcats NBA franchise by placing some their games on News 14, a terrestrially delivered regional news channel.

[6] There are still cases on non-carriage of integrated RSNs on rival MVPDs; one high profile example is Time Warner Cable SportsNet LA, with rights to air Los Angeles Dodger Games, not being carried on DirecTV.

2.2.1 Downstream Prices, Quantities, and Characteristics

We combine data from multiple databases to construct downstream prices, quantities, and characteristics. Our foundational dataset is the Nielsen FOCUS database. It provides, for each cable system in the US, the set of channels offered, the number of homes passed, the total number of subscribers (i.e. to any bundle), the owner of the system, and the zip codes served. We use the years 2000 to 2010. We restrict our analysis to system-years in which the system faced no direct wire-based competition.[7] We construct market shares from the number of subscribers by dividing by the number of households in a market (obtained from 2000 and 2010 Census data). We combine these data with individual-level survey data from household survey firms Mediamark Research & Intelligence (MRI) and Simmons, using MRI data for 2000 to 2007, and Simmons for 2008-2010. Specifically, if a system-year had over 40 survey respondents, we use the average of the market share from the FOCUS data and the cable market share among the survey respondents; otherwise we use only the FOCUS data. We further eliminate any system-year for which the FOCUS subscriber data was not updated from the previous year, or we did not have at least 40 survey respondents in the MRI/Simmons data. We use the remaining system-years to construct our markets.

For our analysis, we define a market for each year to be a set of zip codes served by a single cable system and, by construction, both satellite providers. For cable systems, we aggregate over bundles within a system, focusing on total system subscribers. Our demand model is therefore a distributor choice model, rather than a bundle choice model.[8] We construct satellite shares within each of our markets for DirecTV and Dish Network from the MRI/Simmons survey data.[9] Furthermore, we gathered historical channel offerings and prices for DirecTV and Dish Network through the Internet Archive (archive.org).

We combine multiple sources of information on cable television prices. Systems regularly post prices for their tiers of service on their websites and these websites are often saved in the Internet Archive.[10] We use the price of Expanded Basic Service, the most popular bundle chosen by households and the bundle which typically contains all the channels in our analysis. Furthermore, newspapers often report when prices change at local cable systems. Some newspapers report this information every time cable prices change (typically yearly), providing valuable information about the history of price changes for a single (often large) system or geographic family of systems owned by the same provider. Finally, cable systems typically have "rate cards" describing their current

---

[7] We do so because when a system faces competition from another cable operator we do not know the number of subscribers in the areas where the system faced competition relative to the areas where it did not.

[8] While we would prefer a bundle demand model, our subscriber data was not rich enough to estimate bundle-specific quantities. This isn't overly limiting, as our focus is on the impact of vertical integration on inter-distributor demand.

[9] We use satellite state market shares unless we have at least 5 respondents in the individual-level data, in which case we take the average of the satellite state market shares and within-market market shares, placing greater weight on the market-level survey data the greater the number of observations. We dropped any constructed market whose total market share exceeded one or which had a zero market share for one of the satellite providers (which happens naturally due to sampling error).

[10] Following industry practice, we refer to the set of channels offered at a given (incremental) price as a tier of service and the combination of tiers chosen by households as the bundle that they buy. Thus the expanded basic bundle (service) consists of the limited basic tier and the expanded basic tier.

tiers, channels, and prices which they use for marketing or to inform customers of changes in these offerings; they were used when able to found online. We searched the Internet for all such information about cable prices and linked the information obtained to FOCUS systems by hand based on the provider, principal community served, and other communities served as reported in the newspaper or listed on the rate card. For system-years where we do not find a price from websites, rate sheets, or newspapers, we link to the TNS Bill Harvesting database. These data are individual-level bills for cable service which report the company providing the service, the household's expenditure, and their zip code. For a given system-year if we have at least 5 respondents, we use the mean expenditure for subscribers to that system.[11] These data also provide the level of a tax on satellite television service in states where it exists, which we use as an instrumental variable for price in demand estimation.

Table 7 reports the average price, market share, and number of RSN, cable, and total channels offered across markets and years in our estimation dataset. We use 11 years of data, comprising over 6,000 market-years, with an average coverage of 31.5 million (roughly 30% of) US households per year.[12] Average prices are quite similar across providers, whether on an unweighted basis or weighted by the number of households in the market. The satellite companies generally offer more channels on their Expanded Basic service than the local cable system, but a similar number of RSNs.

We derive MVPD margins for for DirecTV and Dish from their 2007 10K reports.[13]

2.2.2 Viewership

We estimate demand using both bundle purchase and viewing data. We have two types of viewing data: some at the level of individual households and others reporting aggregate viewing decisions at the level of the Designated Market Area (DMA "ratings").[14] Average viewership for various RSNs is reported in Table 6 and average viewership for other cable networks is reported in Tables 8-9.

The first group of data come from our MRI and Simmons datasets described in the previous sub-section. Our MRI data reports the number of hours watched for each of the sampled households

---

[11] We only use bills which clearly delineate video programming costs (i.e., that separate it out from other bundled services such as internet and phone), and use the average of a system's service tier revenue (i.e., excluding pay-per-view or one-time charges) to construct prices.

[12] While we observe the population of channel lineups, incomplete reporting of subscriber information in the FOCUS dataset and the inability to collect cable prices in some markets prevents us from constructing the information we need in every U.S. cable market.

[13] We compute Comcast margins using video, advertising, and franchise fee revenues; programming expenses; and sales, general, and administrative (SG&A) expenses multiplied by both the video revenue share of total revenues (to proportionately allocate expenses across Comcast's other businesses) and the share of SG&A expenses that are subscriber acquisition and retention related (computed from DirecTV's reports). We compute DirecTV margins using total revenues; and programming, subscriber acquisition, upgrade, and retention expenses. For Dish, we use total revenues; subscriber acquisition costs; and the share of subscriber related expenses multiplied by the share of non-SG&A costs (programming and service expenses) that are programming related (computed from DirecTV's reports). The computed values are $\{.539, .396, .413\}$.

[14] DMAs are mutually exclusive and exhaustive definitions of television markets created by Nielsen and used for the purchase of advertising time.

of 96 national channels from 2000 to 2007, while our Simmons data reports the same information for 99 national channels between 2008 and 2010. Our aggregate ratings data come from Nielsen. Reported is the average rating for each of between 63 and 100 channels, of which 18 to 29 are RSNs, depending on the year, in each of the 44 to 56 largest DMAs between 2000 and 2010.

Tables 6, 8, and 9 report summary statistics for our viewing data. Tables 8 and 9 report, for each of our sources of viewing data, the mean rating for each of the 87 non-RSNs in either dataset, as well as additional information from our household data. For example, the average rating for the ABC Family Channel in the Nielsen data across the 747 DMA-years for which the information was recorded is 0.418. This is measured in percentage points, so it suggests a household selected at random in one of these years and DMAs would be watching the ABC Family Channel with probability 0.418 percent. While small, this is above average for cable networks. Similarly, the average rating for the RSN, Yankees Entertainment & Sports (YES) from Table 6, is 0.27. For RSN viewership, we have additional information about the average RSN rating by platform chosen by households (i.e. cable or each satellite operator), which we report there.

Our household-level data provide further details about viewing which are summarized in the remaining columns of Tables 8 and 9. The last column reports the share of households on average across DMAs and years that report *any* viewing of that channel. As noted in Crawford and Yurukoglu (2012), this provides valuable information about whether a household has any interest in a channel that we will use to inform the estimated distribution of preferences for channels across households.[15]

2.2.3 Average Affiliate Fees and Advertising Rates

As described earlier, affiliate fees are the monthly per-subscriber charges paid by distributors to content providers for the ability to distribute the channel. SNL Kagan maintains a database with aggregate information about individual cable television networks, both nationally-distributed networks like CNN and ESPN as well as RSNs like the family of Comcast and Fox networks. For many networks, we use information about the average affiliate fee paid by cable systems to each such network. For cable channels, we have information about affiliate fees paid by between 120 and 210 channels per year between 2000 and 2010. For RSNs, we also have information about the total national subscribers served by each of 88 providers between 2000 and 2010. These are also reported in Tables 6, 8, and 9. The average affiliate fee in our data is \$0.16 per subscriber per month for a nationally distributed channel and \$1.45 for an RSN.

Per-subscriber advertising rates are determined for each channel by dividing total advertising revenues (provided by Kagan) by total subscribers.

---

[15] The MRI/Simmons data allows us to estimate the probability that a given channel is never watched for national channels; we regress this probability on viewership to impute this probability for RSNs.

3 Model

In this section, we present an industry model that predicts: (i) household viewership of channels; (ii) household demand for multichannel television services; (iii) prices and bundles that are offered by distributors; and (iv) negotiated distributor-channel specific affiliate fees. One key output from the specification and estimation of our model is the impact on viewership and demand of adding or removing channels from a bundle, which in turn informs the degree to which firms internalize the profits of integrated units when making strategic decisions, and the incentives of an RSN to provide or withold access to its content to distributors.

3.1 Overview

Index consumer households by i, markets by m, and time periods by t. There are a set of "downstream" multichannel video programming distributors (MVPDs) $\mathcal{F}_t$ and "upstream" channels $\mathcal{C}_t$ active in each period t. MVPDs create and maintain a distribution network and perform retail activities such as billing, packaging, and technical support. Examples include Comcast, Time Warner Cable, Cox, Cablevision, DirecTV, and municipal cable companies.

Let the set of MVPDs active in a given market-period be denoted $\mathcal{F}_{mt}$. We will assume that each distributor $f \in \mathcal{F}_{mt}$ in each period offers a single "bundle" in market m, where a household subscribing to this bundle pays a price p_{fmt} and has access to a set of channels $\mathcal{B}_{fmt} \subseteq \mathcal{C}_t$.[16] We will use f to denote both a firm as well as a bundle offered by that firm.

We assume the following timing: in **stage 1** channels and distributors bargain bilaterally to decide affiliate fees, and distributors also simultaneously set prices and make carriage decisions for each market in which they operate; in **stage 2** households choose which firm, if any, to subscribe to in their market; and in **stage 3** households view television channels. We now provide details of each stage and further assumptions, proceeding in reverse order of timing.

3.2 Stage 3: Household Viewing

Household i in market m and period t subscribing to firm $f \in \mathcal{F}_{mt}$ allocates its time between watching available channels ($\{c\} \subseteq \mathcal{B}_f$) and non-television activity (denoted by $c = 0$) to solve:

$$\begin{aligned} \max_{\boldsymbol{t}_{if}} v_{if}(\boldsymbol{t}_{if}) \quad &= \quad \sum_{c \in \mathcal{B}_f \cup \{0\}} \frac{\gamma_{ict}}{1-\nu_c} (t_{ifc})^{1-\nu_c} \qquad (1) \\ s.t.: \quad & \quad t_{ifc} \geq 0 \; \forall \; c \\ & \quad t_{ifc} = 0 \; \forall \; c \notin \{\mathcal{B}_f \cup \{0\}\} \\ & \quad \sum_{c \in \mathcal{B}_f \cup \{0\}} t_{ifc} \leq T \end{aligned}$$

---

[16] In the previous section we discussed how we deal with firms within a market offering multiple bundles.

Parameters γ_{ict} and $\nu_c \in [0,1)$ are taste parameters for channels, where γ_{ict} sets the level of marginal utility of household i from the first instant of watching channel c, and ν_c controls how fast this marginal utility decays in the amount of time watched. This viewership model is equivalent to the Cobb-Douglas model used in Crawford and Yurukoglu (2012) if $\nu_c \to 1$ for all c. We restrict ν_c to be equal for all non-sport channels and the outside-option, and equal for all sports channels (which include RSNs); i.e., $\nu_c = \nu^S$ if c is a sports channel, and $\nu_c = \nu^{NS}$ otherwise.[17]

We parameterize $\boldsymbol{\gamma}_{it} \equiv \{\gamma_{ict}\}_c$, a $\#C_t \times 1$ vector of household channel preferences, as $\boldsymbol{\gamma}_{it} \equiv \boldsymbol{\chi}_{it} \cdot \tilde{\boldsymbol{\gamma}}_{it}$, where $\boldsymbol{\chi}_{it}$ is a vector whose components χ_{ict} are Bernoulli random variables (i.e., 0 or 1) that equals 0 with probability ρ_{ct}, and $\tilde{\boldsymbol{\gamma}}_{it}$ is a vector where each random component $\tilde{\gamma}_{ict}$ is drawn from an exponential distribution with parameter σ_{ct}.

For RSNs, we scale $\tilde{\gamma}_{ict}$ by $\exp(-\gamma^b b_{ict} - \gamma^d d_{ic})$, where $b_{ict} \in [0,1]$ represents the fraction of teams carried on RSN c that are "blacked-out" (i.e., unable to have games televised in household i's market due to restrictions imposed by the team's league) and d_{ic} is the average distance from household i to the stadiums for the teams shown on RSN c (measured in thousands of miles).[18] These terms allow for households to value an RSN differentially if the household cannot watch some of the carried sport teams, or if the household lives further away from the carried teams' stadiums.

3.3 Stage 2: Household Bundle Choice

Household i considers characteristics of each bundle—including the utility obtained from watching channels in the bundle and its price—when determining which firm, if any, to subscribe to. We specify household i's indirect utility conditional on subscribing to f as:

$$u_{ift} = \beta^v v^*_{ift} + \boldsymbol{\beta}^x \boldsymbol{x}_{ft} + \beta^{sat}_{if} + \alpha p_{ft} + \xi_{ft} + \varepsilon_{ift} \tag{2}$$

where v^*_{if} is the indirect utility from the time allocation problem in (1), $\boldsymbol{x}_{ft}$ are firm dummy and year dummy variables, p_{ft} is the per-month subscription fee for bundle f, and ξ_{ft} is a scalar unobservable demand shock for bundle f. Each consumer has a random preference for each satellite provider, β^{sat}_{if}, which is drawn from an independent exponential distribution with parameter ρ^{sat}_f; we assume that $\beta^{sat}_{if} = 0$ if f is a cable provider. We assume that ε_{ift} is distributed Type I extreme value, that the outside option of no bundle is normalized to $u_{i0} = 0$, and that each household chooses the bundle with the highest value of u_{if}.

The probability that household i subscribes to bundle f in market m is obtained by integrating

---

[17] Allowing for this parameter to differ between sports and non-sports channels is motivated by the observation that sports channels receive higher affiliate fees than national channels for the same viewership ratings; we discuss this futher in Section 4.1.2.

[18] We focus only on blackout restrictions for MLB, NBA, and NHL teams. We ignore the NFL in our analysis since its games have only been aired by national channels since the 1960s (CBS, NBC, Fox, and ESPN currently own its television rights).

over $\{\varepsilon_{ift}\}$ for each household:

$$s_{ifmt} = \frac{\exp(\beta^v v^*_{ift} + \boldsymbol{\beta}^x \boldsymbol{x}_{ft} + \beta^{sat}_{if} + \alpha p_{ft} + \xi_{ft})}{1 + \sum_{k \in \mathcal{F}_{mt}} \exp(\beta^v v^*_{ikt} + \boldsymbol{\beta}^x \boldsymbol{x}_{kt} + \beta^{sat}_{ik} + \alpha p_{kt} + \xi_{kt})} \, . \tag{3}$$

The total market share of each bundle f (in market m at time t) is then $s_{fmt} \equiv \int_i s_{ifmt} dH_{mt}(i)$, where $H_{mt}(i)$ is the joint distribution of household random coefficients $(\boldsymbol{\gamma}, \boldsymbol{\beta})$ in the market, and the demand for the bundle is $D_{fmt} \equiv N_{mt} s_{fmt}$, where N_{mt} is the number of television households in m.

3.4 Stage 1: Affiliate Fee Bargaining, Distributor Pricing, and Bundling

In Stage 1, all distributors and channel conglomerates bargain over affiliate fees $\{\tau_{fct}\}_{\forall f,c}$, where τ_{fct} represents the affiliate fee that distributor f pays the owner of channel c for each of f's household subscribers that receives channel c. Simultaneously, all distributors choose the prices and composition of each of its bundles.[19] That is, we assume that bargaining occurs simultaneously with distributor pricing and bundling.[20,21] We assume that affiliate fees, bundle prices, and bundle compositions are optimal with respect to one another in equilibrium.[22]

3.4.1 Stage 1a. Distributor Pricing and Bundling

Every distributor $f \in \mathcal{F}_t$ chooses prices and bundles $\{p_{fmt}, \mathcal{B}_{fmt}\}_{\forall m: f \in \mathcal{F}_{mt}}$ to maximize its profits given anticipated negotiated affiliate fees $\boldsymbol{\tau}_t \equiv \{\tau_{fct}\}_{\forall f,c}$. Profits for f across all markets are:

$$\Pi^M_{ft}(\{\mathcal{B}_{mt}\}_m, \{\boldsymbol{p}_{mt}\}_m, \boldsymbol{\tau}_t; \mu) \quad = \quad \sum_{m: f \in \mathcal{F}_{mt}} \Pi^M_{fmt}(\mathcal{B}_{mt}, \boldsymbol{p}_{mt}, \boldsymbol{\tau}_t; \mu)$$

---

[19] A given cable distributor f often operates in many markets, and is choosing prices and bundle composition in each of these markets. Satellite firms choose a single national price and channel bundle, with the only potential variation across DMAs being the set of RSNs that are carried.

[20] See also Nocke and White (2007) and Draganska et al. (2010) who use a similar timing assumption. Formally, one can think of separate agents of the distributor bargaining and making the pricing and bundle composition decisions. This sort of timing is also implicit in the analysis described in Rogerson (2014).

[21] An alternative timing assumption would be to assume that affiliate fees are first negotiated, and then distributor prices and bundles are chosen. This would adjust firms perceptions of off-equilibrium actions: e.g., when bargaining, firms would anticipate different bundle prices to immediately be set if off-equilibrium affiliate fees or disagreement were realized. However, there may be reasons to believe that such a rapid response is unrealistic. Absent a fully specified dynamic model of firm bargaining and pricing, which is outside the scope of the current analysis, we believe the approach taken here to be a reasonable approximation. We leverage this assumption to simplify the computation and estimation of our model.

[22] A distributor's optimal carriage decisions for an RSN are indeterminate when no deal is reached for that RSN (i.e., whether or not the distributor would carry the RSN on one of its systems in the event it were available is irrelevant when no deal is reached). In our estimation, we assume that satellite providers, who offer only a single national bundle, adopt the strategy of carrying any channel for which it has negotiated a deal (intuitively, since any deal that is reached should make carriage profitable).

where:

$$\Pi^M_{fmt}(\mathcal{B}_{mt}, \boldsymbol{p}_{mt}, \boldsymbol{\tau}_t; \mu) \quad = \quad D_{fmt}\Big(p_{fmt} - mc_{fmt}\Big) + \mu\Big(\sum_{c \in \mathcal{V}_{ft}} O_{fct} \sum_{g \in \mathcal{F}_{mt}: c \in \mathcal{B}_{gmt}} D_{gmt}(\tau_{gct} + a_{ct})\Big) \quad (4)$$

We denote by $\mathcal{B}_{mt} \equiv \{\mathcal{B}_{fmt}\}_{f \in \mathcal{F}_{mt}}$ and $\boldsymbol{p}_{mt} \equiv \{p_{fmt}\}_{f \in \mathcal{F}_{mt}}$ the set of bundles and associated prices offered in the market, and by a_{ct} the expected advertising revenue obtained by channel c per subscriber to a bundle containing c. The term $\mathcal{V}_{ft}$ represents the set of channels owned by MVPD f in period t, and the term O_{fct} represents MVPD f's ownership share of channel c at time t.[23]

The first component of an MVPD's profit function in a given market m, given by (4), is standard: each bundle has a price and a marginal cost (mc_{fmt}) that determine margins, and this is multipled by demand. We assume that each MVPDs' marginal cost in market m can be decomposed into the sum of the per-subscriber fees that f must pay to the various channels in its market-bundle, and a bundle-specific cost shock that is the sum of non-channel related marginal costs, denoted by ω_{fmt}: i.e., $mc_{fmt} \equiv \sum_{c \in \mathcal{B}_{fmt}} \tau_{fct} + \omega_{fmt}$.[24] The second component of the profit function is non-standard, and represents the degree to which a vertically integrated downstream unit values the profits that accrue to its upstream (i.e., channel) units. These terms include per-subscriber fees and advertising revenues that accrue to integrated upstream channels from its own viewers as well as from viewers of other distributors.[25] The parameter $\mu \in [0, 1]$ represents the extent to which a downstream MVPD f internalizes upstream affiliate fees and advertising revenues from its integrated channels $c \in \mathcal{V}_{ft}$.

In the absence of any frictions, μ would equal one, implying that the downstream firm perfectly internalizes integrated upstream unit profits, and its strategic decisions maximize total firm profit. Parameter μ could also be less than one, potentially representing divisionalization that could arise from ignorance, poor management, optimal compensation under informational frictions, or any other conflict between managers of different divisions within the same firm.

Optimal Pricing and Bundling. We will leverage necessary conditions on the optimality of MVPD pricing and bundling decisions in our estimation. Differentiating (4) with respect to p_{fmt} (and dividing by market size) yields the following pricing first-order condition:

$$\frac{\partial \Pi^M_{fmt}}{\partial p_{fmt}} = s_{fmt} + \Big(p_{fmt} - mc_{fmt}\Big)\frac{\partial s_{fmt}}{\partial p_{fmt}} + \mu\Big(\sum_{c \in \mathcal{B}_{gmt} \cap \mathcal{V}_{ft}} O_{fct} \sum_{g \in \mathcal{F}_{mt}} (\tau_{gct} + a_{ct})\frac{\partial s_{gmt}}{\partial p_{fmt}}\Big) \qquad (5)$$

---

[23] For our analysis, we only include in $\mathcal{V}_{ft}$ the set of integrated RSNs. We will assume that $c \in \mathcal{V}_{ft}$ (and hence, c is integrated with f) if MVPD f owns any percentage of channel c in period t. In the case that a third party has an $x\%$ stake in MVPD f and $y\%$ stake in channel c at time t, we assume that $O_{fct} = x\% \times y\%$. This can be interpreted as the third party having an $x\%$ probability of making strategic decisions on behalf of the MVPD.

[24] Cost shocks include changes in variable costs such as technical service labor, gasoline, and equipment costs that are incurred on a per-subscriber basis.

[25] We omit portions of integrated channels' profits which are not affected by f's pricing and carriage decisions, as they do not affect the analysis. We also assume that channel c's per-subscriber advertising revenues in market m do not vary across MVPDs, and that channel c's marginal costs per-subscriber are zero.

In addition, we assume that the set of channels actually offered by each MVPD f in each market m satisfies:

$$\mathcal{B}_{fmt} = \arg \max_{\mathcal{B}_f \subseteq \mathcal{A}_f} \Pi^M_{fmt}(\{\mathcal{B}_f, \{\mathcal{B}_{gmt}\}_{g \neq f}\}, \boldsymbol{p}_{mt}, \boldsymbol{\tau}_t; \mu) \tag{6}$$

where $\mathcal{A}_{ft} \subseteq \mathcal{C}_t$ is the set of channels available to MVPD f: i.e., the set of channels for which f has reached an agreement.[26]

Satellite Pricing and Bundling. If distributor f is a satellite MVPD (DirecTV or Dish), we assume that the distributor sets a single national price and bundle. We assume that the bundle offered by a satellite MVPD in any given market may differ from the national bundle only in the set of RSN channels that are offered.

3.4.2 Stage 1b: Bargaining over affiliate fees

Before describing how affiliate fees are determined, we specify the profits each channel c contemplates when bargaining with MVPD f in market m as:

$$\begin{aligned}\Pi^C_{cmt}(\mathcal{B}_{mt}, \boldsymbol{p}_{mt}, \boldsymbol{\tau}_t; \mu, \lambda_R) = \sum_{g \in \mathcal{F}_{mt}: c \in \mathcal{B}_{gmt}} D_{gmt} \Bigg(& \tau_{gct} + a_{ct} \ \ldots \\ & + \mu \times \lambda_{R:fct} \Big(O_{gct}(p_{gmt} - mc_{gmt}) + \sum_{d \in \mathcal{B}_{gmt} \setminus c} O^C_{cdt}(\tau_{gdt} + a_{gdt}) \Big) \Bigg)\end{aligned} \tag{7}$$

The first line reflects affiliate fees and advertising revenues obtained from each bundle the channel is available on; the second line incorporates potential profits of an integrated downstream MVPD, as well as profits from other channels also owned by the same owner of channel c. We denote by O^C_{cdt} the common ownership percentage of two channels c and d by a third-party.[27]

Both terms on the second line are multiplied by μ and $\lambda_{R:fct}$, where:

$$\lambda_{R:fct} = \begin{cases} 1 & \text{if } f \text{ and } c \text{ are integrated (i.e., } O_{fct} > 0) \text{ ,} \\ \lambda_R & \text{if } f \text{ and } c \text{ are not integrated .} \end{cases}$$

We assume that $\lambda_{R:fct} = 1$ if c is owned by f and is bargaining with f; this implies that a channel and distributor that are integrated with each other place equal weight (given by μ) on each other's profits when bargaining with each other. However, if c is integrated but bargaining with a rival distributor (i.e., the MVPD that c is bargaining with, f, is not an owner of c), then $\lambda_{R:fct} = \lambda_R \geq 0$; thus λ_R governs the extent to which an integrated upstream unit recognizes and internalizes the effects of foreclosing the rival MVPD on the profits of its other integrated units.

In Figure 2, we provide an illustration of how channel c's perceived profits when bargaining with MVPD f may change depending on whether or not it is integrated with f. In Figure 2a, c is

---

[26] See footnote 22 regarding our treatment of channels not contained in $\mathcal{A}_{ft}$.

[27] Specifically, if each owner $j \in \mathcal{J}$ of channel c owns shares x_j of c and y_j of channel d, then $O^C_{cdt} \equiv \sum_{j \in \mathcal{J}} x_j y_j$.

$\Pi^C_{cmt} = \pi_{cmt} + \mu(\pi_{fmt} + \pi_{dmt})$ $\qquad$ $\Pi^C_{cmt} = \pi_{cmt} + \mu \times \lambda_R(\pi_{gmt} + \pi_{dmt})$

(a) Internal Bargaining. (b) Bargaining with Rival MVPD.

Figure 2: Examples of Π^C_{cmt} when c bargains with MVPD f.

integrated with MVPD f and another channel d (represented by the dashed square); in this case, c will consider when bargaining with f its own profits (denoted by π_{cmt}), consisting of affiliate fees and advertising revenues, as well as profits of its integrated distributor f and channel d (denoted by π_f and π_d) weighted by μ. We assume that π_{fmt} includes f's subscription revenues net its costs; profits π_{dmt} include d's affiliate fees and advertising revenues. In Figure 2b, c is integrated with another MVPD g and d; in this case, c will consider when bargaining with f (a rival MVPD) its own profits π_{cmt}, and those of its integrated units π_{fmt} and π_{dmt} weighted by $\mu \times \lambda_R$.

The parameter λ_R captures the extent to which an upstream unit has incentives to foreclose access to the RSN to a rival distributor and lower the rivals' bundle quality (thereby shifting demand to the integrated distributor), an effect analogous to the "raising-rivals'-cost" effect discussed in Salop and Scheffman (1983) and Krattenmaker and Salop (1986). We thus refer to λ_R as our "rival-foreclosure" or "raising-rivals'-costs" (RRC) parameter.

We assume that, given channel c is carried on some of MVPD f's systems, the affiliate fee τ_{fct} between distributor f and channel c maximizes their respective bilateral Nash products *given the expected negotiated affiliate fees of all other pairs and the expected prices and bundles for all distributors*:

$$\hat{\tau}_{fct}(\boldsymbol{\tau}_{-fc,t}, \mathcal{B}_t, \boldsymbol{p}_t) = \arg\max_{\tau_{fct}} \Bigg[\underbrace{\sum_m [\Delta_{fc}\Pi^M_{fmt}(\mathcal{B}_{mt}, \boldsymbol{p}_{mt}, \{\tau_{fct}, \boldsymbol{\tau}_{-fc,t}\}; \mu)]}_{GFT^M_{fct}} \Bigg]^{\zeta_{fct}} \times \Bigg[\underbrace{\sum_m [\Delta_{fc}\Pi^C_{cmt}(\mathcal{B}_{mt}, \boldsymbol{p}_{mt}, \{\tau_{fct}, \boldsymbol{\tau}_{-fc,t}\}; \mu, \lambda_R)]}_{GFT^C_{fct}} \Bigg]^{1-\zeta_{fct}} \tag{8}$$

where:

$$[\Delta_{fc}\Pi^M_{fmt}(\mathcal{B}_{mt}, \cdot)] \equiv \Big(\Pi^M_{fmt}(\mathcal{B}_{mt}, \cdot) - \Pi^M_{fmt}(\mathcal{B}_{mt} \setminus fc, \cdot)\Big)$$
$$[\Delta_{fc}\Pi^C_{cmt}(\mathcal{B}_{mt}, \cdot)] \equiv \Big(\Pi^C_{cmt}(\mathcal{B}_{mt}, \cdot) - \Pi^C_{cmt}(\mathcal{B}_{mt} \setminus fc, \cdot)\Big)$$

and $\zeta_{fct} \in [0, 1]$ represents a firm-channel-time specific Nash bargaining parameter.

We denote by $\mathcal{B}_{mt} \setminus fc$ the set of all bundles $\mathcal{B}_{mt}$ in which we remove channel c from bundle f. Thus, these terms represent the difference in either MVPD or channel profits in market m if f no longer carries channel c. We will refer to GFT^M_{fct} and GFT^C_{fct}, which is the sum of these terms across all markets, as the *gains from trade* (or *bilateral surplus*) for MVPD f and channel c coming to an agreement.

This bargaining solution in which each pair of distributors and channels agree upon a set of affiliate fees that maximize the Nash product of their gains from trade is motivated by the model put forth in Horn and Wolinsky (1988), and used by Crawford and Yurukoglu (2012) to model negotiations between MVPDs and channel conglomerates.[28] Each MVPD and conglomerate negotiate a single affiliate fee per channel that applies to all markets.

We can write the first-order condition of (8) for each channel c bargaining with MVPD f as:

$$\zeta_{fct} GFT^C_{fct} \left(\frac{\partial GFT^M_{fct}}{\partial \tau_{fct}}\right) + (1-\zeta_{fct}) GFT^M_{fct} \left(\frac{\partial GFT^C_{fct}}{\partial \tau_{fct}}\right) = 0 \tag{9}$$

where the derivative terms in (9) are:

$$\begin{aligned} \frac{\partial GFT^M_{fct}}{\partial \tau_{fct}} &= \sum_m \frac{\partial \Pi^M_{fmt}}{\partial \tau_{fct}} = (-1 + (\mu \times O_{fct})) \sum_{m \in \mathcal{M}_{fct}} D_{fmt} \\ \frac{\partial GFT^C_{fct}}{\partial \tau_{fct}} &= \sum_m \frac{\partial \Pi^C_{cmt}}{\partial \tau_{fct}} = (1 - (\mu \times \lambda_{R:fct} \times O_{fct})) \sum_{m \in \mathcal{M}_{fct}} D_{fmt} \end{aligned}$$

and $\mathcal{M}_{fct} \equiv \{m : c \in \mathcal{B}_{fmt}\}$ denotes the set of markets where c is on f's bundle. As we have assumed that $\lambda_{R:fct} = 1$ whenever $O_{fct} > 0$ (i.e., f and c are integrated and bargaining with one another), it follows that $\partial GFT^M_{fct}/\partial \tau_{fct} = -\partial GFT^C_{fct}/\partial \tau_{fct}$. We can thus re-write (9) as:

$$(\zeta_{fct}) GFT^C_{fct} = (1-\zeta_{fct}) GFT^M_{fct} \qquad \forall f, c \ s.t. \ \exists m : c \in \mathcal{B}_{fmt}. \tag{10}$$

This bargaining solution is not defined if $\mu \times O_{fct} = 1$; under this case, f and c would perfectly internalize each other's profits when bargaining with one another, and the negotiated τ_{fct} would be indeterminate. Also, in deriving (10), we are leveraging the assumption that distributor bundle prices are set simultaneously with affiliate fees, and there is no anticipated change in p_{fmt} if τ_{fct} changes. Nonetheless, in equilibrium, both prices and affiliate fees will satisfy the pricing first-order conditions given by (5) and the bargaining first-order conditions in (10).

The Role of λ_R. In our model, $\mu \times \lambda_R$ captures the internalization of an integrated downstream MVPD's profits when an integrated channel bargains with another distributor. Consider channel c owned by MVPD f bargaining with rival distributor g (e.g., a satellite distributor). When $\lambda_R > 0$,

---

[28] Other empirical papers that use use this bargaining solution include Grennan (2013), Gowrisankaran et al. (forthcoming), and Ho and Lee (2013); Collard-Wexler et al. (2014) provide a non-cooperative foundation for this bargaining solution.

c's desire to the increase downstream profits of f lowers c's gains from trade when bargaining with the non-integrated rival distributor g compared to when $\lambda_R = 0$. This may lead to the elimination of overall gains from trade, and can result in non-supply of c to g. However, even if there are still positive gains from trade, since these gains will be lower for c when $\lambda_R > 0$, the bargaining process will lead to an increased affiliate fee (τ_{gct}) for the rival distributor. Thus, even if g is still supplied with channel c, its costs are raised; in equilibrium, this can lead the rival to increase the price of its bundles to consumers.

Example. Consider the case in which MVPD f and channel c are both non-integrated entities that bargain with one another in period t. The negotiated affiliate fee τ_{fct} that satisfies the Nash bargaining solution given by (10) solves:

$$\begin{aligned}\sum_{m \in \mathcal{M}_{fct}} D_{fmt}\tau_{fct} = (1-\zeta_{fct}) & \sum_{m \in \mathcal{M}_{fct}} \Big([\Delta_{fc} D_{fmt}](p_{fmt} - mc_{fmt} + \tau_{fct}) \Big) \\ -(\zeta_{fct}) & \sum_{m \in \mathcal{M}_{fct}} \Big(D_{fmt} a_{ct} + \sum_{g \neq f : c \in \mathcal{B}_{gmt}} [\Delta_{fc} D_{gmt}] \big(\tau_{gct} + a_{ct}\big) \Big)\end{aligned} \tag{11}$$

where $[\Delta_{fc} D_{gmt}] \equiv D_{gmt}(\mathcal{B}_{mt}, \cdot) - D_{gmt}(\mathcal{B}_{mt} \setminus fc, \cdot)$ denotes the change in firm g's demand in market m and time t if channel c was removed from firm f's bundle.

The left hand side of (11) is the total payment made by f to c. The right hand side is a fraction of the gains from trade due to agreement, where the first term represents f's increased profits (net of payments to c) due to more subscribers induced by the carriage of channel c, and the remaining terms on the second-line represent (the negative of) c's gains from being carried on f. Intuitively, the more f gains from the relationship, the higher the total payment that is made; the more c gains from the relationship, the lower the total payment. If f and c's Nash bargaining parameters were equal, then $\zeta_{fct} = 1/2$ and these gains from trade would be split in half.

4 Estimation and Identification

In this section, we discuss the estimation of our model's parameters and how they are identified (given our modeling assumptions) from patterns in the data. We proceed in two stages:

1. In the first stage, we estimate $\boldsymbol{\theta} \equiv \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3\}$, where:
    (a) $\boldsymbol{\theta}_1 \equiv \{\boldsymbol{\Sigma}, \boldsymbol{\nu}, \boldsymbol{\rho}, \gamma^d, \gamma^b\}$, where $\boldsymbol{\Sigma} \equiv \{\sigma_{ct}\}_{\forall c}$, $\boldsymbol{\nu} \equiv \{\nu^S, \nu^{NS}\}$, and $\boldsymbol{\rho} \equiv \{\rho_{ct}\}_{\forall c,t}$, determines household viewership decisions by governing the distribution of $\boldsymbol{\gamma}$ and how fast marginal utilities from viewership decay.
    (b) $\boldsymbol{\theta}_2 \equiv \{\beta^v, \boldsymbol{\beta}^x, \boldsymbol{\rho}^{sat}, \alpha\}$, where $\boldsymbol{\rho}^{sat} \equiv \{\rho^{sat}_{DirecTV}, \rho^{sat}_{Dish}\}$, determines household bundle choice.
    (c) $\boldsymbol{\theta}_3 \equiv \{\mu\}$ represents the extent to which integrated conglomerates and distributors internalize profits across upstream and downstream units when pricing, bargaining, and

choosing other strategic variables.

Initially, we assume that $\zeta_{fct} = 1/2 \; \forall f, c, t$, and that distributors and channels have the same Nash bargaining parameters.

2. In the second stage, we estimate our RRC parameter, λ_R.

To capture the impact of program access rules, we will assume that $\lambda_R = 0$ in non-loophole markets and estimate our first stage parameters using only these markets. That is, we assume that the program access rules effectively require integrated firms to ignore any foreclosure incentives in dealing with non-integrated rivals. We then estimate λ_R using only the markets in our data in which the terrestrial loophole was used by RSNs (i.e., Philadelphia and San Diego).

4.1 First Stage Estimation

4.1.1 Moments used in Estimation

We estimate the model parameters via GMM, using the following moments derived from the model described in the previous section.

Household Viewership. For every RSN and 38 national channels in each year, we use the difference between the following viewership moments observed in the data and predicted by the model:

1. Summing across markets, the mean viewership for each channel-year;

2. Summing across markets, the number of households with zero viewership for each (non-RSN) channel-year.

To avoid re-solving the viewership problem for every household for every evaluation of a candidate parameter vector, we follow the importance sampling approach of Ackerberg (2009).

Household Bundle Choice. For every year and bundle, we assume that each bundle's unobservable characteristic is orthogonal to a vector of instruments: i.e., $E[\xi_{fmt}(\boldsymbol{\theta})\boldsymbol{Z}^{\xi}_{mt}] = 0$, where the expectation is taken across all markets, firms, and years. For $\boldsymbol{Z}_{mt}$, we include bundle observable characteristics $\boldsymbol{x}_{fmt}$ and predicted indirect utility of channel viewing v^*_{fmt} for the mean consumer; we also include the satellite tax within the market to instrument for p^o_{fmt}. We recover $\xi_{fmt}(\boldsymbol{\theta})$ using the standard Berry et al. (1995) inversion.

Distributor Bargaining, Pricing, and Carriage. First, for any $\boldsymbol{\theta}$, the vector of affiliate fees $\{\tau_{fct}\}$ and bundle-specific marginal costs $\{mc_{fmt}\}$ can be directly computed using the optimal pricing and bargaining conditions given by (5) and (10) (see Appendix for further details). We use these predicted values of $\{mc_{fmt}(\boldsymbol{\theta})\}$ and $\{\tau_{fct}(\boldsymbol{\theta})\}$ in constructing the next set of moments which we form only using 2007 data and values:

1. **Average affiliate fees**: For each RSN active in 2007 and four national channels (ABC Family, ESPN, TNT, and USA), we minimize the difference between the model's predicted average affiliate fees across MVPDs and observed average affiliate fees (τ^o_{ct}):

$$E_f[\tau_{fct}(\boldsymbol{\theta})] - \tau^o_{ct} = \omega^C_{ct}$$

   where deviations, denoted by ω^C_{ct}, reflect measurement error in τ_c. We weight estimated affiliate fees by national MVPD market shares conditional on carriage of the channel to approximate expectations across MVPDs.

2. **Implied markups**: The model's predicted MVPD price-cost markups should match those observed in the data:

$$E_m[(p^o_{fmt} - mc_{fmt}(\boldsymbol{\theta}))/p^o_{fmt}] = markup^o_{ft} \qquad \forall f \in \{Comcast, DirecTV, Dish\}\ .$$

3. **Bundle Optimality and Carriage**: Equation (6) implies that every distributor f chooses the optimal set of channels to include in each bundle in each market m. We will assume that distributor f's true per-household profits (not per-subscriber) in market m are given by $\tilde{\pi}^M_{fmt}(\cdot)$, where:

$$\tilde{\pi}^M_{fmt}(\mathcal{B}_{mt}, \cdot) \equiv [\pi^M_{fmt}(\mathcal{B}_{mt}, \cdot) - \nu^1_{fmt}(\mathcal{B}_{fmt})] + \sum_{c \in \mathcal{B}_{fmt}} \nu^2_{ct}\ . \tag{12}$$

   and $\pi^M_{fmt}(\mathcal{B}_{mt}, \cdot)$ represents our (the econometrician's) estimate of a firm's per-household profits. We introduce two types of disturbances in this definition: the first, $\nu^1_{fmt}(\cdot)$, represents a mean-zero i.i.d. bundle-distributor-market-time specific disturbance which captures potential measurement or specification error between our estimate of a firm's profits and that used by the firm; the second, ν^2_{ct}, is a mean zero channel-time specific disturbance that is known to the distributor when making its carriage decision (but not known during the bargaining stage), unobserved to the econometrician, and may include non-measured per-household fixed incentives or costs of carrying a channel.

   Now consider a channel c that has negotiated an agreement with some firm f: i.e., f carries c on its bundles in some non-empty set of markets. A firm's optimal bundling decision given by (6) implies that:

$$\begin{aligned} [\Delta_{fc}\pi^M_{fmt}(\mathcal{B}_{mt}, \cdot)] - [\Delta_{fc}\nu^1_{fmt}(\mathcal{B}_{fmt})] + \nu^2_{ct} &\geq 0 \qquad \forall m : c \in \mathcal{B}_{fmt} \\ -\Big([\Delta_{fc}\pi^M_{fmt}(\mathcal{B}_{m't} \cup fc, \cdot)] - [\Delta_{fc}\nu^1_{fm't}(\mathcal{B}_{fm't} \cup fc, \cdot)] + \nu^2_{ct}\Big) &\geq 0 \qquad \forall m' : c \notin \mathcal{B}_{fm't} \end{aligned}$$

   where $[\Delta_{fc}\pi^M_{fmt}(\mathcal{B}_{mt}, \cdot)] \equiv \pi^M_{fmt}(\mathcal{B}_{mt}, \cdot) - \pi^M_{fmt}(\mathcal{B}_{mt} \setminus fc, \cdot)$, $[\Delta_{fc}\nu^1_{fmt}(\mathcal{B}_{fmt})] \equiv \nu^1_{fmt}(\mathcal{B}_{fmt}) - \nu^1_{fmt}(\mathcal{B}_{fmt} \setminus fc)$, and $\mathcal{B}_{m't} \cup fc$ denotes the set of all bundles $\mathcal{B}_{m't}$ where c is added to bundle f. That is, these inequalities imply that in any market in which c is carried by f, f obtains higher profits from carrying than by dropping c (holding fixed prices and carriage decisions of other firms); similarly, in any market where c is not carried, f obtains higher profits from not carrying than by carrying c.

This implies for any fm and $f'm'$ pair such that $c \in \mathcal{B}_{fmt}$ and $c \notin \mathcal{B}_{f'm't}$ and both MVPDs f and f' have an agreement with c, the two inequalities above can be added together to yield:

$$[\Delta_{fc}\pi_{fmt}(\mathcal{B}_{mt},\cdot)]-[\Delta_{f'c}\pi_{fmt}(\mathcal{B}_{m't}\cup f'c,\cdot)]-([\Delta_{fc}\nu^1_{fmt}(\mathcal{B}_{fmt})]-[\Delta_{f'c}\nu^1_{f'm't}(\mathcal{B}_{f'm't}\cup f'c,\cdot)]) \geq 0\ .$$

where the ν^2_{ct} disturbances cancel out. Thus, given our assumptions on the distribution of $\{\nu^1_{fct}(\cdot)\}$, for each firm f and RSN c with agreement,

$$\begin{aligned} E_{m\in\mathcal{M}^+_{fct}}\Big[[\Delta_{fc}\pi_{fmt}(\mathcal{B}_{fmt},\cdot)] - [\Delta_{f'c}\pi_{f'm'(fm),t}(\mathcal{B}_{f'm'(fm),t}\cup fc)]\Big] &\geq 0 \\ E_{m\in\mathcal{M}^-_{fct}}\Big[-[\Delta_{fc}\pi_{fmt}(\mathcal{B}_{fmt}\cup fc,\cdot)] + [\Delta_{f'c}\pi_{f'm'(fm),t}(\mathcal{B}_{f'm'(fm),t})]\Big] &\geq 0 \end{aligned}$$

where $\mathcal{M}^+_{fct}$ denotes the set of markets in which f is active and carries channel c, $\mathcal{M}^-_{fct}$ denotes those markets in which f is active but does not carry c, and $f'm'(fm)$ denotes a firm-market pair where f' has an agreement with c but has the opposite carriage decision as firm f in market m for c (i.e., if $c \in \mathcal{B}_{fmt}$, then $c \notin \mathcal{B}_{f'm'(fm),t}$, and vice versa). These inequalities imply that the summed change in f's per-household profits in market m and f''s profits in market m' (where either f carries c in m or f' carries c in m'), when reversing the observed carriage decisions in both markets and averaging across all markets m in which f either carries or doesn't carry c, is positive. If these inequalities did not hold, it would imply that either f or f' would have a profitable deviation by changing its carriage decisions for c in certain markets.

These inequalities motivate maximizing the following moments in estimation:

$$\sum_{f\in\mathcal{F}_t}\frac{1}{\#(\mathcal{M}^+_{fct})}\left[\sum_{m\in\mathcal{M}^+_{fct}}[\Delta_{fc}\pi_{fmt}(\mathcal{B}_{fmt},\cdot)] - [\Delta_{f'c}\pi_{f'm'(fm),t}(\mathcal{B}_{f'm'(fm),t}\cup fc,\cdot)]\right]_- \quad \forall c \tag{13}$$

$$\sum_{f\in\mathcal{F}_t}\frac{1}{\#(\mathcal{M}^-_{fct})}\left[\sum_{m\in\mathcal{M}^-_{fct}}-[\Delta_{fc}\pi_{fmt}(\mathcal{B}_{fmt}\cup fc,\cdot)] + [\Delta_{f'c}\pi_{f'm'(fm),t}(\mathcal{B}_{f'm'(fm),t},\cdot)]\right]_- \quad \forall c \tag{14}$$

where $[\cdot]_- \equiv \min\{\cdot, 0\}$. We choose $f'm'(fm)$ to be a firm-market pair such that: $f' \neq f$, $f'm'$ has the opposite carriage decision for c as firm-market pair fm, and m' is the closest market (in Euclidean distance) to market m in terms of (weighted) distance to the teams carried on

RSN c and fraction of teams on c that are blacked out.[29,30] We construct these moments for each RSN active in 2007. These sets of moments are similar to those used in Crawford and Yurukoglu (2012) and utilizes insights from Pakes et al. (forthcoming).

4.1.2 Identification

We now provide an informal discussion of how the parameters of the model are identified from these moments.

The main parameters governing the distribution of γ_{ict} (i.e., $\mathbf{\Sigma}, \boldsymbol{\rho}$) are primarily identified from viewing behavior: e.g., channels watched more often have higher values of γ_{ict} and lower values of ρ_{ct}. However, since we do not possess ratings for channels at the system level, we identify the black-out and distance parameters γ^b, γ^d primarily from the Bundle Optimality and Carriage moments; we defer discussion of these parameters until the end of this subsection when discussing identification of μ.

Parameters governing household bundle choice, $\boldsymbol{\beta}^x$ and $\boldsymbol{\beta}^v$, are identified from variation in bundle market shares as observed bundle characteristics and channel utility changes: i.e., across firms and years, and as channels are added and dropped from bundles. The satellite tax is an instrument for price, and is used to identify the price sensitivity coefficient α. Information contained in cable and satellite pricing margins helps identify the heterogeneity in preferences for satellite. In particular, the relationship between satellite and cable market shares has strict implications for predicted price elasticities (and hence implied markups) under a standard logit demand system without preference heterogeneity; inclusion of a random preference for satellite (parameterized by $\boldsymbol{\rho}^{sat}$) assists with rationalizing observed markups for a given satellite market share.

In addition to observing how bundle market shares vary based on channel composition (which has limited variation for some channels across markets), matching observed average affiliate fees negotiated for each channel $\{\tau^o_{ct}\}$ to those predicted by the model $\{\tau_{fct}(\boldsymbol{\theta})\}$ is crucial. First, our model relates $\tau_{fct}(\boldsymbol{\theta})$ to the gains from trade created when channel c contracts with firm f: i.e., differences in f and c's profits (primarily realized from subscription and advertising revenues) when f drops c. Thus, our model attempts to rationalize a channel with higher observed affiliate fees τ^o_{ct} by predicting that this channel creates greater surplus from carriage: this is partly through the term $\beta^v v^*_{ift}$ in a household's bundle utility equation given by (2), which in turn is also a function of parameters governing the distribution of γ_{ict}, and how γ_{ict} is scaled to enter into utility by ν_c—i.e., a channel with a higher γ_{ic} and lower decay parameter ν_c than another will contribute more to a viewer's utility from the same amount of time the channel is watched.

To anchor this in an example, consider a single market and bundle with two channels c and

---

[29] The rationale for this matching procedure is to match markets with comparable magnitudes of profitability changes, and to be robust to the possibility that ν^2_{ct} might vary across dissimilar markets.

[30] In practice, we also double the set of inequalities in (13) and (14) by segmenting the set of markets for channel into those within and outside of 100mi to the stadiums of the teams carried by the RSN, and only matching any firm-market to other firm-markets within the same segment. The reason for this is to ensure that the magnitudes of profit violations are comparable: e.g., profit violations in markets far away from an RSN's teams' stadiums might be offset by larger profit violations in markets close to the stadiums.

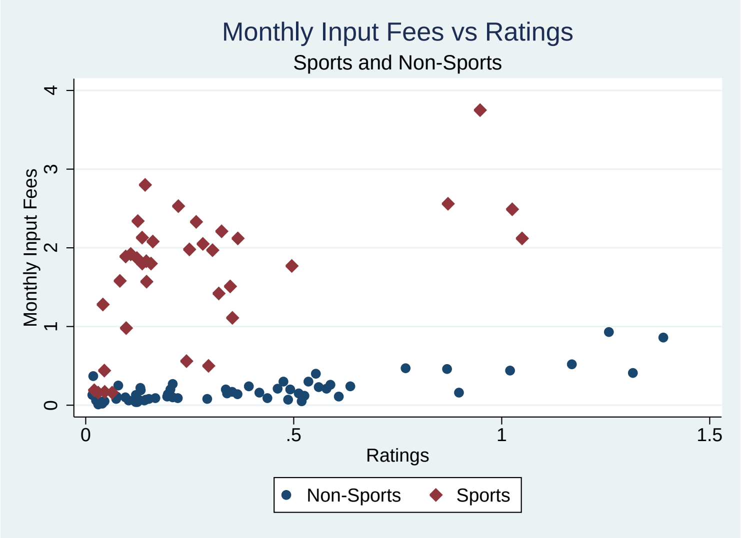


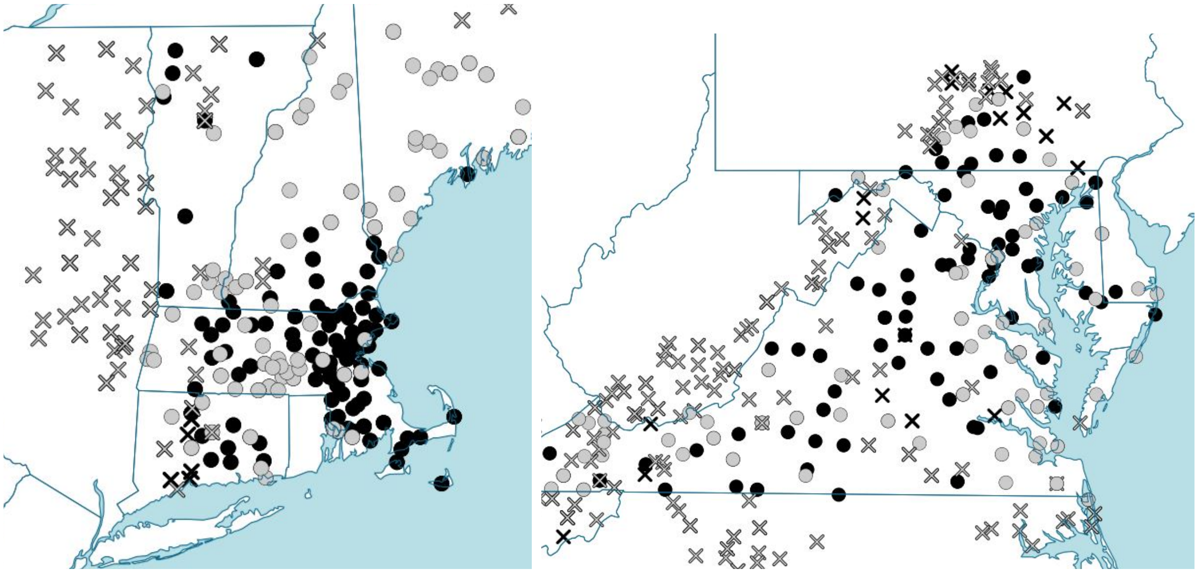
Figure 3: Negotiated monthly affiliate fees and viewership ratings.

d, and a single household i. Assume that the household watches d more than c. This could be induced by many potential combinations of $(\gamma_{ic}, \nu_c, \gamma_{id}, \nu_d)$; e.g., γ_{id} could be higher than γ_{ic} and $\nu_c = \nu_d$. If this were true, however, then d should obtain higher negotiated affiliate fees as it would be predicted to generate a higher surplus for a viewer, and hence there would be higher gains from trade from carriage of d than c. However, if affiliate fees are observed to be the same for the two channels despite the difference in viewership, then the model would predict that the rate of "decay" for channel c, ν_c, was in fact higher than ν_d (thereby allowing c to generate the same utility for consumers—and hence the same negotiated affiliate fees—for a shorter amount of time watched).

Now add to this example two additional markets: one market only has channel c available, and another only has channel d. If viewership patterns for these channels in the new markets were similar to those in the first market, then variation in market shares for the bundle across these markets as the channel composition of the bundles changed would inform the value of β^v.

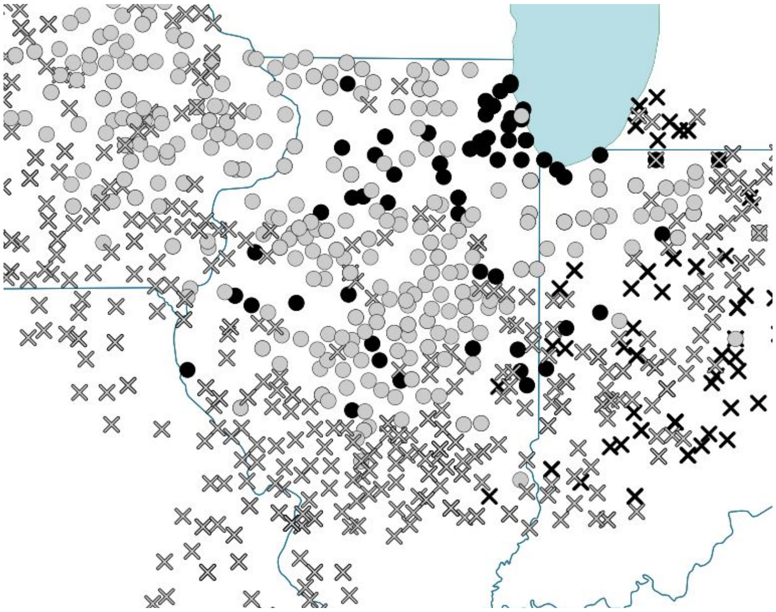
In a sense, the parameters governing the distributions of γ_{ic} and ν_c can be seen as helping the model rationalize variation in both negotiated affiliate fees and the market share of bundles as (both the mean and variance of observed) *viewership* of channels varies across markets, controlling for channel carriage; and β_v can be seen as helping the model rationalize variation in market shares of bundles as *channel carriage* changes across markets, holding fixed patterns of viewership for these channels.

The reason that we allow for consumers to possess two different "decay" parameters $\{\nu^S, \nu^{NS}\}$ for sports and non-sports channels is motivated by the data, illustrated in Figure 3. Sports channels have consistently higher negotiated affiliate fees than non-sports channels with similar viewership patterns (ratings), in cases receiving payments an order of a magnitude higher. Our model rationalizes this by assigning a higher decay rate to sports channels, which predicts higher utility delivered to consumers for a given amount of time the channel is watched; thus, sports channels are able to negotiate higher affiliate fees as they create greater gains-from-trade upon agreement with an MVPD.



(a) Carriage of CSN New England

(b) Carriage of CSN Mid-Atlantic



(c) Carriage of CSN Chicago

Figure 4: Carriage by Comcast and non-integrated cable MVPDs of three Comcast-integrated RSNs across cable systems in 2007. Circles represent carriage by a system, X's represent no carriage. Black markers represent Comcast systems, grey markers represent non-Comcast cable systems.

Although the internalization parameter μ enters into the computation of several moments (including any moment based off of recovered values of $\tau_{fct}(\boldsymbol{\theta})$ and $mc_{fmt}(\boldsymbol{\theta})$), it will primarily be identified off of the Bundle Optimality and Carriage moments. In particular, as μ increases, distributors have a greater incentive to carry an integrated channel for a fixed value $\tau_{fct}(\cdot)$; hence, the

Table 1: Regression of RSN Carriage on Integration Status, Distance, and Blackout Percentage

	Coeff.	SE	t
Integrated with RSN	0.166	0.065	2.56
Avg Distance to RSN's Stadiums (mi)	-0.001	0.000	-5.23
% Teams not Blacked Out	0.512	0.094	5.45
Fixed effects:	RSN, MSO, DMA		
R-squared	0.713		
N	1593		

Notes: Linear probability regression where the dependent variable is whether a system carries an RSN in 2007. SE's are clustered by RSN. Inclusion of system demographic controls (race, population density, average income, household ownership) did not change point estimates.

model will help to rationalize higher carriage rates between integrated distributors and channels (which is observed in the data). We identify our black-out and distance parameters, γ^b, γ^d, in a similar fashion. An example of the variation in the data that we leverage is illustrated in Figure 4, which presents the integrated and non-integrated carriage of a Comcast integrated RSN in three different regions of the U.S. In these three settings, moving further away from the RSN's teams' stadiums yields more likely carriage by a Comcast system of the channel (denoted by a black circle) than carriage by a non-Comcast system (denoted by a grey circle). For example, in Figure 4a, all Comcast systems in northern Vermont carry CSN NE whereas most non-Comcast systems do not (denoted by grey X's); and in Figure 4b and in Figure 4c, non-carriage by non-Comcast systems occurs much closer to the RSN's teams' stadiums than for Comcast systems (as there are more grey X's near Washington DC and Chicago than black X's, which denote non-carriage by Comcast systems). These maps also indicate that non-carriage is much more likely in areas where the teams on the RSN are blacked out (as in New York for CSN NE, Pennsylvania for CSN Mid-Atlantic, and Michigan for CSN Chicago).

Table 1 summarizes this relationship across all RSN's and distributors in our sample: carriage of an RSN by a cable system is strongly increasing with the RSN and distributor being integrated, and strongly decreasing in the distance between the system and the RSN's teams' stadiums and in the fraction of teams that are blacked out.

4.2 Second Stage Estimation

4.2.1 Recovery of λ_R

To recover our RRC parameter λ_R, we will use information provided by markets in which distributors are able to exclude competitors from carrying an integrated RSN channel—i.e., terrestrial loophole markets. The markets we focus on will be Philadelphia and San Diego, the channels in question CSN Philadelphia (owned by Comcast) and 4SD (owned by Cox), and the competitors excluded from carriage are satellite providers DirecTV and Dish.

To describe our approach, consider a channel c that is integrated with cable distributor f and that is "relevant" (i.e., offered and plausibly available to some set of distributors) in markets $\mathcal{M}_c$.[31] If we observe that channel c does not contract with satellite distributor $g \neq f$, we will assume that λ_R must have been sufficiently large for c and g to not contract with one another to be an equilibrium outcome. A *necessary* condition for this is that there is no affiliate fee $\tilde{\tau}_{gct}$ such that c and g would both find it profitable to contract with one another:

$$\sum_{m \in \mathcal{M}_c} \Bigg[\underbrace{[\Delta_{gc}\Pi^M_{gmt}(\{\mathcal{B}^o_{mt} \cup gc\}, \boldsymbol{p}^o_{mt}, \tilde{\boldsymbol{\tau}}; \hat{\mu})]}_{GFT^M_{gcmt}} + \underbrace{[\Delta_{gc}\Pi^C_{cmt}(\{\mathcal{B}^o_{mt} \cup gc\}, \boldsymbol{p}^o_{mt}, \tilde{\boldsymbol{\tau}}; \hat{\mu}, \lambda_R)]}_{GFT^C_{gcmt}} \Bigg] \leq 0 \; \forall \; \tilde{\tau}_{gct} \quad (15)$$

where the o superscript denotes variables that are observed, $\{\mathcal{B}^o_{mt} \cup gc\}$ denotes the set of observed bundles with the modification that g carries c in all (relevant) markets,[32] $\hat{\cdot}$ are estimated values from the first-stage estimation, $\tilde{\boldsymbol{\tau}} \equiv \{\tilde{\tau}_{gct}, \hat{\boldsymbol{\tau}}_{-gct}\}$, $\hat{\boldsymbol{\tau}}_{-gct}$ represents all affiliate fees except those between g and c, and GFT^M_{gcmt} and GFT^C_{gcmt} represent g and c's respective gains-from-trade from agreement in market m.[33,34] If (15) holds for all values of $\tilde{\tau}_{gct}$, the Nash bargaining solution between g and c given by (8) is not defined.

Since we are evaluating a deviation in a model in which bundle composition, bundle prices, and affiliate fees are simultaneously determined, when computing "counterfactual" profits from agreement between channel c and distributor g (the terms with underbraces in (15)), we will hold fixed bundle prices and carriage decisions for all other channels and all other distributors when evaluating counterfactual profits upon carriage of c by g.[35] In that case, condition (15) holds at all $\tilde{\tau}_{gct}$ if and only if the joint profits of the two parties is larger with non-supply. We thus can test whether (15) holds for $\tilde{\tau}_{gct} = 0$ to determine whether or not a deviation for c to supply g is profitable for both parties. Under the Nash-in-Nash (passive beliefs) bargaining assumption, the change in joint profit with satellite distributor g is calculated assuming that satellite distributor $g \neq g'$ does not have access to the RSN.

Multilateral Deviations. In the event of being offered a deviating deal for RSN c, satellite distributor g might instead think that the other satellite distributor g' had also been offered a deal. We instead make use of a necessary condition for non-supply of both satellite distributors to be an equilibrium regardless of the satellite distributors' beliefs. Specifically, we will determine whether, at the observed set of bundles, affiliate fees, and bundle prices, there are no gains from trade between c and *both* satellite providers g and g' (thereby ruling out the presence of this profitable

---

[31] To be specific, we define all markets in a DMA to be relevant for an RSN if, across all systems within that DMA, the average fraction of teams on that RSN that are not blacked out is greater than or equal to .30.

[32] Recall that a satellite distributor offers the same bundle in all markets, and that we assume that it carries any channel with which it has negotiated an agreement.

[33] To be precise, affiliate fees are not directly estimated; instead, we compute their implied values at the estimated parameters $\hat{\boldsymbol{\theta}}$: i.e., $\hat{\boldsymbol{\tau}} \equiv \boldsymbol{\tau}(\hat{\boldsymbol{\theta}})$, where $\boldsymbol{\tau}(\cdot)$ is the solution to the Nash bargaining first-order condition given by (10).

[34] We assume that satellite providers choose to carry an RSN in all relevant markets if supplied with the channel.

[35] The condition that there does not exist a deviation to carriage is not the same as testing whether carriage of c by g would comprise an equilibrium outcome, as this test would require (among other things) computing equilibrium prices and affiliate fees conditional on carriage of c by g being known and anticipated by all firms in the market.

deviation):

$$\sum_{m\in\mathcal{M}_c}\Bigg[[\Delta_{gc,g'c}\Pi^M_{gmt}(\{\mathcal{B}^o_{mt}\cup\{gc,g'c\}\},\boldsymbol{p}^o_{mt},\tilde{\boldsymbol{\tau}};\hat{\mu})]+[\Delta_{gc,g'c}\Pi^M_{g'mt}(\{\mathcal{B}^o_{mt}\cup\{gc,g'c\}\},\boldsymbol{p}^o_{mt},\tilde{\boldsymbol{\tau}};\hat{\mu})]\ldots \\ +[\Delta_{gc,g'c}\Pi^C_{cmt}(\{\mathcal{B}^o_{mt}\cup\{gc,g'c\}\},\boldsymbol{p}^o_{mt},\tilde{\boldsymbol{\tau}};\hat{\mu},\lambda_R)]\Bigg]\leq 0\ , \tag{16}$$

where the three terms on the left-hand side of the inequality represent g, g', and c's gains from trade from both g and g' being supplied with channel c and carrying the channel in all of g's relevant markets, and $\tilde{\boldsymbol{\tau}}$ is equal to $\hat{\boldsymbol{\tau}}$ except that $\tilde{\tau}_{sct}=\tilde{\tau}_{s'ct}=0$. We refer to the sum of these terms as the *three-party-surplus* from carriage of c by satellite providers.[36] As in the case of bilateral deviations before, we test whether or not (16) holds when the negotiated affiliate fees between g and c and between g' and c equal 0.

We estimate a lower bound of λ_R, denoted $\hat{\underline{\lambda}}_R$, by finding the lowest value that ensures that (16) holds for the two cable-integrated RSNs that do not contract with satellite providers in the loophole markets.[37]

Incentives for Exclusion. It is instructive at this point to discuss the competing forces that would induce a cable provider to withold its integrated RSN from a satellite provider. This is equivalent to understanding when the gains created when satellite providers are supplied with the RSN are offset by the losses incurred by the cable provider.

The primary gains created when a satellite provider g is supplied with the RSN are through potential market expansion effects from carriage: i.e., if consumers who previously did not subscribe to an MVPD now would if satellite were to carry the RSN. Each household that substitutes from the outside good to g would generate additional bilateral surplus equal to the level of g's margins plus any additional advertising revenues generated by those households watching the RSN.

The primary losses generated by supplying g with the RSN would be incurred by the RSN's integrated cable owner if households substituted away from the integrated cable provider to g. Although these consumers would generate surplus for g, insofar as cable margins are higher than those of satellite providers (by 10+ percentage points in our data), any household that switched from cable to satellite as a result of supplying satellite with the RSN would reduce industry profits by this difference in margins.

Consequently, factors that would incentivize exclusion by cable of satellite (for $\lambda_r > 0$) would include: a smaller share of consumers that are not subscribers to any MVPD and lower advertising rates (thereby reducing the potential gains generated by market expansion); and a larger cable "footprint" (market share) in the RSN's relevant market area, a larger diversion ratio between satellite and cable distribution, and a larger differential between cable and satellite margins (all

---

[36] Specificatlly, it can be shown that if the three-party-surplus is positive, then RSN c has a deviating pair of offers $\{\tilde{\tau}_g,\tilde{\tau}_{g'}\}$ it can make that the satellite distributors will both accept regardless of their beliefs and increases c's profits.

[37] For now, we will assume away the specification error introduced in (12): i.e., $\nu^2_{ct}=0$.

Table 2: Estimates of Key Parameters

	Parameter Estimate	SE
ν^S	0.60	
ν^{NS}	0.94	
γ_d (Distance Decay)	-3.5263	
γ_b (Blackout Decay)	-1.7108	
α_0	-0.2997	
β_v	0.0404	
$\rho^{sat}_{DirecTV}$	10.295	
ρ^{sat}_{Dish}	13.197	
μ	0.9286	
$\mu \times \underline{\lambda}_r$	0.997	

Notes: Key parameters from the first and second stage estimation of the full model. (Standard Errors TBA).

of which would exacerbate the losses from business stealing by satellite from cable). However, for lower values of λ_r (closer to 0), any losses that would be incurred by the RSN's integrated owner would be internalized less by the RSN when bargaining with g, reducing the likelihood of exclusion occurring.

5 Results

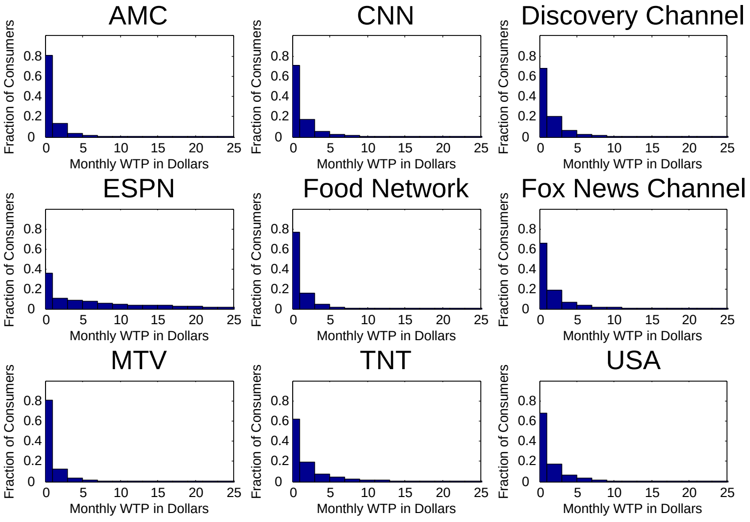
Preliminary estimates of the key parameters of our model are reported in Table 2. We discuss our estimates primarily through how they influence predicted moments relating to consumer viewership and subscription patterns, firm pricing and carraige decisions, and negotiated agreements.

5.1 Channel Valuations

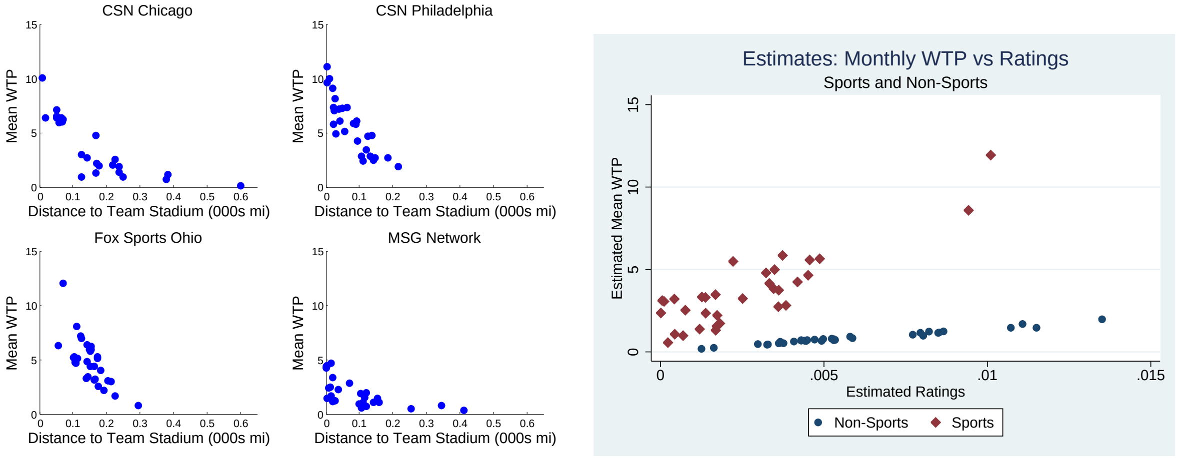
Our model predicts the willingness-to-pay (WTP) for each channel by household by computing the contribution of a given channel to bundle utility (v^*_{ijt} in (2)), and multiplying it by our estimates of β^v/α_i to convert it into dollars.

The distribution of household WTP for nine national channels in 2007 is provided in Figure 5a. In Appendix B, Table 10 reports WTP estimates for all national channels and Table 11 reports WTP estimates for the RSNs. Although most national channels have average WTP values below \$1/month (and other than sports channels ESPN and ESPN2, none exceed \$2), the pattern is very different for RSNs: only 2 out of 30 are predicted to have average WTP values less than \$1/month, and 80% are greater than \$2/month.

Our estimate of the RSN distance-decay is negative, and implies that consumers derive less utility from watching a RSN the further they are from the teams carried on the RSN: increasing the average distance of a household from an RSN's teams' stadiums from 0 to 100 miles reduces that household's value of the channel by 30% ($1 - \exp(-3.53 \times 0.1)$). Figure 5b illustrates this pattern, and plots the predicted mean WTP of households for four different RSNs as the distance from a household to an RSN's team stadium increases. Our blackout parameter is also negative,



(a) Histograms of monthly WTP for selected national channels.



(b) Mean WTP by market-RSN vs distance from market to RSN's teams' stadiums.

(c) Estimated monthly WTP vs estimated ratings for sports and non-sports channels

Figure 5: Predicted WTP for Channels (2007 values).

implying that being unable to watch 50% of the teams that the RSN normally carries due to blackout restrictions reduces the valuation of the channel by 58%.

Finally, we estimate different values of ν^S and ν^{NS}, where the higher value of ν^S implies that consumers' marginal utility from watching sports channels falls faster than for non-sports channels; in turn, this implies that consumers derive higher utility from sports channels than non-sports channels for the same amount of time spent watching each (for a given level of γ). Our model thus predicts that sports channels receive higher negotiated affiliate fees for the same viewership ratings, as depicted in Figure 5c.

Table 3: Elasticities and Margins

Elasticity of row with respect to price of column:	Cable	DirecTV	Dish
Cable	-2.093	0.280	0.192
DirecTV	1.895	-3.130	0.278
Dish	2.111	0.482	-3.333
Mean Cable Margin	0.678		
Mean DirecTV Margin	0.373		
Mean Dish Margin	0.401		
OLS Logit Price Coefficient	-0.0046**	(t: -2.40)	
IV Logit Price Coefficient	-0.0987***	(t: -6.17)	

Notes: This table reports mean price elasticities and margins by cable and the two satellite distributors, as well as the effect of the satellite tax instrument on the price coefficient in a logit demand system.

5.2 Pricing and Subscription Choices

In Table 3, we report average predicted own and cross price elasticities and implied margins for cable and satellite MVPDs predicted by our model. Demand for the average cable system is more inelastic (-2.1) than for satellite (-3.1 and -3.3), which is consistent with its larger market shares and higher predicted margins. Estimated values of $\rho^{sat}_{DirecTV}$ and ρ^{sat}_{Dish}, which govern the distribution of random preferences for satellite bundles, assist the model in predicting price-cost margins that are close to the observed Comcast, DirecTV, and Dish margins.

In addition, the bottom panel of Table 3 reports the effect of instrumenting for bundle prices using the satellite tax instrument that was discussed in the previous section. In a logit demand system, instrumenting for price yields a 20 times larger estimated price coefficient, consistent with the presence of a positive correlation between price changes and unobservable bundle characteristics.

5.3 Internalization and RRC Parameters.

We now turn to the estimates and magnitudes of μ and λ_R.

Our estimated value of μ indicates that firms do internalize the profits of other integrated units when making decisions: i.e., when pricing and determining carriage on its bundles, an MVPD internalizes potential effects on affiliate fees and advertising revenues accruing to integrated channels; and when bargaining internally, an integrated MVPD and channel face reduced double marginalization incentives. Insofar our estimated value of $\mu < 1$, however, such internalization may be imperfect.

Our estimated lower bound for $\mu \times \underline{\lambda}_R$ is .997, which indicates that integrated channels' supply decisions vis-à-vis non-integrated rival distributors are significantly affected by foreclosure incentives. Figure 6 graphs the total three party surplus between the integrated channel and the two satellite distributors in the two loophole markets we examine (Philadelphia and San Diego). We see that for values of lower than .9, it is not an equilibrium for either channel to exclude both satellite distributors as there would be a profitable deviation (for some negotiated set of affiliate fees) for

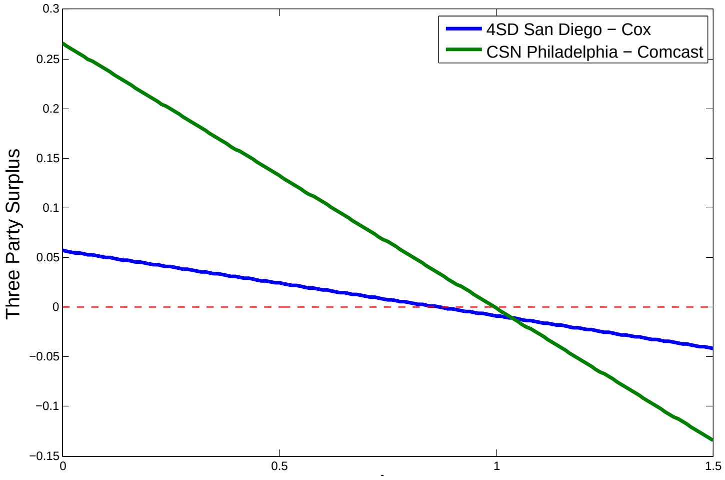


Figure 6: Three-party surplus between the integrated cable MVPD, DirecTV, and Dish as a function of $\mu \times \lambda_r$ in Philadelphia and San Diego.

the channel to be supplied. However, for values between approximately .9 and our estimate, we can rationalize exclusion in San Diego but not Philadelphia. Only for values of $\mu \times \underline{\lambda}_R \geq .997$ does our model rationalize exclusion in both of these loophole markets.

6 The Welfare Effects of Vertical Integration

In this section, we use estimates from our model to perform counterfactual exercises that illustrate how vertical integration affects affiliate fee negotiations, distributors' pricing and carriage decisions, and—ultimately—firm and consumer welfare.

We focus on 30 RSNs that are active in 2007, 16 of which were (at least partially) integrated with a downstream distributor (13 with a cable MVPD, 3 with DirecTV). Of these 16 integrated RSNs, two—CSN Philadelphia and 4SD—were owned by cable distributors in "loophole" markets, and were not provided to satellite. Consequently, there is variation in both integration and ownership as well as whether or not the RSN is subject to program access rules (PARs) during this time period. For every RSN, we aim to simulate market outcomes if the RSN is or is not integrated, and if it is integrated, whether or not PARs are enforced.

For our current analysis, we exclude from the analysis 3 RSNs (CSN NW, CSS, and Cox Sports TV) that did not supply either satellite provider in markets where PARs were in effect, leaving us with 27 RSNs.

6.1 Potential Effects

Before proceeding, it is instructive to highlight the effects of vertical integration that are captured by our model and that we attempt to quantify in our counterfactual exercises.

First, our model emphasizes three primary supply side decisions: negotiations over affiliate fees and supply, bundle pricing, and channel carriage conditional on supply. When an MVPD and a channel are integrated, our estimated value for $\hat{\mu} > 0$ implies that integrated downstream and upstream units (at least partially) internalize joint profits when making all of these decisions; furthermore, our estimated value for $\hat{\underline{\lambda}}_R > 0$ implies that an integrated channel may have incentives to foreclosure a rival downstream distributor.

Assume now that MVPD f integrates with channel c, and that there is a rival MVPD g and another channel d. The following effects of vertical integration are thus admitted in our setting:

1. Bargaining Effects and Foreclosure: When c bargains with a rival MVPD g (since $\hat{\underline{\lambda}}_R > 0$), c internalizes lost revenues to its integrated downstream MVPD f if g is supplied; the gains-from-trade that accrue to c by supplying g are thus reduced, potentially leading to a higher negotiated affiliate fee τ_{gct} or—if gains-from-trade are eliminated altogether—non-supply.[38]
2. Pricing Effects:
   (a) f faces a lower "perceived" marginal cost as it internalizes affiliate fee payments made to c, thereby mitigating double marginalization incentives.
   (b) f internalizes affiliate fees paid by rival MVPD g to integrated channel c, thereby partly alleviating bundle pricing pressure across MVPDs (by increasing f's "effective" marginal cost) as f now partly benefits from customers lost to g (Chen, 2001).
3. Carriage Effects: an MVPD f may be more likely to carry c in markets where the gains-to-carriage are marginal as, again, f internalizes payments made to c and faces a lower perceived marginal cost of carriage.

The welfare effects of some of these incentives may be straightforward to sign ex ante; for others, it is not clear. Downstream foreclosure, for instance, may likely lead to consumer welfare losses: if g loses access to c or pays a higher affiliate fee τ_{gct}, g's subscribers may receive less utility from their bundle of channels (from reduced choice or higher prices); f's price may also increase in response to facing a weaker competitor.[39] However, the two pricing effects have potentially opposite effects: whereas 2a would favor lower bundle prices, 2b may mitigate price competition and push prices higher. Finally, increased carriage of channels may raise consumer welfare.

There are also other potential responses that are not accounted for in our model. Most importantly, we have not modeled investment in channel, programming, and distribution service quality, which may change upon integration (Bolton and Whinston, 1991; Hart, 1995). Consequently, although our counterfactuals are indeed rich, they are still only partial equilibrium results, and thus any interpretation of our findings must be made with this in mind.

---

[38] Although we do not focus on this, our model does admit the potential for "upstream" in addition to "downstream" foreclosure: i.e., when f negotiates with other channels d, f internalizes viewership changes to its integrated channel c when it carries channel d on its own bundles; if c and d are substitutable, f may be willing to pay a lower negotiated affiliate fee τ_{fdt} (and potentially have less of an incentive to carry d), as the gains from trade from f carrying d are partially mitigated by lost viewership and advertising revenues to c.

[39] If g lowers its price as a result of losing c, it is also possible that f may lower its price: e.g., f and g's prices could be strategic complements, and f could potentially reduce its marginal cost by negotiating lower affiliate fees with d (e.g., if g's loss of c reduces d's outside option from disagreement with f).

6.2 Implementation

For each RSN that is active in 2007, we compute market outcomes in that year under the following three scenarios:

1. Integration and no-PARs: In this environment, for any non-integrated RSN, we assume that the largest cable MVPD in that RSN's relevant DMAs is the new full owner of the channel; for any integrated RSN, we do not change its ownership structure. We assume that in this environment, $\mu = \hat{\mu}$ and $\lambda_R = \hat{\underline{\lambda}}_R$ so that all RSNs are allowed to potentially exclude and not supply rival MVPDs.

   For each cable-owned RSN, to determine whether the channel is supplied to satellite distributors, we will determine whether or not at the observed set of bundles, affiliate fees, and bundle prices, the *three-party-surplus* given by (16) is positive at the new values of μ and λ_R.[40] For the three RSNs owned by DirecTV, we use the bilateral surplus given by (15) to determine whether or not each cable MVPD and Dish Network is provided with the channel.[41]

2. Integration and PARs: We follow the same setup as in the Integration and no-PARs case, except that we assume that: $\lambda_R = 0$, the two integrated loophole RSNs—CSN Philadelphia and 4SD—are supplied to both satellite distributors, and the supply decisions of all other RSNs are unchanged.[42]

3. Non-Integration: We follow the same setup as in the Integration and PARs case, except that we assume that $\mu = 0$ so that all RSNs are effectively non-integrated (i.e., no MVPD or channel internalizes the profits of any other unit). This is equivalent to assuming that ownership shares $O_{fct} = 0$ for all MVPDs and RSNs.

In all scenarios, we recompute equilibrium negotiated affiliate fees $\boldsymbol{\tau}_{ct}$ for the RSN in question and bundle prices $\{p_{fmt}\}$ for all cable distributors, and we assume that national satellite prices are unchanged (see Appendix A.3 for further details). To account for potential carriage changes as integration status varies, we make the following adjustments:

1. For RSN c that was previously integrated with MVPD f but is now disintegrated, we take the minimum of: (i) the maximum distance between c's teams' stadiums and a non-integrated cable MVPD system that carried the RSN, and (ii) the minimum distance between c's teams' stadiums and a non-integrated cable MVPD system that did not carry the RSN. We then assume that any system that is beyond this minimum distance for the formerly integrated

---

[40] For the two loophole RSNs (CSN Philadelphia and 4SD), we do not need to redetermine supply as we utilize a value of λ_R that rationalizes non-supply of satellite.

[41] We also test whether or not these supply conditions hold at the updated equilibrium affiliate fees and bundle prices.

[42] Aside from the loophole RSNs, all RSNs are provided to all distributors in 2007 except in four cases: integrated RSNs CSN Northwest, Comcast/Charter Sports Southeast, and Cox Sports TV supply neither satellite provider; YES Network is independent and does not supply Dish. We exclude from our analysis the first three channels, and hold fixed YES's supply decisions when it is integrated.

MVPD f no longer carries c once c is disintegrated. As the distance from an RSN's teams' stadiums and a cable system correlates with the gains-from-carriage of that channel, we thus use the observation that non-integrated MVPDs did not carry c past a certain distance to approximate optimal behavior for f once it is disintegrated. This approach attempts to provide an upper bound on the degree to which carriage can increase between non-integration and integration: e.g., if we believe that the integrated MVPD f has greater gains-from-carriage of c for one of its systems with similar observables (i.e., distance) as a system belonging to a non-integrated MVPD (which is consistent with f choosing to integrate with c in the first place), then we may overstate potential carriage changes.

More sophisticated adjustments to carriage (which may include accounting for profit disturbances in carriage and allowing multiple MVPDs to adjust their carriage decisions) are the focus of ongoing work.

2. For an RSN that was previously non-integrated but becomes integrated, we assume that the RSN is carried by all of the new owner's systems in the RSN's relevant DMA. This also provides an upper bound on the extent to which vertical integration can increase the carriage of a channel by the integrated distributor.

6.3 Results

Table 4 reports market shares, prices, firm surpluses, and consumer welfare across the three different integration environments for six selected RSNs. Panel I contains the two cable-integrated RSNs that operate in terrestrial loophole markets; Panel II contains two selected cable-integrated RSNs located in non-loophole markets; and panel III contains two selected non-integrated RSNs that are assigned a cable owner in the integrated specifications. For each panel, one of the specifications ("(i) VI, no PARs," "(ii) VI, PARs," and "(iii) No VI") corresponds to the actual observed setting: e.g., specification (i) is the setting for the terrestrial loophole RSNs described in Panel I. All reported figures except for market shares are in \$ per household per month, and all percentage changes are relative to the non-integration specification (iii). Below each RSN name is the MVPD that either owns the channel or is assigned ownership of the RSN, the number of households and the MVPD owner's footprint (which is the percentage of households that the MVPD "passes" or plausibly could serve) in the RSN's relevant DMAs, and the predicted percentage change in the number of the integrated MVPD's households that carry the RSN once the channel is integrated. A missing "Aff Fees to Sat" value indicates that the RSN is not supplied to the two satellite distributors. Outcomes for all individual RSNs are reported in the Appendix in Tables 12-15.

Table 5 reports the same market outcomes averaged across all RSNs within each of the three panels in Table 4, as well as across all RSNs in our sample, weighted by the number of households in each RSN's relevant DMAs. "Aff Fees to Rival" represents the affiliate fees charged to the integrated MVPD's rival distributors (both satellite MVPDs if the channel is cable-integrated, each cable MVPD and Dish if the channel is integrated with DirecTV) *conditional on the channel*

Table 4: Simulated Market Outcomes for Selected RSNs

		(i) VI, no PARs		(ii) VI, PARs		(iii) No VI
		Level	% Change	Level	% Change	Level
I. VI, LOOPHOLE						
CSN PHIL	Avg Cable Mkt Share	0.65	4.78%	0.63	2.34%	0.62
Comcast	Avg Sat Mkt Share	0.16	-16.35%	0.19	-1.51%	0.19
Pop 4.25M	Avg Cable Prices	55.36	0.10%	54.57	-1.32%	55.31
Footprint 90%	Aff Fees to Sat	-	-	2.05	0.02%	2.05
	Cable Surplus	28.11	3.87%	27.05	-0.06%	27.06
Carriage +13%	Satellite Surplus	3.82	-8.42%	4.10	-1.54%	4.17
	RSN Surplus	1.91	-19.71%	2.44	2.91%	2.38
	Consumer Welfare	25.65	-4.03%	27.11	1.43%	26.73
	Total Welfare	59.48	-1.41%	60.70	0.61%	60.33
4SD	Avg Cable Mkt Share	0.73	0.81%	0.73	0.22%	0.73
Cox	Avg Sat Mkt Share	0.16	-3.41%	0.16	-0.34%	0.16
Pop 2.81M	Avg Cable Prices	54.62	-0.09%	54.58	-0.16%	54.67
Footprint 100%	Aff Fees to Sat	-	-	0.22	-0.28%	0.22
	Cable Surplus	44.93	0.72%	44.61	0.02%	44.61
Carriage +0%	Satellite Surplus	3.26	-2.66%	3.33	-0.33%	3.34
	RSN Surplus	0.52	-28.17%	0.71	-0.74%	0.72
	Consumer Welfare	26.20	-1.16%	26.56	0.20%	26.51
	Total Welfare	74.90	-0.37%	75.22	0.06%	75.18
II. VI, NON-LOOPHOLE						
CSN BAY AREA	Avg Cable Mkt Share	0.65	1.62%	0.65	1.69%	0.64
Comcast	Avg Sat Mkt Share	0.21	-1.13%	0.21	-1.19%	0.21
Pop 6.03M	Avg Cable Prices	56.00	-0.96%	55.97	-1.01%	56.55
Footprint 54%	Aff Fees to Sat	2.64	60.82%	1.62	-1.65%	1.64
	Cable Surplus	34.55	0.13%	34.56	0.17%	34.51
Carriage +0%	Satellite Surplus	4.21	-3.89%	4.34	-1.10%	4.38
	RSN Surplus	1.56	6.48%	1.42	-3.00%	1.46
	Consumer Welfare	29.14	1.02%	29.16	1.08%	28.85
	Total Welfare	69.46	0.38%	69.48	0.40%	69.20
CSN MID-ATL	Avg Cable Mkt Share	0.66	2.86%	0.66	2.12%	0.64
Comcast	Avg Sat Mkt Share	0.16	-11.84%	0.17	-2.19%	0.18
Pop 6.55M	Avg Cable Prices	58.44	0.01	57.40	-1.09%	58.03
Footprint 70%	Aff Fees to Sat	-	-	1.20	-8.93%	1.32
	Cable Surplus	23.40	2.62%	22.88	0.33%	22.81
Carriage +9%	Satellite Surplus	3.68	-4.44%	3.80	-1.30%	3.85
	RSN Surplus	1.22	-10.94%	1.31	-4.50%	1.37
	Consumer Welfare	25.59	-2.76%	26.67	1.35%	26.31
	Total Welfare	53.89	-0.83%	54.66	0.59%	54.34
III. NON-INTEGRATED						
FS DETROIT	Avg Cable Mkt Share	0.58	3.25%	0.58	3.46%	0.56
*Comcast	Avg Sat Mkt Share	0.16	-1.87%	0.16	-2.04%	0.17
Pop 4.84M	Avg Cable Prices	50.02	-1.81%	49.95	-1.95%	50.94
Footprint 82%	Aff Fees to Sat	2.83	75.78%	1.55	-3.25%	1.61
	Cable Surplus	19.99	0.49%	20.02	0.61%	19.89
Carriage +15%	Satellite Surplus	3.37	-6.49%	3.54	-1.80%	3.61
	RSN Surplus	1.97	8.52%	1.77	-2.32%	1.81
	Consumer Welfare	23.71	2.08%	23.74	2.23%	23.22
	Total Welfare	49.04	1.03%	49.07	1.10%	48.54
NESN	Avg Cable Mkt Share	0.68	9.09%	0.68	8.83%	0.63
*Comcast	Avg Sat Mkt Share	0.10	-20.16%	0.11	-7.83%	0.12
Pop 5.20M	Avg Cable Prices	58.37	-1.63%	55.53	-6.42%	59.34
Footprint 52%	Aff Fees to Sat	-	-	1.92	-11.34%	2.16
	Cable Surplus	29.97	5.27%	28.85	1.35%	28.46
Carriage +0%	Satellite Surplus	2.16	-11.20%	2.27	-6.35%	2.43
	RSN Surplus	3.51	-13.39%	3.59	-11.55%	4.05
	Consumer Welfare	24.95	0.03%	26.93	7.98%	24.94
	Total Welfare	60.58	1.16%	61.64	2.93%	59.89

Notes: This table presents observed and simulated market outcomes for individual RSNs. Specification (i) corresponds to assuming that $\lambda_R = \underline{\hat{\lambda}}_R$ and $\mu = \hat{\mu}$, and allowing the owner of the RSN not provide to rival MVPDs; specification (ii) corresponds to setting $\lambda_R = 0$ and prohibiting the RSN owner from excluding rivals; specification (iii) sets $\mu = 0$ and disintegrates the RSNs. Panel I shows the two cable-integrated RSNs located in terrestrial loophole markets; Panel II shows two selected cable-integrated RSNs located in non-loophole markets; Panel III shows two selected non-integrated RSNs that are assigned a cable owner in specifications (i) and (ii). All reported figures except for market shares are in \$/household/month, and all % changes are relative to specification (iii). Beneath the channel name is the name of the MVPD that owns (or is assigned ownership of) the channel, the number of television households and the MVPD owner's footprint (% of households passed) in the RSN's relevant DMAs, and the % change in the integrated MVPD's households that obtain access to the channel upon moving from (iii) to (ii) in the RSN's relevant DMAs.

Table 5: Average of Simulated Market Outcomes

	(i) VI, no PARs		(ii) VI, PARs		(iii) No VI
	Level	% Change	Level	% Change	Level
I. VI, LOOPHOLE					
Avg Cable Mkt Share	0.68	3.04%	0.67	1.41%	0.66
Avg Sat Mkt Share	0.16	-11.71%	0.18	-1.09%	0.18
Avg Cable Prices	55.06	0.02%	54.58	-0.86%	55.05
Aff Fees to Sat	-	-	1.32	0.00%	1.32
Cable Surplus	34.81	2.22%	34.04	-0.02%	34.05
Satellite Surplus	3.59	-6.42%	3.80	-1.12%	3.84
RSN Surplus	1.35	-21.12%	1.75	2.30%	1.72
Consumer Welfare	25.87	-2.89%	26.89	0.94%	26.64
Total Welfare	65.63	-0.94%	66.49	0.36%	66.25
# Foreclosed to Sat	2/2		Carriage Increase:		7.66%
II. VI, NON-LOOPHOLE					
Avg Cable Mkt Share	0.65	1.33%	0.65	1.07%	0.65
Avg Sat Mkt Share	0.17	-3.80%	0.18	-0.99%	0.18
Avg Cable Prices	58.57	-0.24%	58.34	-0.64%	58.71
Aff Fees to Rivals	1.57	19.45%	1.08	-0.67%	1.09
Cable Surplus	27.30	0.56%	27.16	0.07%	27.15
Satellite Surplus	3.76	-2.34%	3.81	-0.91%	3.84
RSN Surplus	1.14	-2.00%	1.14	-0.45%	1.15
Consumer Welfare	27.15	-0.25%	27.41	0.74%	27.21
Total Welfare	59.34	-0.01%	59.52	0.30%	59.34
# Foreclosed to Sat	4/11		Carriage Increase:		3.33%
III. NON-INTEGRATED					
Avg Cable Mkt Share	0.61	1.79%	0.61	1.83%	0.60
Avg Sat Mkt Share	0.20	-2.04%	0.20	-1.41%	0.20
Avg Cable Prices	56.67	-0.71%	56.49	-1.02%	57.07
Aff Fees to Sat	1.39	38.55%	1.12	-0.50%	1.13
Cable Surplus	20.99	0.56%	20.92	0.28%	20.87
Satellite Surplus	4.44	-2.83%	4.50	-1.30%	4.55
RSN Surplus	1.17	9.26%	1.12	1.79%	1.14
Consumer Welfare	26.20	0.75%	26.34	1.28%	26.01
Total Welfare	52.80	0.44%	52.88	0.58%	52.57
# Foreclosed to Sat	2/14		Carriage Increase:		10.45%
ALL RSNS					
Avg Cable Mkt Share	0.63	1.73%	0.63	1.67%	0.62
Avg Sat Mkt Share	0.19	-3.08%	0.19	-1.36%	0.19
Avg Cable Prices	57.36	-0.54%	57.14	-0.97%	57.64
Aff Fees to Rivals	1.45	32.70%	1.11	-0.92%	1.12
Cable Surplus	23.95	0.65%	23.84	0.30%	23.80
Satellite Surplus	4.14	-3.38%	4.20	-1.24%	4.25
RSN Surplus	1.17	0.19%	1.15	-1.51%	1.16
Consumer Welfare	26.56	0.27%	26.78	1.19%	26.50
Total Welfare	55.82	0.20%	55.97	0.56%	55.71
# Foreclosed to Sat	8/27		Carriage Increase:		7.56%

Notes: This table presents the average across the simulated market outcomes for the RSNs located in each panel, weighted by the number of households in each RSN's relevant DMAs. All reported figures except for market shares are in $/household/month, and all % changes are relative to specification (iii). "Avg Fees to Rival" are computed conditional on supply (and represent average affilliate fees to the satellite MVPDs for cable-integrated RSNs, and to cable MVPDs for satellite-integrated RSNs), "# Foreclosed to Sat" reports the number of RSNs in each panel that are not provided to satellite MVPDs (in the case of a cable-integrated RSN; satellite-owned RSNs are never predicted to not supply cable MVPDs), and "Carriage Increase" is the % increase in the number of the integrated MVPD's households that can access the channel upon integration.

being supplied. "# Foreclosed to Sat" represents the number of RSNs in each panel that are not provided to satellite distributors (we predict that DirecTV-integrated RSNs are never withheld from cable distributors).

Efficiency Effects: Double Marginalization and Carriage. We first focus on the potential efficiency gains from vertical integration, which we capture via the difference between specification (ii), representing integration with PARs in effect ($\lambda_R = 0$), and specification (iii), non-integration. Across all RSNs (bottom panel of Table 5), integration and the elimination of double-marginalization for a single RSN yields on average approximately a \$0.50 (1%) decrease in cable prices. Though integration of most RSNs yields less than a \$1 decrease in cable prices, there are notable exceptions: e.g., integrating NESN with Comcast, reported in Table 4, results in average cable prices falling by nearly \$4 (6%) due to NESN's high estimated affiliate fees to Comcast (over \$5/month). Carriage of the integrated RSN, on average, increases by approximately 8% for the integrated MVPD's households, though again there is significant variation across channels.

Although it appears that cable providers are occasionally made worse off when integrated, this is only because the reported cable surplus counts downstream (distributor) profits alone; when a cable MVPD is integrated (and since $\mu > 0$), its pricing decisions will be optimal with respect to joint RSN and distributor profits. We thus always find that joint RSN and integrated cable surplus increases when moving from non-integration to integration with PARs. Satellite surplus falls by approximately 1% when cable integrates with the RSN.[43]

The net effect across all RSNs for total welfare from integrating each RSN but enforcing PARs is approximately \$0.26/household/month (.56%), with all of this coming from an increase in consumer surplus. As noted before, we believe that our approach for predicting counterfactual carriage may overstate the extent to which carriage increases upon integration, and thus caveat our findings accordingly. Even without carriage changes, for some channels such as NESN, surplus gains can be large: we predict a total and consumer welfare gain of nearly \$2/household/month (3% and 8%) when this channel is integrated (its carriage does not increase upon integration). Nevertheless, no other channel is predicted to generate total or consumer welfare gains from integration of over \$1.00.

Foreclosure Effects: Raising-Rivals'-Costs and Exclusion. Comparing specifications (i) and (ii) across all tables provides the impact of removing PARs (and allowing $\lambda_R > 0$) so that integrated RSNs internalize the profits of their downstream units when bargaining with rival MVPDs. For all RSNs, we find that allowing foreclosure strictly reduces consumer and total welfare from the case where integration was allowed but PARs were enforced ($\lambda_R = 0$).

This reduction is caused by two main effects. The first occurs when an RSN is excluded from rival MVPDs. We predict that none of the three DirecTV-owned RSNs would choose to exclude cable providers; however, we predict that 8 out of the 24 cable-integrated RSNs would wish to exclude satellite. We find that all excluded RSNs have cable owners with at least a 40% footprint, consistent with the discussion in Section 4.2.1 that noted that a larger cable footprint would increase the potential losses incurred by a cable provider upon supply of satellite. For

---

[43] In Table 12, we report market outcomes for the three satellite integrated RSNs. Although we assume that satellite MVPDs set national prices and do not adjust them in our counterfactuals, since specification (ii) also allows for $\mu > 0$ for other integrated-RSNs in the same relevant DMA, cable prices and hence market outcomes can change.

the selected set of RSNs described in Table 4, there are four channels—the two loophole RSNs, previously integrated CSN Mid-Atlantic, and previously independent NESN—that are predicted to exclude satellite distributors if PARs were lifted. In these cases, satellite markets shares and surplus fall.

When rival MVPDs are still supplied by integrated MVPDs but $\lambda_r > 0$, we find that raising-rivals'-costs effects can be large. Table 5 column (i) reports affiliate fees charged to rivals *conditional on supply*; we find that affiliate fees, upon agreement, increase by over 30% on average across all RSNs from the baseline of non-integration; focusing only on previously non-integrated RSNs, this figure is almost 40%. In some cases, including for CSN Bay Area and FS Detroit in Table 4, this increase is over \$1/month/subscriber. Although the impact of an increase in affiliate fees on a rival is not as significant as exclusion, the increase still results in a meaningful reduction in surplus.

However, even though we have assumed that satellite distributors do not adjust their prices in our counterfactuals, a cable-integrated RSN can still negatively harm consumer welfare by charging satellite distributors higher affiliate fees if this in turn induces the integrated cable owner to increase its own downstream prices. Intuitively, if a cable-integrated RSN increases its affiliate fees with satellite distributors, then the RSN's downstream cable MVPD is facing a higher effective marginal cost when pricing its cable bundle since it now internalizes lost affiliate fee revenues to the RSN from satellite. This effect, discussed in Chen (2001), can be seen clearly when comparing specifications (i) and (ii) when a channel is not excluded from its rivals: e.g., Comcast, the assigned owner of FS Detroit in Table 4, increases its own price of a bundle by \$0.07/month as a result of negotiating a 75% higher affiliate fee for FS Detroit from satellite distributors.[44]

Net Effects. By comparing specification (i) to (iii), we can estimate the net impact of integration of RSNs without restrictions on exclusion from a non-integrated baseline case. On average across all RSNs, the efficiency effects slightly dominate the foreclosure effects when examining total and consumer welfare. These averages, however, mask considerable heterogeneity. Foreclosure effects dominate if one focuses only on the already integrated RSNs—for which integration without PARs is predicted to reduce consumer and total welfare—versus RSNs that are not integrated in the data. Foreclosure effects also dominate when examining only the cases in which the RSN is predicted to exclude satellite distributors upon integration with a cable MVPD: consumer and total welfare fall in 6 of the 8 RSN cases where exclusion was predicted, with the two exceptions—Altitude Sports and NESN—being the only two that previously were disintegrated. For these two RSNs in which the consumer welfare losses from exclusion are not enough to offset the welfare gains from integration, the efficiency gains come either from a large increase in carriage (nearly 50% for Altitude) or a large reduction in cable prices due to mitigated double marginalization incentives (nearly \$4 for NESN).

---

[44] The welfare effects of these large increases in affiliate fees charged to satellite may be muted by our consideration of one vertical merger at a time, which motivated our holding satellite distributors' (national) prices fixed. Were integration to increase nationally and lead to higher affiliate fees charged to satellite distributors in many markets, we may expect satellite prices to increase in response, lowering consumer welfare further.

Our analysis suggests that PARs are binding for some RSNs that are already integrated, and widespread enforcement—conditional on vertical integration status—may improve consumer and total welfare by prohibiting otherwise harmful foreclosure activities.

7 Concluding Remarks

This paper examines vertical integration of high value sports content in the U.S. cable and satellite television industry. Our framework accounts for consumer viewership and subscription decisions, distributor pricing and carriage decisions, and channel-distributor bargaining over affiliate fees. The framework allows for vertical integration to reduce double marginalization and increase carriage; to foreclose rivals from carrying integrated content or to raise their costs of carriage; and for the possibility that divisions within an integrated firm do not perfectly internalize each other's incentives.

We use the estimated model to examine the welfare effects of vertical integration regulatory policy towards the supply of integrated sports content. We find that relaxing program access rules would result in foreclosure of cable integrated RSNs to satellite distributors in a handful of large markets including New England, New York, and Washington DC, and enforcing these regulations in "loophole" markets would prevent existing exclusion. We predict that foreclosure decreases consumer surplus, resulting in losses as large as 3 to 4 percentage points per capita when a single channel is not supplied to satellite, as consumers with strong tastes for both satellite television and regional sports are unable to consume regional sports on their chosen distributor. Even if exclusion does not occur, we find evidence that foreclosure incentives can lead to significant increases in the affiliate fees charged to rival distributors. Finally, we find that the net effect of vertical integration—accounting for both foreclosure and efficiency effects—is heterogeneous, and positive for both consumer and total welfare when averaged across our entire sample of channels, but negative when averaged across only those channels that were integrated during our period of analysis.

As we have noted previously, this analysis is partial and can be extended in a number of directions. Incorporating additional responses to vertical integration—including investment and firm entry—and examining how predictions would be impacted by improved information sharing or alignment of incentives within the firm are important extensions. Furthermore, documenting and measuring the strength of these vertical integration effects in other industries remains a promising area for future research.

References

Ackerberg, Daniel A. (2009). 'A New Use of Importance Sampling to Reduce Computational Burden in Simulation Estimation', *Quantitative Marketing and Economics* 7(4), 343–376.

Asker, John. (2004), Measuring Advantages from Exclusive Dealing. Unpublished.

Berry, Steven, Levinsohn, James and Pakes, Ariel. (1995). 'Automobile Prices in Market Equilibrium', *Econometrica* 63(4), 841–890.

Bolton, Patrick and Whinston, Michael D. (1991). 'The "Foreclosure" Effects of Vertical Mergers', *Journal of Institutional and Theoretical Economics* 147, 207–226.

Chen, D. and Waterman, D. (2007). 'Vertical Ownership, Program Network Carriage, and Tier Positioning in Cable Television: An Empirical Study', *Review of Industrial Organization* 30(3), 227–251.

Chen, Yongmin. (2001). 'On vertical mergers and their competitive effects', *RAND Journal of Economics* 32(4), 667–685.

Chipty, Tasneem. (2001). 'Vertical Integration, Market Foreclosure, and Consumer Welfare in the Cable Television Industry', *American Economic Review* 91(3), 428–453.

Collard-Wexler, Allan, Gowrisankaran, Gautam and Lee, Robin S. (2014), Bargaining in Bilateral Oligopoly: An Alternating Offers Representation of the "Nash-in-Nash" Solution. NBER Working Paper 20641.

Conlon, Christopher T. and Mortimer, Julie H. (2013), All Units Discount: Experimental Evidence from the Vending Industry. Unpublished.

Crawford, Gregory S., Lee, Robin S., Vieira, Breno, Whinston, Michael and Yurukoglu, Ali. (2014), Channel 5 or 500? Vertical Integration, Favoritism, and Discrimination in Multichannel Television. Unpublished.

Crawford, Gregory S. and Yurukoglu, Ali. (2012). 'The Welfare Effects of Bundling in Multichannel Television Markets', *American Economic Review* 102(2), 643–685.

Draganska, Michaela, Klapper, Daniel and Villas-Boas, Sofia B. (2010). 'A Larger Slice or a Larger Pie? An Empirical Investigation of Bargaining Power in the Distribution Channel', *Marketing Science* 29(1), 57–74.

FCC. (2012), Report and Order in MB Docket Nos. 12-68, 07-18, 05-192; Further Notice of Proposed Rulemaking in MB Docket No. 12-68; Order on Reconsideration in MB Docket No. 07-29, Technical report. FCC 12-123, released October 5, 2012.

Gowrisankaran, Gautam, Nevo, Aviv and Town, Robert. (forthcoming). 'Mergers When Prices Are Negotiated: Evidence from the Hospital Industry', *American Economic Review* .

Grennan, Matthew. (2013). 'Price Discrimination and Bargaining: Empirical Evidence from Medical Devices', *American Economic Review* 103(1), 147–177.

Grossman, S. and Hart, O. (1986). 'The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration', *Journal of Political Economy* 94, 691–796.

Hart, Oliver. (1995), *Firms, Contracts, and Financial Structure*, Oxford University Press, New York.

Hart, Oliver and Tirole, Jean. (1990). 'Vertical Integration and Market Foreclosure', *Brookings Papers on Economic Activity. Microeconomics* 1990, 205–286.

Hastings, Justine S. (2004). 'Vertical Relationships and Competition in Retail Gasoline Markets: Empirical Evidence from Contract Changes in Southern California', *American Economic Review* 94, 317–328.

Hastings, Justine S. and Gilbert, Richard J. (2005). 'Market Power, Vertical Integration and the Wholesale Price of Gasoline', *Journal of Industrial Economics* 53(4), 469–492.

Ho, Kate and Lee, Robin S. (2013), Insurer Competition and Negotiated Hospital Prices. NBER Working Paper 19401.

Horn, Hendrick and Wolinsky, Asher. (1988). 'Bilateral Monopoly and Incentives for Merger', *The RAND Journal of Economics* 19, 408–419.

Hortacsu, Ali and Syverson, Chad. (2007). 'Cementing Relationships: Vertical Integration, Foreclosure, Productivity, and Prices', *Journal of Political Economy* 115(2), 250–301.

Houde, Jean-Francois. (2012). 'Spatial Differentiation and Vertical Mergers in Retail Markets for Gasoline', *American Economic Review* 102(5), 2147–2182.

Krattenmaker, Thomas G. and Salop, Steven C. (1986). 'Anticompetitive Exclusion: Raising Rivals' Costs to Achieve Power over Price', *The Yale Law Journal* 2, 209–293.

Lee, Robin S. (2013). 'Vertical Integration and Exclusivity in Platform and Two-Sided Markets', *American Economic Review* 103(7), 2960–3000.

Mortimer, Julie H. (2008). 'Vertical Contracts in the Video Rental Industry', 75, 165–199.

Nocke, Volker and White, Lucy. (2007). 'Do Vertical Mergers Facilitate Upstream Collusion', *American Economic Review* 97(4), 1321–1339.

Ordover, J., Saloner, G. and Salop, S. (1990). 'Equilibrium Vertical Foreclosure', *American Economic Review* 80, 127–142.

Pakes, Ariel, Porter, Jack, Ho, Kate and Ishii, Joy. (forthcoming). 'Moment Inequalities and Their Application', *Econometrica* .

Perry, Martin K. (1990), Vertical Integration: Determinants and Effects, *in* **Richard Schmalensee and Robert Willig**., eds, 'Handbook of Industrial Organization', Vol. 1, North Holland, Amsterdam.

Riordan, Michael H. (2008), Competitive Effects of Vertical Integration, *in* **Paolo Buccirossi**., ed., 'Handbook of Antitrust Economics', MIT Press, Cambridge, MA.

Rogerson, William P. (2014), Vertical Mergers in the Video Programming and Distribution Industry: The Case of Comcast-NBCU, *in* **Jr. John E. Kwoka and Lawrence J. White**., eds, 'The Antitrust Revolution, 6th Edition', Oxford University Press, New York.

Salop, Steven C. and Scheffman, David T. (1983). 'Raising Rivals' Costs', *American Economic Review* 73(2), 267–271.

Shepard, Andrea. (1993). 'Contractual Form, Retail Price, and Asset Characteristics in Gasoline Retailing', *RAND Journal of Economics* 24(1), 58–77.

Spengler, Joseph J. (1950). 'Vertical integration and antitrust policy', *The Journal of Political Economy* pp. 347–352.

Suzuki, Ayako. (2009). 'Market foreclosure and vertical merger: A case study of the vertical merger between Turner Broadcasting and Time Warner', *International Journal of Industrial Organization* 27(4), 523–543.

Villas-Boas, Sofia B. (2007). 'Vertical Relationships between Manufacturers and Retailers: Inference with Limited Data', *Review of Economic Studies* 74(2), 625–652.

Waterman, D. H. and Weiss, A. A. (1996). 'The Effects of Vertical Integration between Cable Television Systems and Pay Cable Networks', *Journal of Econometrics* 72(1-2), 357–95.

Whinston, Michael D. (2006), *Lectures on Antitrust Economics*, MIT Press, Cambridge, MA.

Willamson, Oliver. (1985), *The economic institutions of capitalism*, New York, NY.

A Further Estimation and Computational Details

A.1 Solving for Negotiated Input Fees and Bundle Marginal Costs

We will omit the subscript on $\Psi_{fct} \equiv (1-\zeta_{fct})/\zeta_{fct}$ for the expressions in this subsection. Let $\mathcal{B}^R_{fmt}$ be the observed set of RSNs carried by f in market m in period t.

Consider MVPD f bargaining with channel c over input fee τ_{fct}. Closed form expressions for MVPD and channel "GFT" terms defined in (8) can be derived as follows:

$$\begin{aligned} GFT^M_{fct} = \sum_{m \in \mathcal{M}_{fct}} \Bigg[& \Big[\mu_{fct} D_{fmt} - D^{\backslash fc}_{fmt}\Big]\tau_{fct} + \mu_{fct}(D_{fmt} + \sum_{g \neq f : c \in \mathcal{B}_{gmt}} [\Delta_{fc} D_{gmt}]) a_{cmt} \\ & + \mu_{fct} \sum_{g \neq f : c \in \mathcal{B}_{gmt}} [\Delta_{fc} D_{gmt}] \tau_{gct} + \sum_{d \in \mathcal{V}_{ft} \backslash c} \sum_{g \in \mathcal{F}_{mt} : d \in \mathcal{B}_{gmt}} [\Delta_{fc} D_{gmt}] \mu_{fdt} (\tau_{gdt} + a_{dmt}) \\ & + [\Delta_{fc} D_{fmt}] \Big(p_{fmt} - mc_{fmt}\Big) \Bigg] \end{aligned} \tag{17}$$

$$\begin{aligned} GFT^C_{fct} = \sum_{m \in \mathcal{M}_{fct}} \Bigg[& (D_{fmt} - \mu_{fct} D^{\backslash fc}_{fmt})\tau_{fct} + (D_{fmt} + \sum_{g \neq f : c \in \mathcal{B}_{gmt}} [\Delta_{fc} D_{gmt}]) a_{cmt} + \sum_{g \neq f : c \in \mathcal{B}_{gmt}} [\Delta_{fc} D_{gmt}](\tau_{gct}) \\ & + \sum_{g \in \mathcal{F}_{mt}} \lambda_{R:fct} [\Delta_{fc} D_{gmt}] \sum_{d \in \mathcal{B}_{gmt} \backslash c} \mu^C_{cdt} (\tau_{gdt} + a_{dmt}) + \sum_{g \in \mathcal{F}_{mt}} \mu_{gct} \lambda_{R:fct} [\Delta_{fc} D_{gmt}] \Big(p_{gmt} - mc_{gmt}\Big) \Bigg] \end{aligned} \tag{18}$$

where: $D^{\backslash fc}_{fmt}$ is the demand for f in market m if it dropped channel c; $\lambda_{R:fct} = \lambda_R$ if f and c are not integrated, and $\lambda_{R:fct} = 1$ otherwise; $\mu_{fct} = \mu \times O_{fct}$; and $\mu^C_{cdt} = \mu \times O^C_{cdt}$.

Focus on the bargain between an RSN c and MVPD f.[45] Using (17) and (18), the Nash Bargaining first-order condition $\forall f \in \mathcal{F}_{mt}, c \in \mathcal{C}^R_t$ given by (10) ($GFT^C_{fct} = \Psi GFT^M_{fct}$) can be re-written as:

$$\begin{aligned} & \tau_{fct} \sum_{m \in \mathcal{M}_{fct}} \Big[(1+\Psi)(1-\mu_{fct}) D_{fmt}\Big] + \sum_{g \neq f : c \in \mathcal{B}_{gmt}} \tau_{gct} \sum_{m \in \mathcal{M}_{fct}} (1 - \Psi \mu_{fct})[\Delta_{fc} D_{gmt}] \\ & \quad + \sum_{g \in \mathcal{F}_{mt}} \sum_{d \in \mathcal{B}_{gmt} \backslash c} \tau_{gdt} ((\Psi - \mu_{fct}) \mathbb{1}_{g=f} + \mu^C_{cdt} \lambda_{R:fct} - \Psi \mu_{fdt}) \sum_{m \in \mathcal{M}_{fct}} [\Delta_{fc} D_{gmt}] \\ & \quad + (\Psi - \mu_{fct}) \sum_{m \in \mathcal{M}_{fct}} mc^{\backslash R}_{fmt} [\Delta_{fc} D_{fmt}] = \\ & \quad \sum_{m \in \mathcal{M}_{fct}} \Big[(\Psi - \mu_{fct})[\Delta_{fc} D_{fmt}] p_{fmt}\Big] \\ & \quad - \sum_{m \in \mathcal{M}_{fct}} \Bigg[a_{cmt} \Big((1 - \Psi \mu_{fct}) D_{fmt} + (1 - \Psi \mu_{fct}) \sum_{g \neq f : c \in \mathcal{B}_{gmt}} [\Delta_{fc} D_{gmt}]\Big) \\ & \quad + \sum_{g \in \mathcal{F}_{mt}} \sum_{d \in \mathcal{B}_{gmt} \backslash c} a_{dmt} (\mu^C_{cdt} \lambda_{R:fct} - \Psi \mu_{fdt})([\Delta_{fc} D_{gmt}]) \Bigg] \end{aligned} \tag{19}$$

where $mc^{\backslash R}_{fmt}$ represents non-RSN marginal costs: i.e., $mc^{\backslash R}_{fmt} \equiv mc_{fmt} - \sum_{d \in \mathcal{B}^R_{fmt}} \tau_{fdt}$.

We can also re-write the pricing first-order condition in (5), which provides the optimal set of prices for

---

[45] In estimation, we are assuming that $\lambda_R = 0$ in the "non-loophole" markets, and thus omit terms that would otherwise enter (e.g., if c were integrated with a rival MVPD f'). In the counterfactuals, we re-introduce these terms.)

every cable provider f in every market m, as:

$$\sum_{g\in\mathcal{F}_{mt}} \frac{\partial D_{gmt}}{\partial p_{fmt}} \Big(mc^{\backslash R}_{gmt} \mathbb{1}_{g=f} + \sum_{d\in\mathcal{B}^R_{gmt}} (\mathbb{1}_{g=f} - \mu_{fdt})\tau_{gdt} \Big) = \Big[D_{fmt} + \frac{\partial D_{fmt}}{\partial p_{fmt}} p_{fmt} + \sum_{g\in\mathcal{F}_{mt}} \frac{\partial D_{gmt}}{\partial p_{fmt}} \sum_{d\in\mathcal{B}^R_{gmt}} \mu_{fdt} a_{dmt} \Big] \tag{20}$$

However, if f is a satellite provider (denoted $f \in \mathcal{F}^{sat}$), we assume that there is a single national price p_{ft} and non-RSN marginal cost $\hat{mc}^{\backslash R}_{fmt}$ that applies across all markets; this implies that there is only a single pricing first-order condition for satellite firms:

$$\sum_m \sum_{g\in\mathcal{F}_{mt}} \frac{\partial D_{gmt}}{\partial p_{ft}} \Big(mc^{\backslash R}_{gt} \mathbb{1}_{g=f} + \sum_{d\in\mathcal{B}^R_{gmt}} (\mathbb{1}_{g=f} - \mu_{fdt})\tau_{gdt} \Big) = \tag{21}$$

$$\sum_m \Big(D_{fmt} + \frac{\partial D_{fmt}}{\partial p_{ft}} p_{ft} + \sum_{g\in\mathcal{F}_{mt}} \frac{\partial D_{gmt}}{\partial p_{ft}} \sum_{d\in\mathcal{B}^R_{gmt}} \mu_{fdt} a_{dmt} \Big) \qquad \forall f \in \mathcal{F}^{sat}$$

Equations (19), (20), and (21) express input fees and marginal costs as a function of demand parameters, prices, and advertising rates. We thus solve for the vector of RSN input fees $\{\tau_{fct}\}_{\forall f,t,c\in\mathcal{C}^R_t}$ for all RSNs and non-RSN bundle marginal costs $\{mc^{\backslash R}_{fmt}\}_{\forall fmt}$ via matrix inversion when evaluating the objective for any parameter vector $\boldsymbol{\theta}$.

National Channels. We use our estimates of RSN input fees and non-RSN bundle marginal costs to recover $\{\tau_{fct}\}_{\forall ft,c\notin\mathcal{C}^R_t}$ for non-RSN channels via matrix inversion on the following:

$$\tau_{fct} \sum_{m\in\mathcal{M}_{fct}} \Big[D_{fmt} + \Psi D^{\backslash fc}_{fmt} \Big] + \sum_{g\neq f: c\in\mathcal{B}_{gmt}} \tau_{gct} \sum_{m\in\mathcal{M}_{fct}} [\Delta_{fc} D_{gmt}] = \tag{22}$$

$$\sum_{m\in\mathcal{M}_{fct}} \Bigg[(\Psi)[\Delta_{fc} D_{fmt}](p_{fmt} - \hat{mc}_{fmt}) \Bigg] + \sum_{g\in\mathcal{F}_{mt}} \sum_{d\in\mathcal{B}_{gmt}\backslash c} \mu_{fdt} \Psi \hat{\tau}_{gdt} \sum_{m\in\mathcal{M}_{fct}} [\Delta_{fc} D_{gmt}]$$

$$- \sum_{m\in\mathcal{M}_{fct}} \Bigg[a_{cmt} \Big(D_{fmt} + \sum_{g\neq f: c\in\mathcal{B}_{gmt}} [\Delta_{fc} D_{gmt}] \Big) + \sum_{g\in\mathcal{F}_{mt}} \sum_{d\in\mathcal{B}_{gmt}\backslash c} a_{dmt} (-\Psi \mu_{fdt}) ([\Delta_{fc} D_{gmt}]) \Bigg]$$

where we construct estimates of each bundle's marginal costs from our recovered non-RSN marginal costs as follows: $\hat{mc}_{fmt} \equiv \hat{mc}^{\backslash R}_{fmt} + \sum_{d\in\mathcal{B}^R_{fmt}} \hat{\tau}_{fdt}$. We assume away integration incentives for non-RSNs so that $\mu_{fct} = 0 \; \forall ft, c \notin \mathcal{C}^R_t$.

A.2 Computation of Disagreement Payoffs

Computation of several moments requires estimating $\Delta_{fc}[\Pi^M_{fmt}(\mathcal{B}_{mt}, \boldsymbol{p}_{mt}, \{\hat{\boldsymbol{\tau}}_{fct}, \boldsymbol{\tau}_{-fc,t}\})]$ and
$\Delta_{fc}[\Pi^C_{cmt}(\mathcal{B}_{mt}, \boldsymbol{p}_{mt}, \{\hat{\boldsymbol{\tau}}_{fct}, \boldsymbol{\tau}_{-fc,t}\}; \lambda_R)]$ for each MVPD f and channel c that contract in each period. These "gains from trade" for each pair are comprised of agreement and disagreement profits.

Profits from agreement (as a function of $\boldsymbol{\theta}$) can be computed from observed prices and bundle composition using MVPD and Channel profits specified by (4) and (7). Profits from disagreement between MVPD f and channel c are recomputed in each market given the following assumptions:

1. Bundle composition does not change for other MVPDs: $\mathcal{B}'_{gmt} = \mathcal{B}_{gmt} \; \forall \; g \neq f$; bundles for MVPD f just drop c, but do not adjust otherwise;
2. Input prices $\hat{\tau}_{-fc,t}$ for all other MVPD-conglomerate pairs do not adjust;
3. Bundle prices for satellite and cable providers do not adjust.

The second and third assumptions are consistent with the timing of our game and the simultaneous determination of input and bundle prices.

A.3 Recomputing Counterfactual Equilibria when Channels are Added or Removed from Satellite

When we explore counterfactuals when a RSN channel c is either added or removed from satellite providers (and potentially un-integrated), we compute market outcomes when input and bundle prices are allowed to re-equilibrate. Note that this is different than in the previous subsection, where we explore the computation of disagreement points which occur off the equilibrium path, since here changes are anticipated by all players (e.g., if the terrestrial loophole were closed). We assume that:

1. satellite distributors either carry or do not carry c in all (relevant) markets, and (with national pricing) do not change the prices of its bundles;
2. cable systems may change their prices (since demand elasticities may be affected by changes in carriage) but do not change any carriage or bundling decisions;
3. input prices of RSNs (but not national channels) are allowed to adjust.

We compute the new counterfactual equilibrium where c is either now supplied or removed from satellite in a given period t as follows:

1. Given new bundles $\{\mathcal{B}^{R,CF}_{fmt}\}$ and potentially new values for $\{\lambda^{CF}_{R:fct}, \mu^{CF}_{fct}\}$, we iterate on the following until we obtain convergence on counterfactual input prices $\{\tau^{CF}_{fct}\}$, bundle prices $\{p^{CF}_{fmt}\}$, bundle demands $\{D^{CF}_{fmt}\}$, and elasticities $\{\partial s^{CF}_{fmt}/\partial p_{gmt}\}$:
   (a) Solve for the values of $\{\tau^{CF}_{fct}\}_{c\in\mathcal{C}^{RSN}_t}$ given values of $\{D^{CF}_{fmt}\}$, $\{\partial s^{CF}_{fmt}/\partial p_{gmt}\}$, $\{\hat{mc}^{\setminus R}_{fmt}\}$, μ, λ_R, Ψ using the following system of equations:

$$\begin{aligned}
&\tau^{CF}_{fct}\sum_{m\in\mathcal{M}_{fct}}\Big[(1+\Psi)(1-\mu_{fct})D^{CF}_{fmt}\Big]+\sum_{g\neq f:c\in\mathcal{B}^{R,CF}_{gmt}}\tau^{CF}_{gct}\sum_{m\in\mathcal{M}_{fct}}(1-\Psi\mu_{fct}-\mu_{gct}\lambda_R)[\Delta_{fc}D^{CF}_{gmt}] \\
&\quad+\sum_{g\in\mathcal{F}_{mt}}\sum_{d\in\mathcal{B}^{R,CF}_{gmt}\setminus c}\tau^{CF}_{gdt}((\Psi-\mu_{fct})\mathbb{1}_{g=f}+\mu^{C}_{cdt}\lambda_{R:fct}-\Psi\mu_{fdt}-\mu_{gct}\lambda_R)\sum_{m\in\mathcal{M}_{fct}}[\Delta_{fc}D^{CF}_{gmt}]= \\
&\quad\sum_{m\in\mathcal{M}_{fct}}\Big[(\Psi-\mu_{fct})(p^{CF}_{fmt}-\hat{mc}^{\setminus R}_{fmt})[\Delta_{fc}D^{CF}_{fmt}]-\mu_{f'ct}\lambda_R(p^{CF}_{fmt}-\hat{mc}^{\setminus R}_{f'mt})[\Delta_{fc}D^{CF}_{f'mt}]\Big] \\
&\quad-\sum_{m\in\mathcal{M}_{fct}}\Bigg[a_{cmt}\Big((1-\Psi\mu_{fct})D^{CF}_{fmt}+(1-\Psi\mu_{fct})\sum_{g\neq f:c\in\mathcal{B}^{R,CF}_{gmt}}[\Delta_{fc}D^{CF}_{gmt}]\Big) \\
&\quad+\sum_{g\in\mathcal{F}_{mt}}\sum_{d\in\mathcal{B}^{R,CF}_{gmt}\setminus c}a_{dmt}(\mu^{C}_{cdt}\lambda_{R:fct}-\Psi\mu_{fdt})([\Delta_{fc}D^{CF}_{gmt}])\Bigg]\qquad\forall f,c
\end{aligned}\tag{23}$$

   where f and f' represent the MVPDs with which c is potentially integrated. Equation (23) differs from (19) insofar that we now allow for the possibility that $\lambda_R > 0$, and that c may be integrated with a rival MVPD f' when bargaining with f.

   (b) Market by market, update bundle prices $\{p^{CF}_{fmt}\}$ for all cable distributors to maximize profits given new values of $\{\tau^{CF}_{fct}\}$. Update bundle demands $\{D^{CF}_{fmt}\}$ and elasticities $\{\partial s^{CF}_{fmt}/\partial p_{gmt}\}$ at the new computed prices.

   Currently we only update $\{\tau_{fct}\}_{\forall f}$ for the given channel c that is being examined, and not for other channels d that may be active in c's relevant markets.

B Additional Figures and Tables

Table 6: Regional Sports Networks Availability, Affiliate Fees, and Viewership

	Kagan Availability		Kagan Affiliate Fees					Nielsen Viewing			
	Systems Served	HH Served	Years	Mean	StDev	Min	Max	Obs	All HH	Has DTV	Has Dish
Comcast RSNs											
Comcast SportsNet Bay Area	137	4.7	11	$1.70	$0.53	$1.01	$2.52	720	0.41	0.45	0.33
Comcast SportsNet California	1,960	59.4	7	$0.91	$0.14	$0.75	$1.10	720	0.17	0.17	0.17
Comcast SportsNet Chicago	67	0.9	7	$2.02	$0.18	$1.90	$2.37	360	0.54	0.59	0.36
Comcast SportsNet Mid-Atlantic	23	1.7	11	$2.03	$0.74	$0.85	$3.10	1,440	0.13	0.09	0.03
Comcast SportsNet New England	15	1.0	11	$1.26	$0.32	$0.90	$1.89	1,080	0.27	0.30	0.17
Comcast SportsNet Northwest	137	4.7	4	$1.93	$0.09	$1.81	$2.04	—	—	—	—
Comcast SportsNet Philadelphia	135	10.0	11	$1.94	$0.61	$1.05	$2.85	360	0.91	0.06	0.05
Comcast SportsNet Southwest	335	5.7	—	—	—	—	—	—	—	—	—
Comcast/Charter Sports Southeast	194	6.2	11	$0.36	$0.09	$0.20	$0.50	3,600	0.04	0.00	0.00
The mtn	195	7.0	5	$0.20	$0.02	$0.19	$0.23	720	0.04	0.05	0.00
News Corp RSNs											
Fox Sports Arizona	106	3.7	11	$1.58	$0.50	$0.82	$2.28	—	—	—	—
Fox Sports Chicago	342	4.8	7	$1.45	$0.44	$1.08	$2.13	—	—	—	—
Fox Sports Detroit	284	5.3	11	$1.75	$0.45	$1.05	$2.34	360	1.02	0.94	0.68
Fox Sports Florida	152	6.7	11	$1.34	$0.33	$0.90	$1.95	2,160	0.14	0.12	0.12
Fox Sports Houston	48	3.3	—	—	—	—	—	—	—	—	—
Fox Sports Midwest	695	7.4	11	$1.42	$0.44	$0.57	$2.01	1,800	0.31	0.31	0.26
Fox Sports North	620	4.5	11	$1.97	$0.60	$1.15	$2.88	720	0.79	1.04	0.70
Fox Sports Ohio	306	7.0	11	$1.61	$0.49	$0.75	$2.42	2,160	0.34	0.31	0.29
Fox Sports South	905	15.3	17	$1.63	$0.52	$0.52	$2.17	3,600	0.13	0.08	0.07
Fox Sports Southwest	924	12.7	11	$1.68	$0.50	$0.80	$2.43	5,040	0.14	0.15	0.12
Fox Sports West	167	9.2	11	$1.80	$0.44	$0.87	$2.35	1,080	0.16	0.12	0.07
Fox Sports Wisconsin	136	2.2	—	—	—	—	—	—	—	—	—
Big Ten Network	1,960	59.4	—	—	—	—	—	—	—	—	—
Prime Ticket (New)	132	8.2	11	$1.52	$0.46	$0.60	$2.07	720	0.16	0.12	0.09
SportSouth (New)	532	11.3	11	$0.31	$0.13	$0.15	$0.52	—	—	—	—
Sun Sports	234	8.3	11	$1.36	$0.54	$0.55	$2.27	2,160	0.20	0.16	0.12
Liberty RSNs											
Root Sports Northwest	281	5.4	11	$1.73	$0.52	$0.70	$2.54	—	—	—	—
Root Sports Pittsburgh	316	4.5	11	$1.81	$0.53	$1.05	$2.55	—	—	—	—
Root Sports Rocky Mountain	479	5.4	11	$1.58	$0.42	$0.75	$2.06	—	—	—	—
Cablevision RSNs											
Madison Sq. Garden (MSG)	219	9.9	11	$1.82	$0.30	$1.45	$2.44	1,080	0.23	0.24	0.17
MSG Plus	165	7.5	11	$1.24	$0.15	$1.01	$1.61	360	0.07	0.05	0.06
Cox RSNs											
Channel 4 San Diego	15	1.0	11	$0.87	$0.26	$0.53	$1.32	360	0.48	0.03	0.00
Cox Sports Television	70	2.1	9	$0.55	$0.05	$0.50	$0.64	360	0.22	0.01	0.08
Time Warner RSNs											
Metro Sports Network	8	0.6	—	—	—	—	—	—	—	—	—
SportsNet New York	314	20.1	5	$1.91	$0.18	$1.71	$2.20	1,080	0.13	0.13	0.09
Independent/Other RSNs											
Altitude Sports & Entertainment	130	2.8	7	$1.99	$0.29	$1.70	$2.47	360	0.24	0.21	0.22
Bright House Sports Network	—	—	—	—	—	—	—	360	0.02	0.00	0.00
Empire Sports Network	87	1.9	—	—	—	—	—	—	—	—	—
Mid-Atlantic Sports Network (MASN)	109	5.2	6	$1.58	$0.12	$1.45	$1.77	1,440	0.13	0.10	0.13
New England Sports Network (NESN)	213	4.5	11	$1.99	$0.49	$1.30	$2.72	1,080	0.95	1.00	0.48
Royals Sports	18	0.2	6	$0.19	$0.02	$0.16	$0.21	—	—	—	—
SportsTime Ohio	196	9.0	5	$1.51	$0.17	$1.30	$1.73	720	0.33	0.40	0.20
Yankees Entertainment & Sports (YES)	304	15.8	9	$2.13	$0.41	$1.18	$2.62	1,440	0.27	0.30	0.00

Notes: Reported are availability, affiliate fees and average viewing of the major Regional Sports Networks (RSNs) in the United States. Affiliate fees are the monthly per-subscriber fees paid by cable and satellite distributors to television networks for the right to distribute the network's programming to subscribers. Availability and affiliate fee information is provided by SNL Kagan as part of its Media & Communications Package. RSN viewership is from Nielsen and covers 2000-2010.

Table 7: Sample Statistics - Prices, Market Shares, and Channels

		Unweighted				**Weighted by Households**			
	Obs	**Mean**	**StdDev**	**Min**	**Max**	**Mean**	**StdDev**	**Min**	**Max**
Total Markets	6,138	6,138							
Average Households (M)	6,138					31.5			
Cable									
Year	6,138	2004	2.9	2000	2010	2004	2.8	2000	2010
Price	6,138	$51.40	$10.33	$8.67	$130.96	$53.02	$8.84	$8.67	$130.96
Market Share	6,138	0.624	0.161	0.005	0.965	0.630	0.137	0.005	0.965
Cable Networks	6,138	42.6	15.4	0	87	44.9	14.0	0	87
RSNs	6,138	1.6	0.9	0	5	1.8	0.9	0	5
Total Channels	6,138	44.2	15.9	1	90	46.6	14.5	1	90
DirecTV									
Year	6,138	2004	2.9	2000	2010	2004	2.8	2000	2010
Price	6,138	$53.25	$6.57	$46.05	$76.73	$53.27	$6.34	$46.05	$76.73
Market Share	6,138	0.092	0.062	0.002	0.499	0.094	0.064	0.002	0.499
Cable Networks	6,138	80.5	10.3	66	97	81.2	10.1	66	97
RSNs	6,138	1.7	0.9	0	6	1.9	0.9	0	6
Total Channels	6,138	82.2	10.5	66	103	83.0	10.3	66	103
Dish									
Year	6,138	2004	2.9	2000	2010	2004	2.8	2000	2010
Price	6,138	$53.89	$4.75	$44.28	$68.33	$53.96	$4.53	$44.28	$68.33
Market Share	6,138	0.064	0.055	0.000	0.406	0.059	0.052	0.000	0.406
Cable Networks	6,138	70.8	13.2	54	91	71.8	12.9	54	91
RSNs	6,138	1.6	0.8	0	5	1.7	0.7	0	5
Total Channels	6,138	72.4	13.3	54	96	73.5	13.0	54	96

Notes: Reported are the price, market share, and cable, Regional Sport Network (RSN), and total channels for each of the local cable operators and two national satellite providers serving each of our markets. Markets are defined as the set of continuous zip codes within a cable system facing the same portfolio of competitors. We exclude (the relatively few) markets facing competition between cable operators. All the data cover the years 2000-2010. To be included, we required information on each of price, market share, and channels. Cable system subscriber and channel information is from the Nielsen FOCUS dataset. Cable system price information is drawn from the Internet Archive, newspaper reports, and the TNS Bill Harvesting database. Satellite system channel and price information is drawn from the Internet Archive. Cable and satellite subscriber market shares are estimated from the MRI (2000-2007) and Simmons (2008-2010) household surveys. We restrict attention to those markets with at least 5 observations in any year. See the text for more details.

Table 8: Sample Statistics - National Cable Channel Affiliate Fees and Viewership, Part 1

	Affiliate Fees					**Viewership**					
	Kagan					**Nielsen**		**Combined MRI / Simmons**			
	Years	**Mean**	**StDev**	**Min**	**Max**	**Obs**	**Mean**	**Obs**	**Mean**	**SDev**	**Percent Positive**
Channels (A-L)											
ABC Family Channel	11	\$0.19	\$0.02	\$0.16	\$0.22	747	0.418	277,535	0.344	1.149	0.176
AMC	11	\$0.22	\$0.02	\$0.20	\$0.25	747	0.491	277,535	0.351	1.183	0.156
Animal Planet	11	\$0.07	\$0.01	\$0.06	\$0.09	747	0.275	277,535	0.344	1.108	0.203
A&E	11	\$0.21	\$0.03	\$0.16	\$0.26	747	0.664	277,535	0.472	1.373	0.230
BBC America	11	\$0.09	\$0.03	\$0.03	\$0.12	703	0.053	225,618	0.091	0.617	0.041
BET	11	\$0.14	\$0.02	\$0.11	\$0.17	747	0.382	277,535	0.184	1.017	0.070
Bio	11	\$0.07	\$0.03	\$0.00	\$0.11	447	0.082	98,567	0.104	0.618	0.023
Bloomberg Television	11	\$0.04	\$0.02	\$0.02	\$0.06	—	—	150,165	0.029	0.373	0.010
Boomerang	10	\$0.05	\$0.03	\$0.00	\$0.08	280	0.131	—	—	—	—
Bravo	11	\$0.15	\$0.03	\$0.11	\$0.20	747	0.277	277,535	0.169	0.804	0.092
Cartoon Network	11	\$0.14	\$0.03	\$0.08	\$0.18	747	0.989	277,535	0.231	1.098	0.106
CMT	11	\$0.06	\$0.02	\$0.01	\$0.08	—	—	277,535	0.120	0.732	0.067
CNBC	11	\$0.24	\$0.04	\$0.16	\$0.30	747	0.217	277,535	0.313	1.185	0.170
CNN	11	\$0.43	\$0.05	\$0.35	\$0.52	747	0.550	277,535	0.701	1.744	0.319
CNN en Espanol	—	—	—	—	—	463	0.013	—	—	—	—
CNN International	11	\$0.11	\$0.02	\$0.09	\$0.13	567	0.012	—	—	—	—
Comedy Central	11	\$0.11	\$0.02	\$0.08	\$0.14	747	0.449	277,535	0.280	0.997	0.162
Discovery Channel	11	\$0.27	\$0.04	\$0.22	\$0.35	747	0.535	277,535	0.628	1.462	0.327
Disney Channel	11	\$0.81	\$0.06	\$0.75	\$0.91	747	1.171	277,535	0.246	1.074	0.116
E! Entertainment TV	11	\$0.19	\$0.02	\$0.15	\$0.21	747	0.315	277,535	0.201	0.788	0.137
ESPN	11	\$2.81	\$1.12	\$1.14	\$4.34	747	0.836	277,535	0.675	1.767	0.257
ESPN 2	11	\$0.37	\$0.14	\$0.17	\$0.58	747	0.262	277,535	0.334	1.220	0.151
ESPN Classic Sports	11	\$0.14	\$0.03	\$0.10	\$0.18	636	0.037	277,535	0.072	0.521	0.047
ESPN deportes	—	—	—	—	—	280	0.035	—	—	—	—
ESPNews	11	\$0.10	\$0.06	\$0.02	\$0.17	636	0.043	277,535	0.143	0.782	0.084
ESPNU	6	\$0.14	\$0.03	\$0.10	\$0.17	280	0.037	—	—	—	—
Fine Living Network	—	—	—	—	—	55	0.003	150,165	0.025	0.324	0.009
FitTV	11	\$0.05	\$0.02	\$0.02	\$0.07	205	0.005	—	—	—	—
Flix	—	—	—	—	—	—	—	101,275	0.013	0.165	0.004
Food Network	11	\$0.06	\$0.03	\$0.03	\$0.14	747	0.411	277,535	0.396	1.364	0.175
Fox News Channel	11	\$0.32	\$0.18	\$0.17	\$0.70	747	0.785	277,535	0.697	1.961	0.267
Fuse	11	\$0.06	\$0.01	\$0.05	\$0.08	747	0.024	225,618	0.018	0.308	0.009
FX	11	\$0.34	\$0.06	\$0.27	\$0.43	747	0.463	277,535	0.258	0.976	0.137
G4	9	\$0.07	\$0.02	\$0.05	\$0.09	591	0.051	225,618	0.036	0.411	0.016
GSN	11	\$0.07	\$0.03	\$0.04	\$0.10	747	0.154	277,535	0.088	0.703	0.036
Golf Channel	11	\$0.20	\$0.05	\$0.13	\$0.26	580	0.065	277,535	0.084	0.633	0.041
Hallmark Channel	11	\$0.04	\$0.02	\$0.01	\$0.06	699	0.307	225,618	0.301	1.268	0.088
Headline News	—	—	—	—	—	747	0.214	277,535	0.278	0.983	0.173
HGTV	11	\$0.08	\$0.04	\$0.03	\$0.14	747	0.500	277,535	0.397	1.446	0.162
History Channel	11	\$0.18	\$0.04	\$0.13	\$0.23	747	0.531	277,535	0.531	1.462	0.251
HSN	—	—	—	—	—	580	0.038	252,217	0.044	0.395	0.031
IFC	11	\$0.18	\$0.01	\$0.17	\$0.19	—	—	277,535	0.045	0.424	0.023
Investigation Discovery	11	\$0.04	\$0.03	\$0.00	\$0.07	441	0.121	174,621	0.067	0.628	0.018
Lifetime	11	\$0.21	\$0.06	\$0.13	\$0.29	747	0.679	277,535	0.554	1.650	0.199
Lifetime Movie Network	11	\$0.07	\$0.03	\$0.00	\$0.09	328	0.185	225,618	0.250	1.174	0.068

Notes: Reported are affiliate fees and average viewing of the major cable television networks included in our demand system. Affiliate fees are the monthly per-subscriber fees paid by cable and satellite distributors to television networks for the right to distribute the network's programming to subscribers. Affiliate fee information is provided by SNL Kagan as part of its Media & Communications Package. Nielsen viewership data reports the average rating on each channel across between 44 and 56 Designated Market Areas (DMAs) between 2000 and 2010. MRI / Simmons viewership data reports the average viewership and the percent of households with any (positive) viewership of each channel by households in the MRI (2000-2007) and Simmons (2008-2010) household surveys.

Table 9: Sample Statistics - National Cable Channel Affiliate Fees and Viewership, Part 2

	Affiliate Fees					**Viewership**					
	Kagan					**Nielsen**		**Combined MRI / Simmons**			
	Years	**Mean**	**StDev**	**Min**	**Max**	**Obs**	**Mean**	**Obs**	**Mean**	**SDev**	**Percent Positive**
Channels (M-Z)											
MSNBC	11	$0.14	$0.02	$0.12	$0.17	747	0.343	277,535	0.330	1.181	0.182
MTV	11	$0.27	$0.05	$0.20	$0.35	747	0.568	277,535	0.235	0.983	0.127
MTV Hits	11	$0.01	$0.00	$0.01	$0.01	280	0.030	—	—	—	—
MTV Jams	11	$0.01	$0.01	$0.01	$0.02	280	0.038	—	—	—	—
MTV2	11	$0.03	$0.01	$0.01	$0.05	601	0.082	277,535	0.070	0.542	0.042
Nat Geo Wild	6	$0.07	$0.02	$0.04	$0.09	112	0.068	—	—	—	—
Nat Geo Channel	11	$0.17	$0.06	$0.00	$0.21	608	0.136	225,618	0.212	0.883	0.096
NBA TV	11	$0.31	$0.06	$0.19	$0.37	280	0.035	—	—	—	—
NBC Sports / Versus	8	$0.50	$0.33	$0.11	$0.85	55	0.047	—	—	—	—
NFL Network	4	$0.45	$0.09	$0.32	$0.53	56	0.027	—	—	—	—
NHL Network	11	$0.11	$0.06	$0.00	$0.18	376	0.082	—	—	—	—
Nickelodeon	11	$0.37	$0.05	$0.29	$0.47	747	1.555	277,535	0.200	0.991	0.096
NickToons TV	9	$0.05	$0.03	$0.00	$0.07	447	0.128	—	—	—	—
OWN	11	$0.06	$0.03	$0.00	$0.09	280	0.130	—	—	—	—
Outdoor Channel	11	$0.04	$0.01	$0.03	$0.05	—	—	174,621	0.068	0.594	0.021
Ovation	11	$0.06	$0.02	$0.03	$0.08	280	0.027	—	—	—	—
Oxygen	11	$0.07	$0.04	$0.00	$0.10	656	0.131	225,618	0.114	0.658	0.052
ReelzChannel	—	—	—	—	—	280	0.033	—	—	—	—
Science Channel	11	$0.04	$0.02	$0.00	$0.07	592	0.072	174,621	0.092	0.635	0.030
ShopNBC	—	—	—	—	—	280	0.025	—	—	—	—
SoapNet	11	$0.11	$0.05	$0.02	$0.15	656	0.135	174,621	0.109	0.833	0.022
Speed Channel	11	$0.17	$0.03	$0.11	$0.21	747	0.091	277,535	0.097	0.679	0.046
Style Network	11	$0.10	$0.04	$0.03	$0.14	646	0.063	225,618	0.040	0.416	0.019
Sundance Channel	11	$0.23	$0.04	$0.16	$0.27	—	—	174,621	0.037	0.397	0.012
SyFy	11	$0.17	$0.04	$0.12	$0.22	747	0.427	277,535	0.301	1.207	0.126
TBS	11	$0.37	$0.12	$0.19	$0.54	747	0.905	277,535	0.497	1.345	0.243
TechTV	4	$0.02	$0.01	$0.00	$0.03	47	0.006	51,917	0.012	0.202	0.002
The Hub	11	$0.04	$0.02	$0.01	$0.06	441	0.037	—	—	—	—
TLC	11	$0.16	$0.01	$0.14	$0.17	747	0.422	277,535	0.342	1.151	0.173
Toon Disney	—	—	—	—	—	376	0.146	177,590	0.096	0.644	0.034
Travel Channel	11	$0.07	$0.02	$0.04	$0.11	747	0.166	277,535	0.157	0.712	0.106
truTV	11	$0.09	$0.01	$0.08	$0.10	747	0.384	277,535	0.233	1.081	0.101
Turner Classic Movies	11	$0.22	$0.03	$0.16	$0.27	580	0.286	277,535	0.268	1.142	0.105
TNT	11	$0.83	$0.16	$0.55	$1.10	747	1.219	277,535	0.592	1.553	0.263
TV Guide Network	11	$0.03	$0.01	$0.02	$0.05	656	0.101	277,535	0.082	0.488	0.082
TV Land	11	$0.08	$0.03	$0.01	$0.12	376	0.412	277,535	0.190	0.979	0.086
TV One	7	$0.03	$0.03	$0.00	$0.08	280	0.129	123,885	0.050	0.572	0.008
USA	11	$0.46	$0.07	$0.36	$0.57	747	1.081	277,535	0.503	1.442	0.230
VH1	11	$0.12	$0.02	$0.09	$0.16	747	0.336	277,535	0.151	0.717	0.101
VH1 Classic	11	$0.05	$0.01	$0.02	$0.07	55	0.024	149,303	0.044	0.422	0.016
Weather Channel	11	$0.10	$0.01	$0.08	$0.12	747	0.234	204,189	0.380	0.879	0.266
WE	11	$0.09	$0.01	$0.07	$0.11	328	0.084	225,618	0.096	0.621	0.041

Notes: Reported are affiliate fees and average viewing of the major cable television networks included in our demand system. Affiliate fees are the monthly per-subscriber fees paid by cable and satellite distributors to television networks for the right to distribute the network's programming to subscribers. Affiliate fee information is provided by SNL Kagan as part of its Media & Communications Package. Nielsen viewership data reports the average rating on each channel across between 44 and 56 Designated Market Areas (DMAs) between 2000 and 2010. MRI / Simmons viewership data reports the average viewership and the percent of households with any (positive) viewership of each channel by households in the MRI (2000-2007) and Simmons (2008-2010) household surveys.

Table 10: Monthly WTP for Non-RSNs

Channel Name	Mean WTP	Fraction Positive	Mean Among Positive
ABC Family Channel	0.60	0.79	0.76
AMC	0.68	0.70	0.97
Animal Planet	0.46	0.67	0.68
Arts Entertainment AE	1.06	0.64	1.64
BET	0.60	0.36	1.65
Bravo	0.49	0.58	0.85
Cartoon Network	0.79	0.54	1.45
CMT	0.24	0.40	0.60
CNBC	0.55	0.62	0.89
CNN	1.21	0.77	1.58
Comedy Central	0.76	0.59	1.30
Discovery Channel	1.23	0.91	1.36
Disney Channel	0.74	0.62	1.20
E Entertainment TV	0.45	0.59	0.77
ESPN	8.67	0.71	12.15
ESPN 2	4.23	0.57	7.41
ESPN Classic	1.39	0.43	3.22
Food Network	0.92	0.71	1.29
Fox News Channel	1.61	0.81	2.01
FX	0.70	0.65	1.08
Golf Channel	0.19	0.36	0.53
Hallmark Channel	0.69	0.60	1.15
Headline News	0.52	0.64	0.82
HGTV	0.97	0.75	1.29
History Channel	1.18	0.78	1.51
Lifetime	1.13	0.67	1.70
MSNBC	0.70	0.67	1.05
MTV	0.74	0.56	1.32
Nickelodeon	1.40	0.51	2.75
SyFy, Sci-Fi	0.82	0.55	1.48
TBS	1.21	0.77	1.58
TLC	0.80	0.69	1.15
truTV, Court TV	0.72	0.46	1.56
Turner Classic Movies	0.68	0.57	1.18
TNT	1.94	0.81	2.41
USA	1.50	0.74	2.02
VH1	0.45	0.55	0.81
Weather Channel	0.66	0.87	0.75

Notes: This table presented estimated mean monthly willingness-to-pay in dollars for 38 national channels in 2007. The first column is the unconditional mean. The second column is the fraction of consumers with positive valuations. The third column is the mean conditional on having a positive valuation.

Table 11: Monthly WTP for RSNs

RSN Name	Mean WTP	Fraction Positive	Mean Among Positive
Altitude Sports	3.31	0.85	3.87
Channel 4 San Diego (4SD)	3.48	0.26	13.48
Comcast SportsNet (CSN) Bay Area	5.85	0.78	7.54
CSN California	3.21	0.52	6.15
CSN Chicago	3.74	0.78	4.79
CSN Mid-Atlantic	3.83	0.65	5.90
CSN New England	4.16	0.81	5.13
CSN Northwest	5.49	0.71	7.71
CSN Philadelphia	5.65	1.00	5.66
ComcastCharter Sports Southeast (CSS)	2.36	0.91	2.59
Cox Sports TV	3.06	0.64	4.79
Fox Sports Detroit	4.65	0.85	5.50
Fox Sports Florida	0.98	0.78	1.26
Fox Sports Midwest	0.56	0.21	2.72
Fox Sports North	2.74	0.95	2.87
Fox Sports Ohio	4.79	1.00	4.81
Fox Sports South	2.35	0.97	2.42
Fox Sports Southwest	1.73	0.99	1.75
Fox Sports West	3.24	1.00	3.25
MSG Plus	3.33	0.66	5.08
Madison Square Garden Network (MSG)	1.55	0.67	2.34
Mid-Atlantic Sports Network (MASN)	2.53	0.51	4.92
New England Sports Network (NESN)	11.94	0.51	23.35
Prime Ticket	2.22	0.59	3.75
Root Sports Northwest	5.58	0.94	5.94
Root Sports Pittsburgh	4.99	0.92	5.45
Root Sports Rocky Mountain	1.07	0.90	1.19
SportsNet New York (SNY)	1.31	0.52	2.55
Sun Sports	3.11	0.71	4.38
Yankees Entertainment Sports (YES)	2.82	0.60	4.68

Notes: This table presented estimated mean monthly willingness-to-pay in dollars for RSNs in 2007. The first column is the unconditional mean. The second column is the fraction of consumers with positive valuations. The third column is the mean amongst those with positive valuations.

Table 12: Simulated Market Outcomes for Integrated, Non-Loophole RSNs (1/2)

		(i) VI, no PARs		(ii) VI, PARs		(iii) No VI
		Level	% Change	Level	% Change	Level
CABLE OWNED RSNs (1/2)						
CSN BAY AREA	Avg Cable Mkt Share	0.65	1.62%	0.65	1.69%	0.64
Comcast	Avg Sat Mkt Share	0.21	-1.13%	0.21	-1.19%	0.21
Pop 6.03M	Avg Cable Prices	56.00	-0.96%	55.97	-1.01%	56.55
Footprint 54%	Aff Fees to Sat	2.64	60.82%	1.62	-1.65%	1.64
	Cable Surplus	34.55	0.13%	34.56	0.17%	34.51
Carriage +0%	Satellite Surplus	4.21	-3.89%	4.34	-1.10%	4.38
	RSN Surplus	1.56	6.48%	1.42	-3.00%	1.46
	Consumer Welfare	29.14	1.02%	29.16	1.08%	28.85
	Total Welfare	69.46	0.38%	69.48	0.40%	69.20
CSN CA	Avg Cable Mkt Share	0.70	0.48%	0.70	0.48%	0.69
Comcast	Avg Sat Mkt Share	0.18	-0.37%	0.18	-0.37%	0.18
Pop 3.86M	Avg Cable Prices	53.73	-0.28%	53.73	-0.28%	53.88
Footprint 10%	Aff Fees to Sat	1.15	68.57%	0.68	-0.38%	0.68
	Cable Surplus	44.57	-0.01%	44.57	-0.01%	44.57
Carriage +0%	Satellite Surplus	3.79	-0.52%	3.79	-0.37%	3.81
	RSN Surplus	0.08	11.86%	0.07	2.96%	0.07
	Consumer Welfare	28.07	0.27%	28.07	0.27%	27.99
	Total Welfare	76.50	0.08%	76.50	0.08%	76.44
CSN CHICAGO	Avg Cable Mkt Share	0.59	0.60%	0.59	0.65%	0.59
Comcast	Avg Sat Mkt Share	0.23	-0.40%	0.23	-0.44%	0.23
Pop 9.62M	Avg Cable Prices	61.54	-0.29%	61.53	-0.31%	61.72
Footprint 76%	Aff Fees to Sat	1.96	34.72%	1.45	-0.56%	1.45
	Cable Surplus	19.77	0.09%	19.77	0.13%	19.75
Carriage +1%	Satellite Surplus	5.27	-1.71%	5.34	-0.40%	5.36
	RSN Surplus	1.53	4.46%	1.45	-1.06%	1.47
	Consumer Welfare	27.70	0.35%	27.71	0.38%	27.61
	Total Welfare	54.27	0.16%	54.28	0.17%	54.18
CSN MID-ATL	Avg Cable Mkt Share	0.66	2.86%	0.66	2.12%	0.64
Comcast	Avg Sat Mkt Share	0.16	-11.84%	0.17	-2.19%	0.18
Pop 6.55M	Avg Cable Prices	58.44	0.70%	57.40	-1.09%	58.03
Footprint 70%	Aff Fees to Sat	-	-	1.20	-8.93%	1.32
	Cable Surplus	23.40	2.62%	22.88	0.33%	22.81
Carriage +9%	Satellite Surplus	3.68	-4.44%	3.80	-1.30%	3.85
	RSN Surplus	1.22	-10.94%	1.31	-4.50%	1.37
	Consumer Welfare	25.59	-2.76%	26.67	1.35%	26.31
	Total Welfare	53.89	-0.83%	54.66	0.59%	54.34
CSN NE	Avg Cable Mkt Share	0.64	3.37%	0.64	2.92%	0.62
Comcast	Avg Sat Mkt Share	0.11	-10.19%	0.12	-2.33%	0.12
Pop 5.2M	Avg Cable Prices	59.38	-0.30%	58.33	-2.06%	59.56
Footprint 85%	Aff Fees to Sat	-	-	1.19	-0.53%	1.20
	Cable Surplus	28.88	1.99%	28.36	0.17%	28.31
Carriage +4%	Satellite Surplus	2.32	-5.24%	2.39	-2.29%	2.44
	RSN Surplus	1.57	-8.07%	1.67	-2.30%	1.70
	Consumer Welfare	24.59	-0.95%	25.45	2.49%	24.83
	Total Welfare	57.35	0.11%	57.86	1.00%	57.29
MSG	Avg Cable Mkt Share	0.69	1.92%	0.68	0.85%	0.68
Cablevision	Avg Sat Mkt Share	0.14	-7.98%	0.15	-0.93%	0.15
Pop 11.7M	Avg Cable Prices	59.78	-0.10%	59.52	-0.53%	59.84
Footprint 41%	Aff Fees to Sat	-	-	0.72	2.48%	0.71
	Cable Surplus	27.99	1.32%	27.62	-0.01%	27.63
Carriage +0%	Satellite Surplus	2.86	-4.50%	2.96	-1.03%	2.99
	RSN Surplus	0.82	-10.75%	0.92	0.58%	0.92
	Consumer Welfare	26.45	-1.10%	26.90	0.59%	26.74
	Total Welfare	58.12	-0.28%	58.41	0.22%	58.28

Notes: This table presents observed and simulated market outcomes for individual RSNs in non-loophole markets. All reported figures except for market shares are in $/household/month, and all % changes are relative to specification (iii). Beneath the channel name is the name of the MVPD that owns the channel, the number of television households and the MVPD owner's footprint (% of households passed) in the RSN's relevant DMAs, and the % change in the integrated MVPD's households that obtain access to the channel upon moving from (iii) to (ii) in the RSN's relevant DMAs.

Table 13: Simulated Market Outcomes for Integrated, Non-Loophole RSNs (2/2)

		(i) VI, no PARs		(ii) VI, PARs		(iii) No VI
		Level	% Change	Level	% Change	Level
CABLE OWNED NON-LOOPHOLE RSNs (2/2)						
MSG PLUS	Avg Cable Mkt Share	0.70	1.16%	0.70	0.67%	0.70
Cablevision	Avg Sat Mkt Share	0.14	-4.58%	0.15	-0.83%	0.15
Pop 9.46M	Avg Cable Prices	59.58	-0.20%	59.46	-0.41%	59.70
Footprint 49%	Aff Fees to Sat	-	-	0.83	-0.01%	0.83
	Cable Surplus	29.98	0.93%	29.71	0.02%	29.70
Carriage +0%	Satellite Surplus	2.89	-0.78%	2.89	-0.81%	2.91
	RSN Surplus	0.62	-22.45%	0.79	-0.75%	0.79
	Consumer Welfare	27.67	-0.49%	27.95	0.52%	27.81
	Total Welfare	61.15	-0.10%	61.33	0.20%	61.21
SNY	Avg Cable Mkt Share	0.68	0.90%	0.68	0.90%	0.68
Comcast, TWC	Avg Sat Mkt Share	0.15	-0.97%	0.15	-0.97%	0.15
Pop 11.7M	Avg Cable Prices	59.52	-0.54%	59.52	-0.54%	59.84
Footprint 35%	Aff Fees to Sat	0.53	0.75%	0.53	-0.03%	0.53
	Cable Surplus	27.62	0.02%	27.62	0.02%	27.62
Carriage +14%	Satellite Surplus	2.96	-0.99%	2.96	-0.97%	2.99
	RSN Surplus	0.80	1.84%	0.80	1.76%	0.79
	Consumer Welfare	26.90	0.61%	26.90	0.61%	26.74
	Total Welfare	58.29	0.27%	58.29	0.27%	58.13
SATELLITE OWNED RSNs						
ROOT NW	Avg Cable Mkt Share	0.63	0.36%	0.63	0.90%	0.62
DirecTV	Avg Sat Mkt Share	0.22	-0.31%	0.22	-0.73%	0.22
Pop 4.15M	Avg Cable Prices	53.66	-0.30%	53.46	-0.68%	53.82
	Aff Fees to Rivals	2.74	22.94%	2.23	0.09%	2.23
	Cable Surplus	23.92	-1.67%	24.33	-0.01%	24.33
	Satellite Surplus	4.59	-1.52%	4.63	-0.73%	4.66
	RSN Surplus	2.95	18.99%	2.49	0.70%	2.48
	Consumer Welfare	28.53	0.28%	28.64	0.67%	28.45
	Total Welfare	59.99	0.12%	60.09	0.29%	59.92
ROOT PITT	Avg Cable Mkt Share	0.65	0.77%	0.65	1.02%	0.65
DirecTV	Avg Sat Mkt Share	0.16	-0.64%	0.16	-0.85%	0.17
Pop 5.09M	Avg Cable Prices	56.27	-0.35%	56.19	-0.49%	56.47
	Aff Fees to Rivals	2.59	12.41%	2.30	-0.52%	2.31
	Cable Surplus	27.63	-0.48%	27.76	0.00%	27.76
	Satellite Surplus	3.82	-1.04%	3.83	-0.85%	3.86
	RSN Surplus	1.78	9.06%	1.63	0.07%	1.63
	Consumer Welfare	25.50	0.42%	25.55	0.58%	25.40
	Total Welfare	58.73	0.13%	58.77	0.20%	58.65
ROOT ROCKY MTN	Avg Cable Mkt Share	0.52	-0.04%	0.52	0.00%	0.52
DirecTV	Avg Sat Mkt Share	0.31	0.03%	0.31	0.00%	0.31
Pop 4.19M	Avg Cable Prices	58.42	0.03%	58.40	0.00%	58.40
	Aff Fees to Rivals	-0.03	-62.93%	-0.08	0.00%	-0.08
	Cable Surplus	16.46	-0.09%	16.47	0.00%	16.47
	Satellite Surplus	7.30	-0.08%	7.31	0.00%	7.31
	RSN Surplus	0.44	5.41%	0.42	0.00%	0.42
	Consumer Welfare	29.84	-0.03%	29.85	0.00%	29.85
	Total Welfare	54.04	-0.01%	54.05	0.00%	54.05

Notes: This table presents observed and simulated market outcomes for individual RSNs in non-loophole markets. All reported figures except for market shares are in $/household/month, and all % changes are relative to specification (iii). Beneath the channel name is the name of the MVPD that owns the channel, the number of television households and the MVPD owner's footprint (% of households passed) in the RSN's relevant DMAs, and the % change in the integrated MVPD's households that obtain access to the channel upon moving from (iii) to (ii) in the RSN's relevant DMAs. "Aff Fees to Rivals" for the three satellite integrated RSNs are an average across cable MVPDs and Dish Network.

Table 14: Simulated Market Outcomes for Non-Integrated RSNs (1/2)

		(i) VI, no PARs		(ii) VI, PARs		(iii) No VI
		Level	% Change	Level	% Change	Level
NON-INTEGRATED RSNs (1/2)						
ALTITUDE SPORTS	Avg Cable Mkt Share	0.58	4.12%	0.58	3.72%	0.56
*Comcast	Avg Sat Mkt Share	0.26	-5.00%	0.27	-2.74%	0.27
Pop 7.12M	Avg Cable Prices	57.60	-1.62%	57.15	-2.38%	58.54
Footprint 74%	Aff Fees to Sat	-	-	1.98	4.46%	1.90
	Cable Surplus	17.00	3.85%	16.69	1.99%	16.37
Carriage +49%	Satellite Surplus	6.24	-2.71%	6.24	-2.68%	6.41
	RSN Surplus	0.48	-35.40%	0.57	-23.36%	0.74
	Consumer Welfare	28.16	0.99%	28.72	3.00%	27.88
	Total Welfare	51.87	0.92%	52.22	1.59%	51.40
FS DETROIT	Avg Cable Mkt Share	0.58	3.25%	0.58	3.46%	0.56
*Comcast	Avg Sat Mkt Share	0.16	-1.87%	0.16	-2.04%	0.17
Pop 4.84M	Avg Cable Prices	50.02	-1.81%	49.95	-1.95%	50.94
Footprint 82%	Aff Fees to Sat	2.83	75.78%	1.55	-3.25%	1.61
	Cable Surplus	19.99	0.49%	20.02	0.61%	19.89
Carriage +15%	Satellite Surplus	3.37	-6.49%	3.54	-1.80%	3.61
	RSN Surplus	1.97	8.52%	1.77	-2.32%	1.81
	Consumer Welfare	23.71	2.08%	23.74	2.23%	23.22
	Total Welfare	49.04	1.03%	49.07	1.10%	48.54
FS FLORIDA	Avg Cable Mkt Share	0.63	1.34%	0.63	1.52%	0.62
*Comcast	Avg Sat Mkt Share	0.23	-1.28%	0.23	-1.49%	0.23
Pop 6.20M	Avg Cable Prices	61.77	-0.71%	61.72	-0.81%	62.22
Footprint 67%	Aff Fees to Sat	1.46	87.94%	0.78	0.23%	0.77
	Cable Surplus	20.29	0.01%	20.30	0.02%	20.29
Carriage +0%	Satellite Surplus	5.94	-3.96%	6.09	-1.50%	6.19
	RSN Surplus	0.98	20.23%	0.81	-0.55%	0.81
	Consumer Welfare	27.73	0.90%	27.77	1.01%	27.49
	Total Welfare	54.95	0.31%	54.97	0.34%	54.78
FS MIDWEST	Avg Cable Mkt Share	0.61	0.66%	0.61	0.67%	0.60
*Comcast	Avg Sat Mkt Share	0.21	-0.46%	0.21	-0.47%	0.21
Pop 10.40M	Avg Cable Prices	54.46	-0.29%	54.46	-0.30%	54.62
Footprint 26%	Aff Fees to Sat	0.32	17.73%	0.27	0.04%	0.27
	Cable Surplus	16.83	-0.01%	16.83	-0.01%	16.83
Carriage +15%	Satellite Surplus	4.89	-0.55%	4.89	-0.43%	4.91
	RSN Surplus	0.11	7.62%	0.11	1.72%	0.11
	Consumer Welfare	27.19	0.33%	27.19	0.34%	27.10
	Total Welfare	49.02	0.14%	49.02	0.15%	48.95
FS NORTH	Avg Cable Mkt Share	0.61	0.13%	0.61	0.14%	0.61
*Charter	Avg Sat Mkt Share	0.15	-0.16%	0.15	-0.17%	0.15
Pop 5.77M	Avg Cable Prices	55.69	-0.08%	55.69	-0.09%	55.74
Footprint 12%	Aff Fees to Sat	1.28	21.18%	1.06	-0.06%	1.06
	Cable Surplus	22.65	0.00%	22.65	0.02%	22.65
Carriage +0%	Satellite Surplus	3.57	-0.86%	3.60	-0.16%	3.60
	RSN Surplus	1.62	1.59%	1.59	-0.19%	1.59
	Consumer Welfare	22.03	0.12%	22.03	0.14%	22.00
	Total Welfare	49.86	0.04%	49.86	0.05%	49.84
FS OHIO	Avg Cable Mkt Share	0.62	1.94%	0.62	2.10%	0.61
*TWC	Avg Sat Mkt Share	0.16	-1.36%	0.16	-1.52%	0.16
Pop 8.16M	Avg Cable Prices	53.44	-0.99%	53.39	-1.07%	53.97
Footprint 51%	Aff Fees to Sat	2.75	35.13%	2.01	-0.95%	2.03
	Cable Surplus	18.35	0.28%	18.36	0.37%	18.29
Carriage +2%	Satellite Surplus	3.49	-4.02%	3.58	-1.49%	3.64
	RSN Surplus	1.80	5.00%	1.69	-1.70%	1.72
	Consumer Welfare	24.13	1.30%	24.15	1.41%	23.82
	Total Welfare	47.77	0.63%	47.79	0.68%	47.46
FS SOUTH	Avg Cable Mkt Share	0.60	1.13%	0.60	1.17%	0.60
*TWC	Avg Sat Mkt Share	0.22	-0.74%	0.22	-0.78%	0.22
Pop 13.20M	Avg Cable Prices	58.17	-0.50%	58.16	-0.52%	58.46
Footprint 33%	Aff Fees to Sat	1.49	15.76%	1.28	-0.24%	1.29
	Cable Surplus	21.09	0.00%	21.10	0.01%	21.10
Carriage +2%	Satellite Surplus	5.06	-1.70%	5.11	-0.75%	5.15
	RSN Surplus	1.02	5.84%	0.97	0.36%	0.97
	Consumer Welfare	26.85	0.59%	26.86	0.61%	26.70
	Total Welfare	54.03	0.23%	54.03	0.24%	53.91

Notes: This table presents observed and simulated market outcomes for individual non-integrated RSNs. All reported figures except for market shares are in \$/household/month, and all % changes are relative to specification (iii). Beneath the channel name is the name of the MVPD that is assigned ownership of the channel, the number of television households and the MVPD owner's footprint (% of households passed) in the RSN's relevant DMAs, and the % change in the integrated MVPD's households that obtain access to the channel upon moving from (iii) to (ii) in the RSN's relevant DMAs.

Table 15: Simulated Market Outcomes for Non-Integrated RSNs (2/2)

		(i) VI, no PARs		(ii) VI, PARs		(iii) No VI
		Level	% Change	Level	% Change	Level
NON-INTEGRATED RSNs (2/2)						
FS SOUTHWEST	Avg Cable Mkt Share	0.57	1.50%	0.57	1.53%	0.56
*Cox	Avg Sat Mkt Share	0.22	-0.64%	0.22	-0.67%	0.23
Pop 12.70M	Avg Cable Prices	53.88	-0.73%	53.88	-0.75%	54.28
Footprint 37%	Aff Fees to Sat	1.38	9.51%	1.27	0.48%	1.26
	Cable Surplus	15.52	-0.02%	15.52	-0.01%	15.52
Carriage +0%	Satellite Surplus	5.28	-1.11%	5.31	-0.62%	5.34
	RSN Surplus	1.00	2.66%	0.97	-0.25%	0.97
	Consumer Welfare	28.32	0.68%	28.33	0.70%	28.13
	Total Welfare	50.12	0.31%	50.12	0.32%	49.96
FS WEST	Avg Cable Mkt Share	0.56	1.40%	0.56	1.53%	0.55
*TWC	Avg Sat Mkt Share	0.24	-0.79%	0.24	-0.90%	0.24
Pop 8.43M	Avg Cable Prices	55.58	-0.71%	55.55	-0.78%	55.98
Footprint 53%	Aff Fees to Sat	1.39	57.00%	0.88	-0.78%	0.89
	Cable Surplus	22.55	0.06%	22.56	0.09%	22.54
Carriage +0%	Satellite Surplus	4.93	-3.25%	5.05	-0.86%	5.10
	RSN Surplus	1.34	9.76%	1.20	-1.25%	1.22
	Consumer Welfare	26.56	0.64%	26.58	0.70%	26.39
	Total Welfare	55.39	0.24%	55.40	0.26%	55.25
MASN	Avg Cable Mkt Share	0.67	1.57%	0.67	1.73%	0.66
*Comcast	Avg Sat Mkt Share	0.16	-1.78%	0.16	-2.09%	0.17
Pop 8.25M	Avg Cable Prices	58.23	-0.26%	58.18	-0.34%	58.38
Footprint 52%	Aff Fees to Sat	1.48	93.22%	0.75	-2.46%	0.77
	Cable Surplus	23.74	0.42%	23.74	0.43%	23.64
Carriage +69%	Satellite Surplus	3.56	-4.35%	3.65	-1.85%	3.72
	RSN Surplus	0.56	92.00%	0.45	54.94%	0.29
	Consumer Welfare	25.98	1.44%	26.01	1.55%	25.61
	Total Welfare	53.84	1.07%	53.85	1.11%	53.26
NESN	Avg Cable Mkt Share	0.68	9.09%	0.68	8.83%	0.63
*Comcast	Avg Sat Mkt Share	0.10	-20.16%	0.11	-7.83%	0.12
Pop 5.20M	Avg Cable Prices	58.37	-1.63%	55.53	-6.42%	59.34
Footprint 85%	Aff Fees to Sat	-	-	1.92	-11.34%	2.16
	Cable Surplus	29.97	5.27%	28.85	1.35%	28.46
Carriage +0%	Satellite Surplus	2.16	-11.20%	2.27	-6.35%	2.43
	RSN Surplus	3.51	-13.39%	3.59	-11.55%	4.05
	Consumer Welfare	24.95	0.03%	26.93	7.98%	24.94
	Total Welfare	60.58	1.16%	61.64	2.93%	59.89
PRIME TICKET	Avg Cable Mkt Share	0.56	0.87%	0.56	0.95%	0.55
*TWC	Avg Sat Mkt Share	0.24	-0.49%	0.24	-0.56%	0.24
Pop 8.32M	Avg Cable Prices	55.73	-0.44%	55.71	-0.49%	55.98
Footprint 53%	Aff Fees to Sat	0.89	62.14%	0.55	0.28%	0.55
	Cable Surplus	22.54	0.03%	22.55	0.04%	22.54
Carriage +0%	Satellite Surplus	4.99	-2.09%	5.07	-0.56%	5.10
	RSN Surplus	0.77	11.36%	0.68	-0.90%	0.69
	Consumer Welfare	26.50	0.40%	26.51	0.44%	26.39
	Total Welfare	54.80	0.15%	54.81	0.17%	54.72
SUN SPORTS	Avg Cable Mkt Share	0.65	-0.19%	0.65	-0.19%	0.65
*TWC	Avg Sat Mkt Share	0.16	0.55%	0.16	0.54%	0.16
Pop 3.41M	Avg Cable Prices	64.77	-0.05%	64.77	-0.05%	64.80
Footprint 65%	Aff Fees to Sat	1.75	0.30%	1.75	0.16%	1.75
	Cable Surplus	17.67	-0.08%	17.67	-0.08%	17.69
Carriage +0%	Satellite Surplus	4.40	0.52%	4.40	0.52%	4.37
	RSN Surplus	0.82	2.51%	0.82	2.45%	0.80
	Consumer Welfare	19.97	0.08%	19.97	0.08%	19.95
	Total Welfare	42.86	0.10%	42.86	0.10%	42.82
YES[a]	Avg Cable Mkt Share	0.68	1.24%	0.68	1.27%	0.68
*TWC	Avg Sat Mkt Share	0.15	-1.39%	0.15	-1.42%	0.15
Pop 11.40M	Avg Cable Prices	59.48	-0.72%	59.47	-0.73%	59.91
Footprint 35%	Aff Fees to Sat	0.93	25.88%	0.75	1.18%	0.74
	Cable Surplus	27.75	0.04%	27.75	0.05%	27.74
Carriage +0%	Satellite Surplus	2.90	-2.52%	2.93	-1.46%	2.97
	RSN Surplus	1.69	1.79%	1.65	-0.34%	1.66
	Consumer Welfare	26.92	0.85%	26.93	0.86%	26.70
	Total Welfare	59.26	0.33%	59.27	0.33%	59.07

Notes: This table presents observed and simulated market outcomes for individual non-integrated RSNs. All reported figures except for market shares are in \$/household/month, and all % changes are relative to specification (iii). Beneath the channel name is the name of the MVPD that is assigned ownership of the channel, the number of television households and the MVPD owner's footprint (% of households passed) in the RSN's relevant DMAs, and the % change in the integrated MVPD's households that obtain access to the channel upon moving from (iii) to (ii) in the RSN's relevant DMAs.

[a] YES does not supply DISH Network in 2007; we assume that it does not supply Dish in any specification.