An Open Access Wireless Market

Supporting Competition, Public Safety, and Universal Service

Peter Cramton and Linda Doyle[1]

3 April 2016

Abstract

We present a market design for an open access wireless market. Open access means that in real time, network capacity cannot be withheld—network throughput is priced dynamically by the marginal demand during congestion. In unconstrained times and locations, a nominal fee is paid for network throughput. As in electricity markets, the real-time market provides the foundation for forward markets. Monthly forwards are auctioned before the start of each month; yearly forwards are auctioned before the start of each year. Market participants, both operators and traders, take positions in forward auctions to manage risk and optimize portfolios. Deviations from forward positions are settled at real-time prices based on actual use. The independent system operator runs the network and conducts the real-time, monthly, and yearly auctions of network throughput. An independent market monitor observes the market, identifies problems, and suggests solutions. A board—including affiliated directors representing important stakeholders together with independent directors with subject matter expertise—governs the market. A goal of the market is to provide a secure, robust, wide-coverage platform for mobile communications supporting public safety and universal service. Public safety has pre-emptive rights during emergencies and otherwise has economic use like wholesale operators. A complementary goal is competition. The open access provision brings vibrant competition through low-cost, non-discriminatory entry into the wireless market. The market provides a natural remedy for mergers, allowing operational efficiency gains while increasing competition. Critical funding is provided through efficient congestion pricing that balances supply and demand at every time and location. The market, enabled by flexible handsets and the LTE technology, radically reforms current spectrum policy. The market coexists and complements the dedicated networks of incumbent carriers, promoting efficient spectrum use and essential innovation.

[1] Peter Cramton is Professor of Economics at the University of Maryland, European University Institute, and University of Cologne; since 1983, he has conducted widely-cited research on market design; he has applied that research to design auction-based markets of radio spectrum, electricity, financial securities, and other products. Linda Doyle is Professor of Engineering and the Arts and Director of CONNECT at Trinity College Dublin; since 1996, she has conducted research on wireless networking, cognitive radio, reconfigurable networks, dynamic spectrum access, spectrum trading, and spectrum regulation. We have benefited from helpful comments from Ofcom staff, especially Geoffrey Myers. Thanks also to Brian Carney, Ken Fields, Rich Fusco, Ulrich Gall, Declan Ganley, Irene Macaluso, John Meyer, Fran O'Flaherty, and Clint Smith for valuable discussions. We are grateful to Rivada Networks for supporting this research. The views expressed are our own.

Contents

Summary 3
Introduction 7
The time is right 10
  Flexibility as an enabler 10
  Emergence of new traffic patterns 11
  The requirement to share resources 11
Objectives and market principles 12
  Efficiency 12
  Simplicity 13
  Transparency 13
  Fairness 13
  Open access as a key market principle 13
Overview of market design 14
Product design 16
  The demand side 17
  The supply side 18
  A worked example 20
Three opportunities to trade 22
  Price floor 27
  Real-time balancing of supply and demand 28
Secondary market for forward products 29
Coexistence of open access and existing markets 29
Promising opportunities for open access markets 29
  Mexico: enhancing competition 29
  European Union: merger remedy 30
  United States: public safety and universal service 32
  Singapore: new entrant set aside 32
Independent System Operator 33
Governance 34
Development of the market over time 34
Spectrum auctions and competition policy 35
Conclusion 35
References 36

Summary

Today's mobile communications are built around a straightforward regulatory model: the government finds suitable radio spectrum, organizes the spectrum into licenses defined by frequencies and geographic coverage, and then auctions the licenses to wireless carriers. This "spectrum auction model" provides outstanding investment incentives for the carriers, since the auction enables carriers to have guaranteed use of the spectrum on a long-term basis (in the US, licenses are effectively perpetual due to a strong expectancy of renewal without fee). The model has a history of success beginning with the first spectrum auction in 1994. But the model also has serious shortcomings. The main challenge is competition in the market for wireless services. Typically, there are two or three carriers that dominate the wireless market due to the enormous economies of scale in network infrastructure. A robust wireless network covering much of the US requires an investment of tens of billions of dollars for spectrum licenses and tens of billions more for physical network infrastructure. As a result, the spectrum auction model inevitably creates an oligopoly where the regulator has a constant fight to maintain competition and promote innovation.

We present a market design for an open access wireless market that addresses this challenge. The open access market would substantially enhance competition and efficiency in mobile broadband. With open access, network throughput cannot be withheld. This has profound implications for pricing—spot market pricing reflects congestion. Only a nominal fee is paid for throughput at times and locations without congestion; however, during congestion throughput is priced at the value of the marginal demand. Many modern wholesale electricity markets, such as those in the US, operate on this open access principle and price energy at every time and location in this way. This is called *locational marginal pricing* in the real-time market. It works extremely well. The high level of price transparency not only leads to efficient short-run decisions, but provides a wealth of market information for longer-term planning including future network investments.

Since real-time congestion prices tend to be volatile, market participants have a desire to manage risks. Forward auctions, conducted in advance of real time, allow market participants to make plans and lock in prices consistent with their needs. The market also allows traders to arbitrage across related products—such as yearly, monthly, and hourly products in the same area—to improve price signals and resource allocation. As in modern electricity markets, we can expect the vast majority of network capacity to trade in advance of the real-time market. The forward markets enable planning and hedging of risks, while the real-time market sends just the right price signal to efficiently manage congestion.

The portfolio of products and markets is established to support the key market objectives: efficiency, transparency, simplicity, and fairness. In the case of a wireless network, yearly and monthly forward products likely are sufficient to lock in prices and quantities in advance of the real-time (hourly) market. Deviations from forward positions are settled at the real-time price.

The products are standard as is the method of settlement. The yearly product is an aggregation of monthly products and the monthly product is an aggregation of real time products. The aggregation is both in time (yearly is a twelve-month strip; monthly is a strip of all hours in the month) and space. For example, Manhattan might be a yearly product, whereas the monthly product might divide Manhattan into upper

and lower. The real-time product would be at a more granular level both in time and space—gigabytes of data throughput in a small geographic area in a particular hour. This disaggregation in real time is desirable, since congestion occurs at this fine division of time and space.

The auctions are also standard and best-practice, drawing on decades of experience in similar markets. All auctions are single-price auctions. Each buyer expresses a downward-sloping demand curve. Each seller expresses an upward-sloping supply curve. The system operator then forms the aggregate supply and demand curves and determines the market clearing price, the point where supply and demand balance. Network throughput is assigned to buyers at the clearing price for the quantity that was bid at that price and sellers are paid at the clearing price. The real-time market is sealed-bid—the system operator merely finds the point where supply and demand cross at the designated trading time. The real-time market is special in that it is a physical market—supply and demand are represented by physical resources. The yearly and monthly markets are important for longer term planning. For this reason, these forward markets are conducted as a simultaneous ascending clock auction. The forward markets remain a single-price auction, but the demand bids are collected over a number of rounds. This allows the demand side to benefit from a simple price and assignment discovery process. The outcome discovery improves buyer decision-making and therefore the efficiency of the auction.

The open access market model is not new. Wholesale electricity markets have operated on these principles for more than a decade with great success. A key difference is that the wireless communications setting is much *simpler*. Electricity markets have large "lumpy" resources that are expensive to turn on and limited in the speed with which they can make adjustments. In contrast, wireless network elements are fast to respond and are efficiently controlled with marginal prices. The success of wholesale electricity markets in a more complex economic setting then goes a long way toward proof of concept.

An open access network can coexist and indeed is complementary with the traditional oligopoly structures in which a few mobile carriers provide distinct networks with exclusive spectrum. An incumbent carrier can take advantage of the open access network to buy throughput when the carrier's network is congested or otherwise unavailable, for example customers may be located outside of the carrier's coverage but within the coverage of the open access network. Similarly, a carrier with excess capacity can sell the surplus to buyers in the open access network. The incumbent carrier, like other market participants, can tailor any sale or purchase to the carrier's needs. And the open access network easily accommodates the inputs of non-carriers. For example, a company with a spectrum license may offer some or all of that spectrum to the market. This allows companies to make pure-spectrum plays and receive competitive returns on that spectrum asset without making any infrastructure investment. Similarly, a federal agency with spectrum rights can offer spectrum to the market, consistent either with priority access or a feasible and incremental schedule of vacating the spectrum. All that is required for the spectrum sharing is that the handsets be capable of utilizing the particular band and that the network be enabled with LTE compliant software that allows the ISO to effect delivery.

The open access market is suitable in all countries; however, two countries stand out as likely first-movers: Mexico and the United States. Mexico has set aside 90 MHz of low-band spectrum for this purpose. The chief motivation for Mexico is addressing a serious competition problem in mobile communications. In the United States, public safety provides an avenue for open access. The US has set aside 20 MHz of highly

desirable, low-band spectrum for public safety use. The spectrum has sat largely unused for eight years. In addition to the spectrum, Congress has allocated $7 billion for network build. The set-aside spectrum and funds would go a long way to initiate an open access network. Major network builders and tower vendors with dominant inventory, stand ready to execute the FirstNet build. Sharing agreements with existing carriers may further speed the initial deployment of a network with unrivaled coverage. The US also has substantial universal service funds that could be allocated—like the public safety funds—to provide network coverage where coverage is currently not economic. Mobile broadband has become an essential service and universal service policy should reflect this fundamental change in basic needs.

To be clear, neither public safety nor universal service are needed for the open access market. However, both public safety and universal service are highly complementary with the market as both create value for the network and provide funds or spectrum for its support. Some countries will adopt the open access model in conjunction with satisfying both public safety and universal service goals; others will link the open access network with either public safety or universal service; and still other countries will create an open access network without links to either public safety or universal service.

Merger is another avenue for the introduction of an open access market. In recent years mobile communications has experienced consolidation. Merger brings network operating efficiencies, but a reduction of competition. Regulators typically oppose mergers that would significantly reduce competition unless the merged entity proposes an acceptable remedy that would restore competition. As a merger remedy, the open access market has many advantages relative to the capacity MVNO remedy adopted in some recent four-to-three mergers in the European Union. In particular, the open access market leaves to competitive forces the number and identity of the MVNOs, as well as the wholesale prices paid for network capacity. Moreover, these outcomes adjust flexibly in response to market fundamentals.

The proposed market design draws on best-practice across many industries and countries. This is done not only to optimize the design with respect to objectives, but also to avoid past mistakes. Much has been learned over the last twenty years in designing markets like communications and electricity, where products are defined by time and space. In addition, to the technical questions of product and auction design, the market design must establish rules for operating, monitoring, and governing the market.

The independent system operator (ISO) runs the network, manages physical delivery, and conducts the real-time, monthly, and yearly auctions. In addition, the ISO plans and executes network development and performs network upgrades. Independent means that the ISO has no ownership interest in the market participants and does not take a position in the market. The ISO is charged by its board to operate the market to maximize the market's core objectives of providing a secure, robust, wide-coverage platform for mobile communications which supports public safety and universal service. Open access and efficient trade are two key principles that govern the ISO's decision making. For example, it is the efficiency objective that determines the ISO's timing of the primary sale of network throughput across the three markets: yearly, monthly, and real time.

An independent market monitor observes the market, identifies problems, and suggests solutions. Independent in this case means that the market monitor, in addition to being independent of the market

participants is also independent of the ISO. The market monitor brings expert market knowledge. Importantly, the market monitor is not a judge; the market monitor cannot enforce market rules and inflict penalties. Rather the market monitor is an observer who writes reports and makes recommendations. In electricity markets, this has allowed the market monitor to quickly identify and address problems. The same would be true here. The market monitor reports to the ISO board.

The board oversees the ISO. To ensure that the board includes the knowledge and views of a diversity of market participants, the board has affiliated directors who are affiliated with a particular stakeholder group. The board also has a number of independent directors who are independent of the stakeholder groups, but bring essential subject matter expertise. Unaffiliated directors are approved by the regulator; affiliated directors are approved by the stakeholder group. This governance structure has the advantage of a more streamlined stakeholder process in response to key decisions of the ISO, since these decisions have already been reviewed by a board that spans important stakeholders' interests.

A main reason that open access wireless markets do not yet exist is that until recently, both handsets and the underlying network technology were not sufficiently flexible to accommodate an open access market. That is no longer true. Today's handsets powered with LTE are now fully capable of operating efficiently in an open access market—indeed open access expands the capabilities of handsets and promotes efficient use. Network technology has similarly advanced. Taking advantage of these advances is especially important in a time where the demand for data throughput is growing exponentially and the share of mobile data traffic is increasing rapidly (Cisco 2015). An open access market provides an Internet-like ecosystem upon which unrestrained innovation can occur to maximize the value society gets from scarce network resources.

Introduction

A quiet revolution is taking place with the expansion of Internet technologies from desktops to tablets to phones to virtually any device whose features can be enhanced with connection to the Internet. This Internet of Things puts mobile communications at the center. Pervasive connectivity is only possible with massive innovation in mobile communications and the best hope of accomplishing this is to endow mobile communications with the same Internet ecosystem that has led to the rapid development of Internet technologies and services that we all enjoy.

The chief weakness of the current mobile communications markets is the oligopoly structure in which a small number, typically between one and four, of mobile network operators build separate networks and then operate them in exclusive spectrum purchased in infrequent spectrum auctions. This is an extreme version of the "walled garden" approach to competition, as users face enormous switching costs to move from one garden to the next, and there are few gardens to choose from.

This paper outlines the design of an alternative approach that still allows the walled gardens to exist and indeed flourish, yet decimates the high entry barriers that prevent Internet-style competition within the current oligopoly structure. The trick is to create an open access network that allows anyone with a good idea to gain access to mobile communications at competitive rates. Indeed, the cornerstone to open access is that use of the network cannot be withheld. Just like the Internet, anyone can use it on a nondiscriminatory basis. Of course, the capacity of the network is scarce, and so prices are needed to assign network resources to users. An open access network adopts efficient pricing. Supply is not withheld. Price is set at the value of the marginal demand.

There are many possibilities for the creation of an open access network. Perhaps the most immediate and direct is the approach in Mexico, where 90 MHz of 700 MHz spectrum has been set aside to create an open access network. This low-band spectrum is especially well-suited to rapidly and economically build a network with excellent coverage. Such a network would provide vibrant competition in a market that today suffers from a lack of competition. Service providers, both new entrants and existing smaller incumbents, could make use of the open access network to offer excellent coverage and service quality at competitive rates. The network would also provide a means for existing incumbents to monetize network resources, both spectrum and physical network infrastructure. The dominant incumbent would, at least initially, be hurt by the new competition. But over time even the dominant incumbent may benefit from a growing and innovative market in mobile communications. Economic welfare—both for producers and consumers—will soar.

The creation of an open access market in the United States is following a different pattern. In large part because the competition problem is less severe in the US, the key motivation for the open access network is public safety. The US has set aside 20 MHz of 700 MHz spectrum for public safety as well as initial funds to build a network that has the coverage and reliability demanded by public safety. It is desirable that the network be structured as an open access network in which public safety has preemptive rights during an emergency. Doing so provides funds to build, operate, and maintain a much more robust network. Traditional public safety networks have long suffered from poor funding. Open access allows a state-of-

the-art network with a steady source of funding for frequent upgrades. It also makes efficient use of the network at times and locations absent a public safety emergency.

In both the US and Mexican markets, there is a third motivation for the open access network: universal service. Since the beginning of telecommunications, it has been understood that there are positive externalities from expanding communications to as many as possible. Communication is more valuable to each individual if that individual can reach more people. The same is true today. Indeed, mobile communications has now become an essential service that enables access to a host of complementary services in the flourishing Internet ecosystem. As a result nearly all countries have public policies geared at fostering universal service. Unfortunately, these well-meaning regulations are poorly conceived for the modern era of mobile communications and are in desperate need of reform. One basic problem is universal service has focused on landline voice communications—a sector of communications that is largely dead.

The good news is that the regulatory shortcomings of universal service are readily overcome. What is needed is to replace ill-conceived subsidies for voice communications with market-based subsidies to bring quality communication to areas that are uneconomic—where the cost of the network build cannot be fully compensated from user fees. An open access network nicely complements efforts toward universal service based on least-cost subsidies to provide coverage in uneconomic areas. An open access network lets network operators compete to be the party that provides coverage. Universal service amounts to identifying the network operator via competitive procurement who will provide quality coverage at least cost in the uneconomic area. The open access network then provides the means to allocate the network resource among competing services providers. In sharp contrast to current policies, this competitive-procurement approach is both simple and effective. The approach utilizes two levels of competition: competition among network operators to provide the network resources in uneconomic areas and then competition among service providers to sell the procured communications resource to users.

Another motivation for the open access network is that it greatly facilitates the economic *repurposing* of previously allocated spectrum to mobile communications, as well as the *sharing* of spectrum in situations where demands vary greatly by time and location. The reality is that spectrum best-suited for mobile communications has been allocated already. The only way to get more is either repurposing or sharing existing spectrum. The open access network facilitates this challenging task in two important ways. The first is information. By revealing the marginal value of demand at each time and location, an open access network provides powerful price information that informs both network build and spectrum valuation. Existing users of spectrum, both government and commercial, see the opportunity cost of failing to share spectrum with the open access network. Second, the open access network provides an easy and direct means to enable both repurposing and sharing. Once a willingness to share or repurpose is expressed, equipment manufacturers can add the relevant band to those supported by ever more flexible devices. On the network side, the additional bands are readily supported. Software defined radios mean that network support can involve little more than a software upgrade and retuning of antennas.

The enabling possibilities of an open access network for sharing and repurposing of spectrum are great. A large quantity of spectrum is set aside for government use. Much of this spectrum could be economically

shared or repurposed. For example, military spectrum could follow the public safety model, which would allow preemption by the military as needed, but in the vast majority of times and locations the spectrum could be made available for commercial use. To facilitate this economic transition of spectrum from government use, the United Kingdom adopted a policy that requires agencies to pay for spectrum use, where the administrative price is intended to approximate the opportunity cost of the agency use. An open access network provides a way to precisely calculate the opportunity cost and then a means for the agency to share or repurpose the spectrum if the opportunity cost exceeds the agency's value.

It is interesting to contrast the current spectrum markets with the electricity market design discussed earlier. In today's spectrum markets, there is essentially only a long-term market. Opportunities to balance supply and demand in the medium- and short-term are largely lost. This is undesirable from an efficiency perspective, but it also creates competition problems in the market for wireless services. In contrast, today's electricity markets begin with a foundation of open access (in the US FERC order 888) and a real-time market that efficiently prices energy at every time and location, and end with well-developed forward markets that enable participants to make plans and manage risk.

Short-term markets for spectrum have only recently been made possible as a result of technological innovation at both the handset and network level. Regulators should take advantage of these new capabilities to promote short-term spectrum markets. These technology advances are not something that we must wait decades for. The technology is here today and simply has to be harnessed. Regulatory leadership likely is required, because the development of these short-term markets is pro-competitive and incumbents generally dislike competition, especially dominant incumbents. There are two fundamental forces in economics—competition and monopoly—that pull in opposite directions. Regulators must constantly look for ways to encourage competition as that is the force that improves economic welfare.

What would the wireless market of the future look like? Much like the electricity markets of today, there would be multiple opportunities to contract. We would have the long-term investment market—the existing spectrum auctions. There would be medium-term markets in which service providers buy and sell network capacity with a one-month to one-year duration. And there would be a spot market in which mobile operators buy and sell capacity on a real-time basis, such as hourly. To some this may sound like science fiction, but it is not. The technology is here today. It simply is a matter of overcoming the inertia of the status quo.

This paper builds on ideas that have been well-understood for decades. Noam (1998) articulates the limitations of spectrum auctions and how they will lead to oligopoly and weak competition. He then goes on to explain the inevitable evolution to an open access market in which spectrum is accessed in spot and futures markets based on congestion. Noam's vision, expressed at the onset of spectrum auctions in the mid-1990s, is now here. One important difference with our approach is that in Noam's spectrum access market (see also Berry et al. 2010), spectrum is unbundled from network infrastructure; whereas, our open access market bundles spectrum and other network resources into a product that consumers value, mobile communications. This is the approach that has proven successful in other markets such as electricity, where the primary product is energy (MWh), which is directly consumed by users, rather than unbundling fuel and the other resources that combine to produce the commodity valued by users.

MacKie-Mason and Varian (1994, 1995) explain dynamic pricing of Internet access. Their paper anticipates the need for congestion pricing to efficiently assign scarce network resources. They describe how congestion pricing works to maximize the economic value of the network. Our paper builds on their insights to develop a practical market design for mobile communications. The market design also draws on our extensive experience in designing related markets over the last twenty years (see, for example, Allaz and Vila 1993; Ausubel and Cramton 2004; Ausubel et al. 2015; Budish et al. 2015; Cramton 2013, 2015; Klemperer 2004; Milgrom 2004; and Stoft 2002).

Our paper is structured as follows. We begin with a more detailed discussion on why now is the time to implement an open access market. Then we discuss the objectives that should guide the market design. Then we address the essential issue of product design. Next we present a straw man design that we believe will work well both initially and as it evolves with experience and the advancement of technology. We then discuss how the open access market complements existing markets. We then describe promising applications of open markets in major countries, focusing on Mexico and the United States. Implementation of the market requires a system operator. We describe the role of the independent system operator and issues of governance. We end with some thoughts on how the market is apt to evolve over time.

The time is right

The open access market is based on the efficient allocation of network capacity. Network capacity goes to those willing to pay the most for it, and the price paid for capacity is set to balance supply and demand. When capacity is scarce, a high price is paid to efficiently allocate the available supply; when capacity is abundant (uncongested) there is no need to ration demand and the price is set to a nominal floor. This kind of market requires the technological and communications ecosystem to support a level of flexibility that did not exist with early-generation mobile communications. Technological, business and policy advances make the time right for the market we propose.

Flexibility as an enabler

Early handsets contained only one radio frontend. Now they contain a large number of frontends that target different technologies and different frequency bands. Hence, not only are handsets designed to work across a wide range of technologies, such as WiFi, 2G, 3G and 4G, they can function in different frequency bands associated with each of these technologies. This is especially important in the case of 4G, for which there are 44 different bands. Typically manufacturers design radios for bands as those bands come into popular usage, and these radios get included in next generation handset products. The implication of flexible handsets is that users can easily migrate from one network to another. In essence there is no reason why users cannot take advantage of additional coverage and capacity on other networks beyond their home network.

This flexible handset is an important enabling factor for the open access market. It means that devices can make use of all capacity offered on an open access network. The flexible handset is already a reality. However, there are other critical technological advances that extend flexibility beyond the handset and enable physical delivery, without which, open access markets would not be viable.

Increasingly, the entire network is becoming more software defined and hence more flexible. This is evident at the basestation and right into the core network. For example, Network Function Virtualization involves implementing network functions in software that can run on a range of industry standard server hardware, and that can be moved to, or instantiated in, various locations in the network as required, without the need to install new equipment.[2] The type of agile network that is needed to support an open access market, i.e. one that requires that capacity gets delivered to those willing to pay the most for it, is one that requires not just flexible handsets but can, depending on implementation, call for changes at the basestation and in the core network. It is increasingly easy, because of this focus on software defined architecture, to insert the hooks needed within the network so that it can divert resources to those who value them most.

Emergence of new traffic patterns

The growth in machine-to-machine and Internet-of-Things traffic is another factor that points to the time being right for action. There is an ever-increasing number of machine-to-machine and Internet of Things applications emerging.[3] All of these require connectivity. That connectivity is delivered both by dedicated networks as well as cellular networks. It is the latter that is of interest in this paper. Ericsson notes that there were around 230 million cellular machine-to-machine subscriptions at the end of 2014[4] and point to many avenues that will encourage that growth further such as the reduction in cost of LTE modems, and new developments and 5G capabilities which will extend the range of addressable applications for massive machine-to-machine communications.

Of relevance for this paper is the type of traffic patterns that are associated with machine-to-machine and Internet of Things applications. A portion of that traffic will undoubtedly be real-time traffic that needs to be served instantly. But a significant proportion is non-real-time traffic, is highly predictable and can be managed in time. Hence carriers can support applications that the can be well served at times of the carriers' choosing. Hence a machine-to-machine service provider can seek capacity for these kinds of applications when capacity is abundant (uncongested), there is no need to ration demand and the price is nominal.

The requirement to share resources

The increased focus from a technical, business, policy and political perspective on resource sharing, is also a pointer for the time being right. This focus on sharing manifests itself in different ways.

Spectrum is perceived as a scarce resource and is becoming scarcer as bandwidth-hungry applications increase. In many jurisdictions around the world, initiatives focusing on spectrum sharing are of huge interest because they are seen as a means of addressing the scarcity. For example, in Europe there is much ongoing work on License Shared Access. License Shared Access is a concept that allows spectrum that has been licensed for international mobile telecommunications to be used by more than one entity. In the US, a report from the President's Council of Advisors on Science and Technology led to the development of

[2] www.etsi.org/images/files/ETSITechnologyLeaflets/NetworkFunctionsVirtualization.pdf
[3] 2 June 2015 - *Internet of Things Market to Hit* $1.7 Trillion by *2020*: *IDC*.
[4] www.ericsson.com/res/docs/2015/ericsson-mobility-report-june-2015.pdf

Federal Communications Commission rules to support sharing of spectrum with federal incumbents. The rules support an approach based on a three-tiered structure consisting of incumbents, of secondary users who have some sort of priority (exclusive sharing) and of other users who share in a non-exclusive manner (PCAST 2012). Though the approaches differ across the globe, the message remains the same: Resources increasingly need to be shared to deliver the types of services the world needs.

In the above examples, spectrum is unbundled from network infrastructure. Sharing of spectrum in this manner has yet to gain significant practical traction. As mentioned in the introduction, we focus on how spectrum and other network resources are combined into a product that consumers value, namely mobile communications. Examples of bundled approaches, albeit from a more limited perspective, do already exist. For example, new coalitions of players are combining bundled resources to offer services to customers through such initiatives as Project Fi. Project Fi is a service that allows users to avail of two different 4G networks (T-Mobile and Sprint) and WiFi. The user gets the best service by automatically connecting to the network with the strongest signal in a given location. The user is no longer committed to just one carrier. This user only cares that the service is provided rather than the identity of the carrier. The two carriers effectively have pooled their resources for the benefit of the user.

Whatever the approach, there is a growing expectation that resources need to be pooled or shared so that they can more effectively and efficiently be used to serve the growing user demands the world is experiencing and the needs of new competing service providers. The open access market proposed in this paper is timely in that it meets the expectations to drive more efficient resource usage through exploiting shared resources. It goes beyond spectrum sharing by recognizing that spectrum is linked to the infrastructure through which it is accessed. It can be seen as a natural evolution and more dynamic version of the kinds of coalitions we see emerging through initiatives such as Project Fi.

Objectives and market principles

In any market design application, it is best to start with the objectives of the market. The market designer evaluates alternative market designs with respect to these objectives, choosing the rules of market interaction to best meet the objectives given the details of the setting. Common objectives in regulated markets are efficiency, transparency, simplicity, and fairness. We describe each.

Efficiency

Efficiency is the most basic objective for economists. A market design is efficient if it yields outcomes that maximize social welfare. In a trading environment, this means that all gains from trade are realized. In the case of mobile communications, efficiency can be broken down into two components:

- Short-run efficiency—existing network resources are put to their best use in satisfying demand.
- Long-run efficiency—long-run network investments are made to satisfy demand at least cost.

In mobile communications, the focus has been on the long-run assignment of spectrum rights in major spectrum auctions conducted once or twice every ten years. Here efficiency is about assigning spectrum rights to the carriers best able to use it. Spectrum auctions are readily designed to award spectrum to those who value it most highly. However, there often are issues of competition in the downstream

communications market that create a divergence between highest private value (as bid in the auction) and highest social value (reflecting the benefits of improved competition in downstream markets). Competition policy typically is needed to avoid excessive concentration of spectrum among dominant incumbents. This competition policy is often crude, complex and controversial.

Moreover, even if well designed spectrum auctions and effective competition policy succeed in assigning spectrum rights to those who will best use it, the spectrum auctions essentially ignore the fluctuations in carrier demand for spectrum and other network resources that occur on a yearly, monthly, and real-time basis. Efficiently addressing these fluctuations in demand and supply over time and at every location is where the open access market can create value and indeed transform the market. Both long-run and short-run efficiency become achievable goals.

Simplicity

The auction should be as simple as possible, but not simpler. Mobile communications is a complex setting; hence, it should not be surprising that some level of complexity is needed in an efficient design. Nonetheless, it is important that the market be made as simple as possible to solve the economic problem of the setting.

Simplicity is best measured in terms of the simplicity of participating in the market. Are the needs of market participants satisfied as simply as possible? Simpler designs let participants express preferences more simply and effectively. Simpler designs have straightforward incentives. Simpler designs also reduce risks of participants.

Transparency

A first requirement of transparent markets is clear and unambiguous rules that map bids into outcomes. With a transparent design, participants know why they won or lost and understand why their payments are what they are. Participants are able—at least after the event—to confirm that the market rules were followed.

Fairness

Equal opportunity is a basic requirement of fairness. All potential participants have access to the market rules and the rules do not inappropriately discriminate among parties.

To a large extent, these four objectives are complementary. The market designer can choose a design that gets high marks with respect to each objective. That will certainly be true in the market design that we propose here. For concreteness, we will make specific recommendations in situations where there are other choices that would work well. By being specific, we establish a benchmark design that then is readily compared with alternatives. Some adjustments to the benchmark design will be required for specific settings.

Open access as a key market principle

One element of the market design that is essential is open access, the idea that the network is open to all on nondiscriminatory terms, and that network capacity cannot be withheld. Open access is the foundation

of today's restructured electricity markets. It is the key force that has led to competitive wholesale electricity markets that have provided reliable electricity supply while saving consumers many tens of billions of dollars from an efficient and competitive market for electricity (O'Connor and O'Connell-Diaz 2015).

Open access does more than knock down the nearly insurmountable entry barriers in mobile communications. It provides an ecosystem of transparent, efficient, and understandable pricing. Pricing is no longer arbitrary, but rather consistent with the marginal value of demand at the time and location in question. This allows service providers to make plans, manage risks, and establish innovative service agreements for users. It also allows for the efficient use of the scarce resource. Open access is not only complementary with, but indeed requires, high levels of efficiency, simplicity, transparency, and fairness. Figure 1 illustrates the evolution of the market model from monopoly, to oligopoly, to vibrant competition.

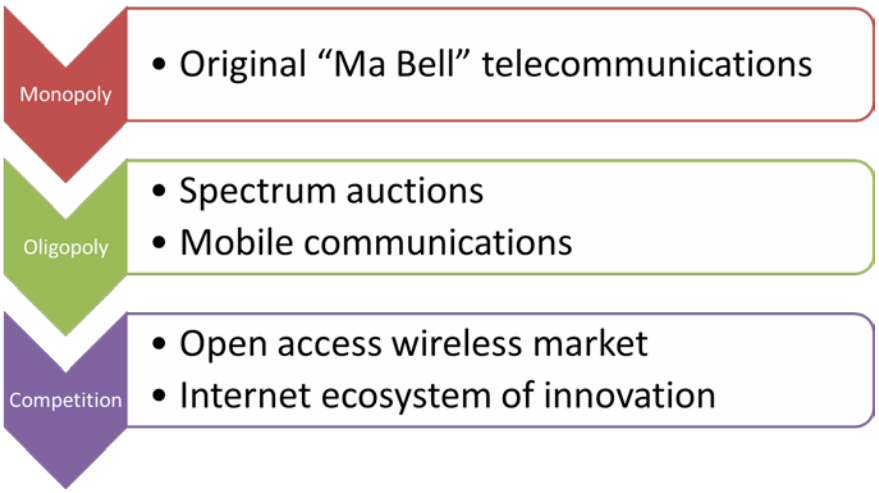


Figure 1: Evolution of communications markets from monopoly to vibrant competition

Overview of market design

The key innovation of the market is the creation of an open access network operated by the Independent System Operator (ISO). Figure 2 is a simple depiction of the market structure. In this example, there are three distinct mobile networks—the open access network and two proprietary networks, operated by MNO_1 and MNO_2. The mobile networks have the physical capability to provide mobile communications. The service providers are the entities that provide these mobile communication services to users (or devices). In the figure, the two mobile network operators, MNO_1 and MNO_2, also act as service providers; they are vertically integrated (AT&T and Verizon would be examples in the US). A third (and potentially many more) service provider operates as a mobile *virtual* network operator. MVNOs provide mobile communications to users by securing physical network resources from the wholesale market, for example by buying these resources from the open access network.

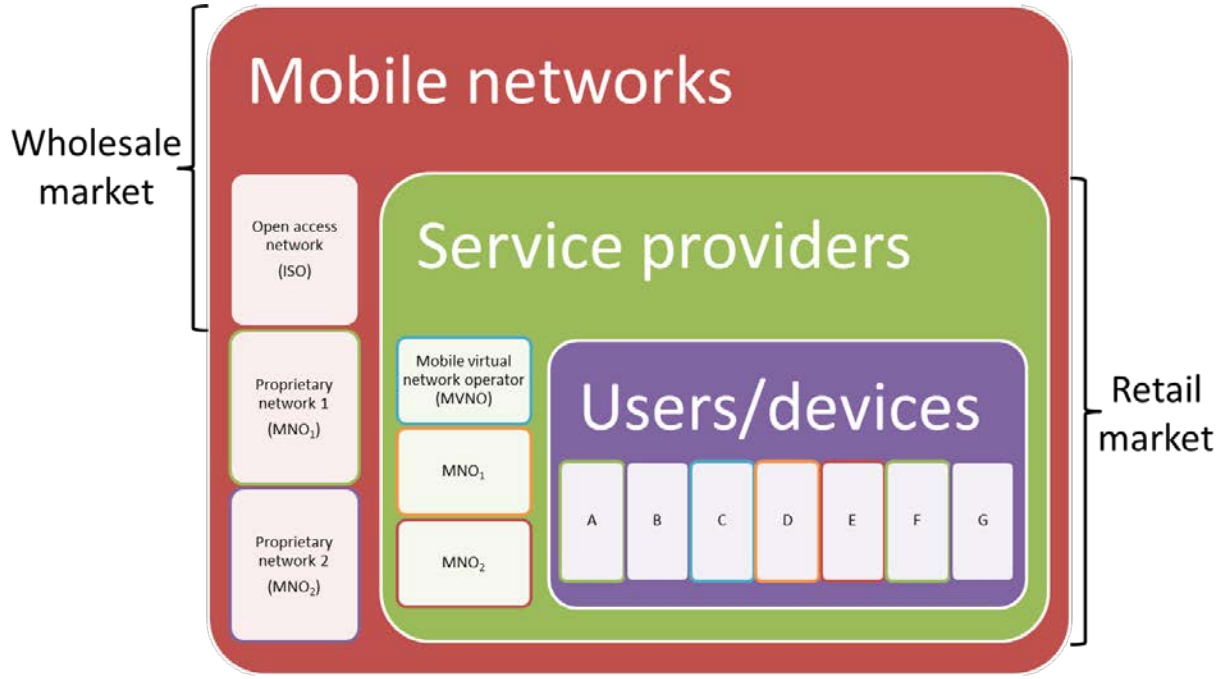


Figure 2: Wholesale and retail markets for mobile communications

Thus, there are two distinct markets, the wholesale market, in which mobile networks offer network resources to service providers, and a retail market, in which service providers offer mobile communications to retail users. This is exactly how restructured electricity markets are organized. By contrast, existing mobile communications markets create an oligopoly structure in which a few mobile network operators offer mobile communications directly to users. This structure, much like the monopoly utility structure of old, limits competition among service providers. The modern structure has no such limits.

This paper focuses on the wholesale market, the trading of network resources between mobile networks and service providers. As in electricity markets, it is necessary to assign and price network resources at each time and location to make sure that gains from trade are maximized and supply and demand balance at every time and location. This means that a time and locational "real-time" market is the backbone of the market, standing on the foundation of open access.

One issue of real-time markets is that they may involve significant risk. Prices become more volatile as we get closer to real time (shorter time intervals) as market participants have fewer and fewer options, which make supply and demand curves steeper, and prices therefore more sensitive to even modest quantity changes.

To allow market participants to make plans and better manage risk, the real-time market is complemented with forward markets in which market participants can buy or sell network resources in advance for longer time intervals and larger geographic areas. The forward markets are financial markets (cash settled), allowing participants to take positions well in advance of real time. The real-time market is physical; that is, it involves the physical delivery of mobile communications. Deviations from forward positions are settled at real-time prices.

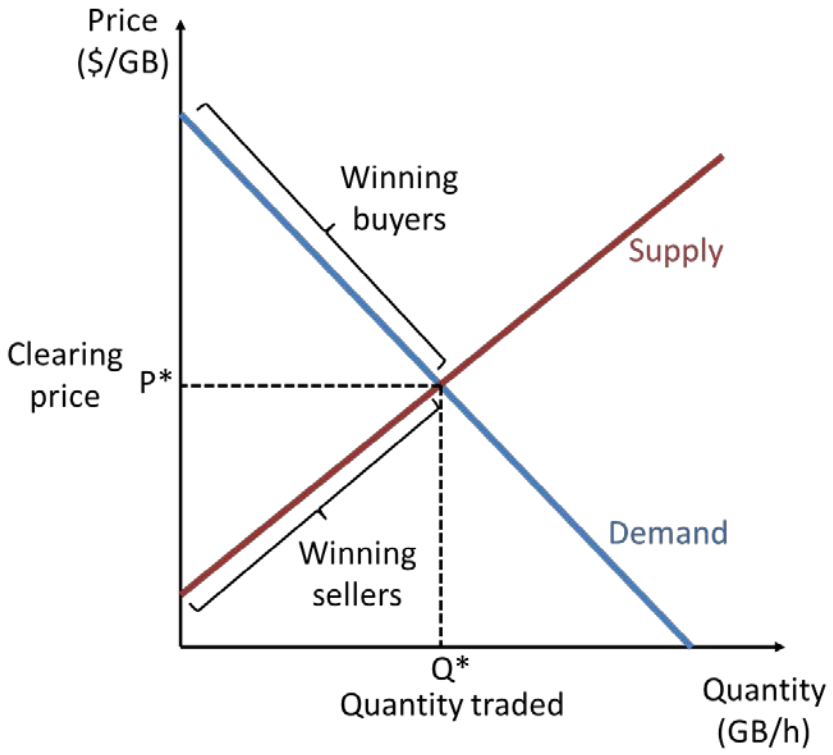


Figure 3: Market clearing at a particular time and location

The forward and real-time markets are structured as two-sided auctions in which network resources are traded. Those selling network resources provide a supply curve indicating the quantity they would like to sell at various prices (more at higher prices); those buying network resources provide a demand curve indicating the quantity they would like to buy at various prices (more at lower prices). The ISO aggregates both the supply and demand curves for each product. The intersection of the aggregate supply and demand curves determines the clearing price (P*) and the quantity traded (Q*), as shown in Figure 3.

Product design

In the design of any new market, one of the most critical decisions is what is the product being traded. The basic idea is to simplify to a more limited number of products from the near-infinite possibilities that exist in practice. Although every kernel of wheat is unique, it makes sense to trade wheat within a small number of quality grades. This creates liquidity for the product being traded, resulting in a more robust and effective market. The same is true with mobile communications. We will want to aggregate time and location to some extent to create liquidity and avoid having too many products. Here the product is gigabytes of mobile data communications at a particular location and time interval. At issue is the length of the time interval and the partition of geographic market.

A more complex variation would involve multiple grades of service at each time and location. Our view is that quality of service, which certainly is important for consumers, may be best managed within the retail market, rather than at the wholesale level. This simplifies the wholesale market.

The challenge is to choose the parameters of the complementary forward and real-time markets in a manner that maps to the cadence and spatial dimensions of the mobile communications world, while creating the desired liquidity and a manageable number of products.

In general service providers think of longer-term usage trends on a yearly basis. It therefore makes sense to offer forward products on the open access market that allow the purchase of capacity for yearlong periods. A second level of refinement can happen through medium-term products. There are seasonal variations in demand that manifest themselves on a monthly basis. Growth/decline in customer bases may also be evident at the monthly level. Therefore, the ability to buy or sell for the month-ahead makes for a reasonable medium-term product. The ultimate fine-tuning would involve resolving demands in real-time. In areas where inadequate throughput has been bought in advance, bids for extra capacity can be made. And in cases where excess capacity is held, surpluses can be sold on the market. We will return to the exact definition of real-time later in this section.

To design a set of products in line with the above approach requires a deeper understanding of the service provider demands and of the network supply.

The demand side

Today service providers typically request capacity on a host network in the number of Terabytes desired. This type of request takes no account of time or spatial variation in usage patterns. In reality user demands for network capacity vary with time and space. The simplified patterns in Figure 4 show user demand for three different hours of the day, on a per cell basis in a specific geographical area. Red indicates a high demand, amber a medium demand and green a low demand. In the off-peak hour, the user demand is much lower, as expected.

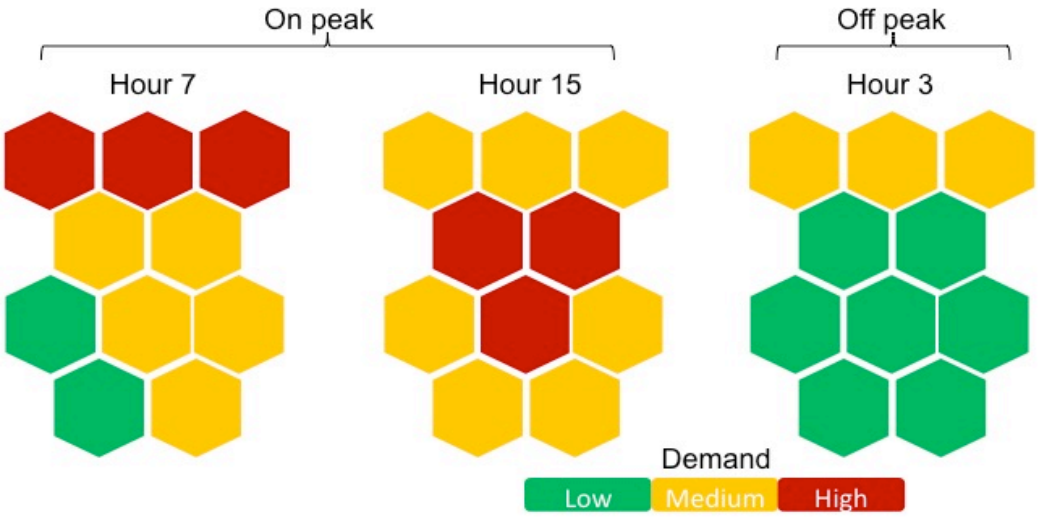


Figure 4: Demand patterns in three sample hours of ten sub-areas of a specific region

In an ideal world, a service provider would be able to make requests for capacity at specific times in specific locations. This would benefit both the service provider seeking the capacity and the network operator supplying the resources. The service provider would only pay for the resources needed and the network operator would be better able to support multiple service providers and more effectively share resources.

The first step in facilitating this is to express demand in a more textured way. Because expressing demand as an absolute amount of data suppresses the temporal variation in demand, it is therefore important to think of "capacity over time" and this leads to some measure of *throughput*. A data metric such as 1 Gigabyte (GB) in an hour takes this into account. However thinking of capacity in terms of throughput is not enough. The spatial dimension must be accounted for by linking this throughput to a specific location. Hence we think of capacity in terms of *throughput at a location*.

The supply side

To be able to bid for throughput at a location, the network operator needs to make products available that are expressed in terms of throughput and location. In principle, the wholesale network can be divided into its smallest geographical basic unit; that is, the cell or more precisely the cell-sector, together with the available supply in terms of GB/h determined for each of these basic units. This granularity of product could be used for the yearly, monthly and real-time markets. In practice, however, such extreme granularity would not lend itself well to the efficient realization of the market. The service provider would need to request throughput per cell or cell-sector for every cell-sector in the network, leading to intractable and unnecessarily onerous forward auction processes. In line with the driving principle that simplicity is desirable, it makes sense to group the smallest geographical building blocks (cell-sectors) into larger areas for the purposes of the yearly and monthly auctions and indeed to aggregate cell sectors in the real-time market.

There are many options for aggregating the basic geographical units into bigger areas. Our approach is twofold. First, we adopt the principle that the areas form a hierarchy: the real-time market is the most

granular, each monthly area is an aggregation of real-time areas, and each yearly area is an aggregation of monthly areas. Second, we are guided by jurisdictional boundaries such as boroughs (New York) or departments (Mexico) in making decisions about the geographic areas. Thus the aggregated areas still will be at a granularity that makes it feasible for small service providers to bid for resources to support local services. Accommodating the needs of a local provider makes the market more competitive, while providing equal access of opportunity.

It is important to make two additional points relating to the real-time market and the basic geographical unit over which it will apply. The technology deployed in the wholesale network has implications for what real time means. It is technically feasible to be able to support real-time markets on *a per second basis*. However in the near-term, current network deployments do not support this level of agility. Hence in the near-term *real time means responding to hourly demands*. Then with further technical advance we can reduce this interval to minutes or even seconds. The same point can be made about the basic geographic unit. Based on our current understanding of architectures and our understanding of the ability to effect control, that *basic geographic unit is an aggregation of cell-sectors*, but as technology develops and architectures change that may become more granular. Figure 5 gives an example of areas over which the yearly, monthly and hourly auctions will take place.

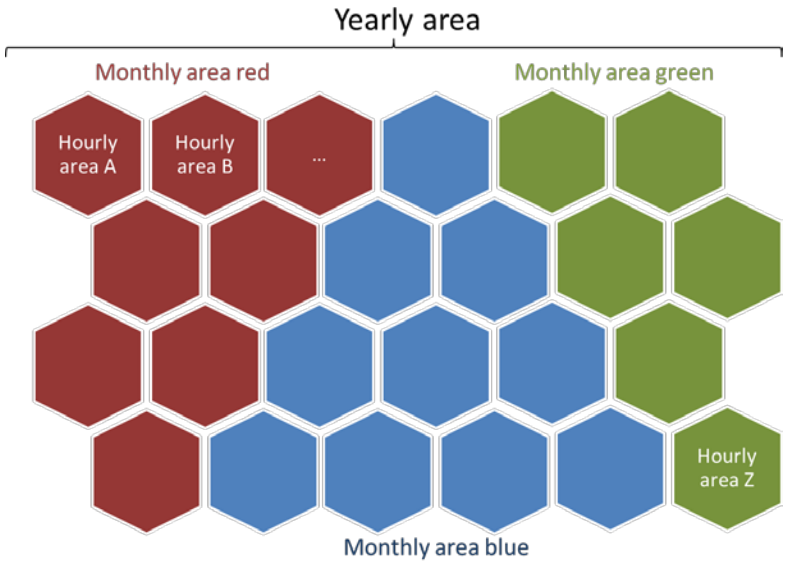


Figure 5: Yearly, monthly, and hourly geographic areas form a hierarchy — yearly > monthly > hourly

The products on offer over these different areas and for these different time intervals depend on the available supply in the wholesale network. Typically, there will be a baseline level of supply that the open access network has at its disposal. Investment in the network or pooling of additional resources will increase capacity from this baseline. Isolated events, such as public safety needs, may cause a temporary reduction in supply. This picture matches well to the demand side cadence. The baseline supply should be obvious for the year ahead. Investment or new partners coming on line will manifest themselves in changes in monthly supply and transitory events will manifest themselves in hourly supply variations.

As a final broad point, it will make sense to have separate products for peak and off-peak hours. This is common in electricity markets, as the demand profile is quite different between peak and off-peak hours. Service providers will want to bid separate prices and quantities for peak and off-peak hours in the yearly and monthly auctions.

The auction details will be introduced in later sections of the paper, but to provide insight into how the markets work, a worked example follows.

A worked example

For the yearly market, the network operator predicts the average capacity that will be available in the network over the yearly area. Not all hourly areas in the yearly area will have the same capacity. Hence the total expected supply for the yearly market—in GBs for every hour of the year—is *the sum of the individual hourly throughputs in GB/hr of each hourly area in the yearly area*. The network operator may decide to put all or a percentage of this capacity on the market, the latter being the more common scenario. The product that goes on the market therefore is a total hourly rate for every hour of the year for the yearly area, disaggregated across the hourly areas.

The service providers will bid for all or a portion of this product based on their own average demand projections for the yearly area. Service providers' bids are price dependent, reflecting the quantity desired at various prices. Figure 6 shows the outcome of the yearly auction for a specific service provider for the peak-hour product. In this example the service provider successfully bids for a total of 165 GB per hour for every hour of the year disaggregated across the different cells as depicted in Figure 6. Each hexagon represents a particular hourly area. The disaggregation of the yearly bids is based on pre-announced breakdowns by the ISO and is not subject to individual service provider preference.

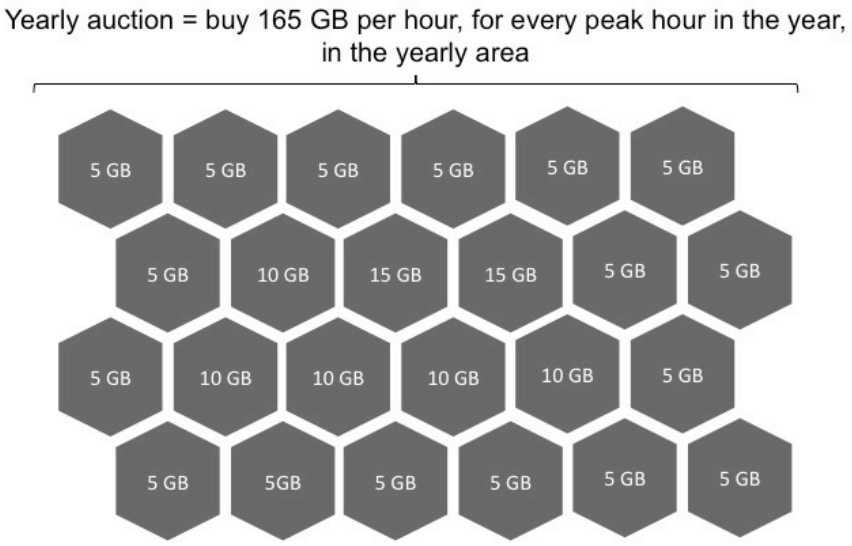


Figure 6: The forward position for the year for every peak hour in the year for a particular yearly area

The network operator goes through the same process for the monthly auctions, except that now the network operator offers products per disaggregated monthly area rather than the larger yearly areas. The network operator can put portions of the total capacity that was held back on the yearly market on the

monthly market or in cases where there has been investment in the network, place newly available capacity on the monthly markets. For the purposes of this example the yearly area is divided into three monthly areas, namely red, green and blue, as was used in Figure 5. For simplicity of illustration, the capacity on offer by the network operator is distributed evenly across the cells.

Each month the service provider has the opportunity to adjust the position up or down over smaller areas. Therefore in the monthly auction, the service provider can also be a seller and place unwanted capacity back on the market. The service provider predicts different average demands for each area for the month ahead and adjusts the forward positions accordingly. The service provider now can choose to take a *different* monthly position in each monthly area. In the example here, the service provider decides to lower its position in the red monthly area and hence sell capacity, and to increase its position in both the blue and green monthly areas. Figure 7 shows the monthly position of the service provider following the monthly auction based on the following preferences:

- In the red monthly area the service provider sells 1GB of throughput for each peak hour in the month at each hourly area (-8GB in total).
- In the blue monthly area the service provider buys 2GB of throughput for each peak hour in the month at each hourly area (+20 GB in total).
- In the green monthly area the service provider buys 3GB of throughput for each peak hour in the month at each hourly area (+18 GB in total).

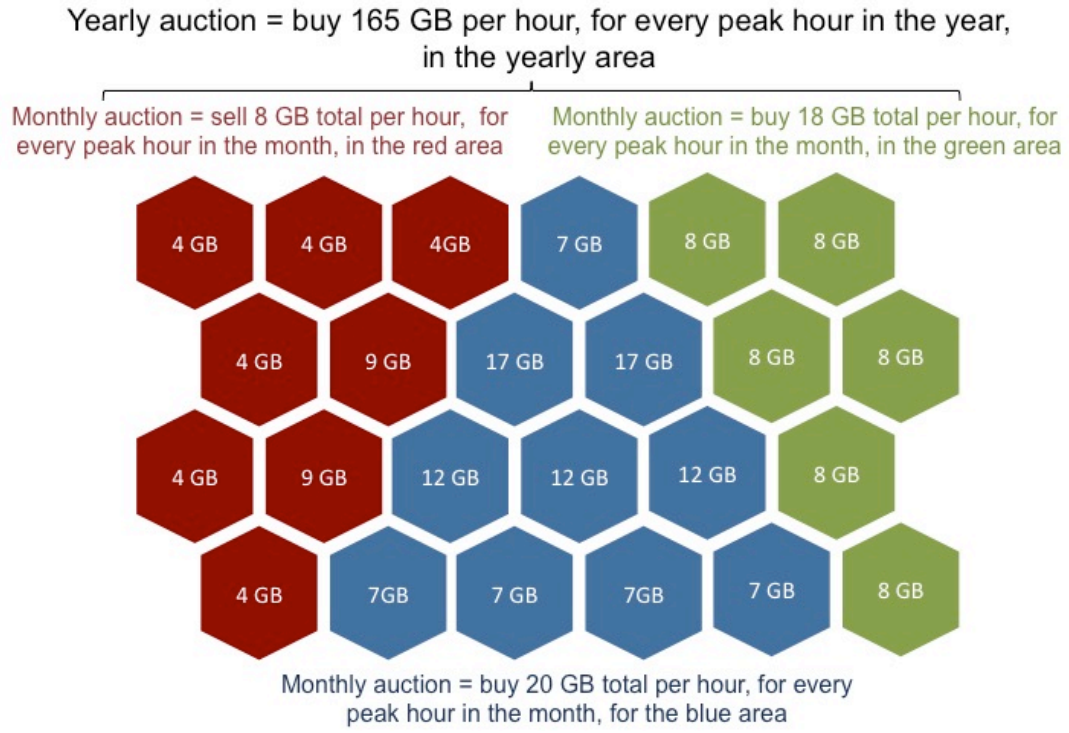


Figure 7: Monthly auction alters forward position in each monthly area

The forward markets are financial instruments. In the real-time (hourly) market, the discrepancies between the throughput that has been bought in advance and the required throughput are resolved.

Figure 7 focuses in on one cell, and the per-hour variation in demand. The reserved forward position meets demand in some cases but in other cases is surplus to demand or there is a deficit. The service provider makes bids at the hourly area level to resolve any discrepancies.

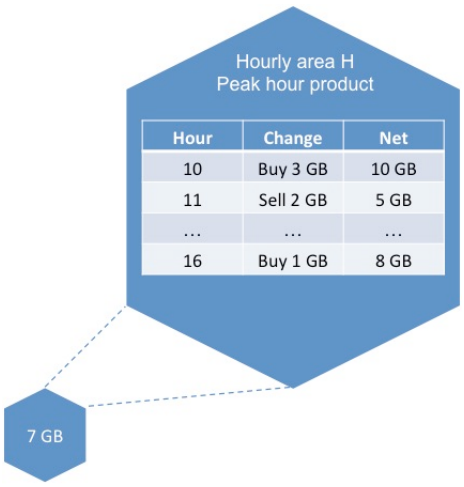


Figure 8: Hourly auction further alters demand for each peak hour for hourly area H

If no congestion exists on the network additional throughput can be accessed at nominal cost. This is what open access means. This drives efficiency of resource usage and leads to new levels of transparency around the cost of those resources. Obviously if the network resources are heavily in demand, the service provider may not be in a position to meet all its users' requirements. This can be addressed by buying more throughput on the forward markets, but with the potential downside of higher costs. Overall the level of dynamism and fine-tuning that is supported by the open access market far outweighs the static systems that are currently in place and that are predicated on the naïve over-provisioning of resources.

Three opportunities to trade

The market allows both forward and spot trading. Specifically, as discussed in the previous section, there are three opportunities to trade, yearly, monthly, and hourly. The hourly market is the current spot or real-time market. The sequence of auctions is illustrated in Figure 9.

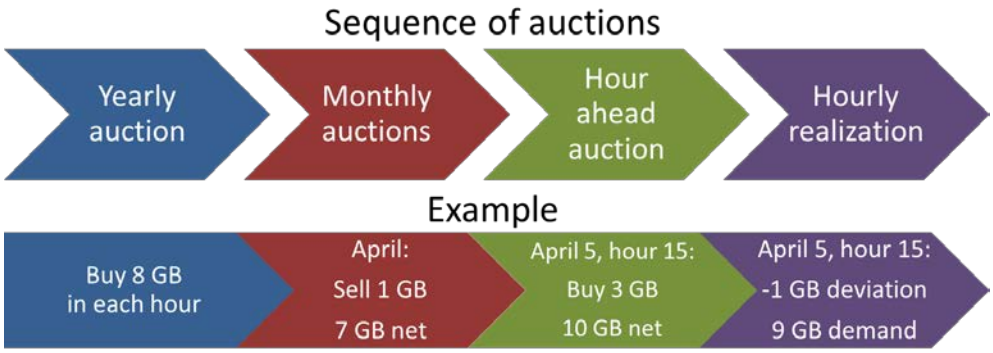


Figure 9: The sequence of auctions with an example

The motivation for forward trading is best appreciated from the point of view of a service provider. The service provider has a demand for throughput based on the anticipated data needs of its customers. Although these demands cannot be perfectly known for the year ahead, they can be estimated. By satisfying some of its demand in the yearly auction, the service provider is able to lock in prices for a significant fraction of its need, reducing risk. Then just before the month, the service provider can better estimate its need in the month-ahead. The service provider can buy additional data or sell any surplus in the monthly auction. Finally, the service provider makes a final determination of its needs and uses the hourly auction to make any desired adjustments. This sequence of three auctions gives the service provider flexibility in timing its purchase of data throughput to meet its demand at least cost and minimal risk.

Speculators may play a useful role arbitraging between markets. A speculator will look for price anomalies in the yearly and monthly markets. The speculator looks to buy underpriced products and sell overpriced products. Since the hourly market is a physical market, the speculator sells any positive position and buys any negative position in the hourly market, leaving the speculator with zero net demand in each hour. To the extent that the speculator makes money, the speculator's bids drive prices to more competitive levels. In equilibrium, the yearly price is the expectation of monthly prices, which is the expectation of hourly prices. Of course service providers, armed with detailed knowledge of the market, are apt to play the speculator role as well.

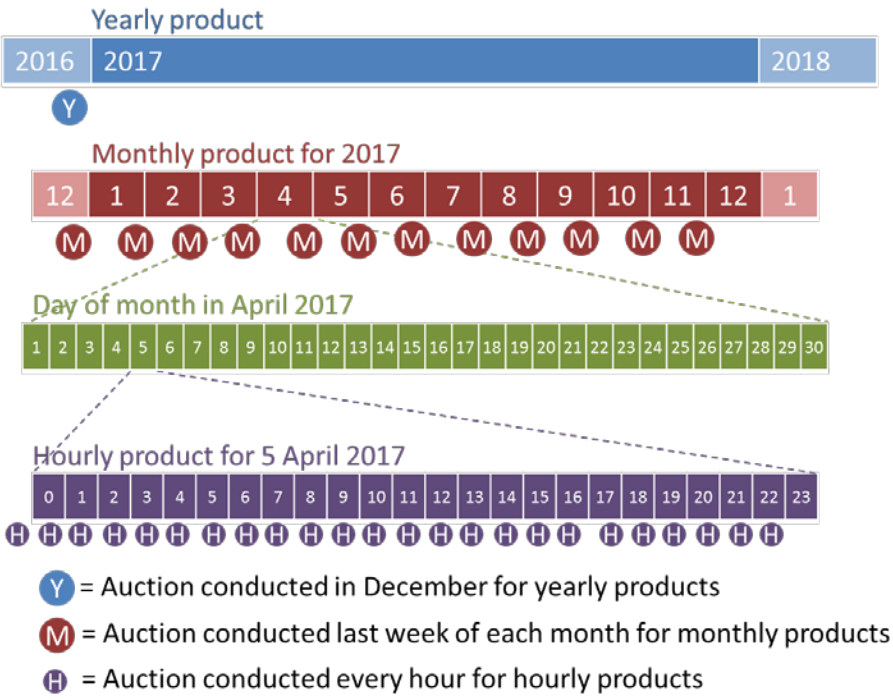


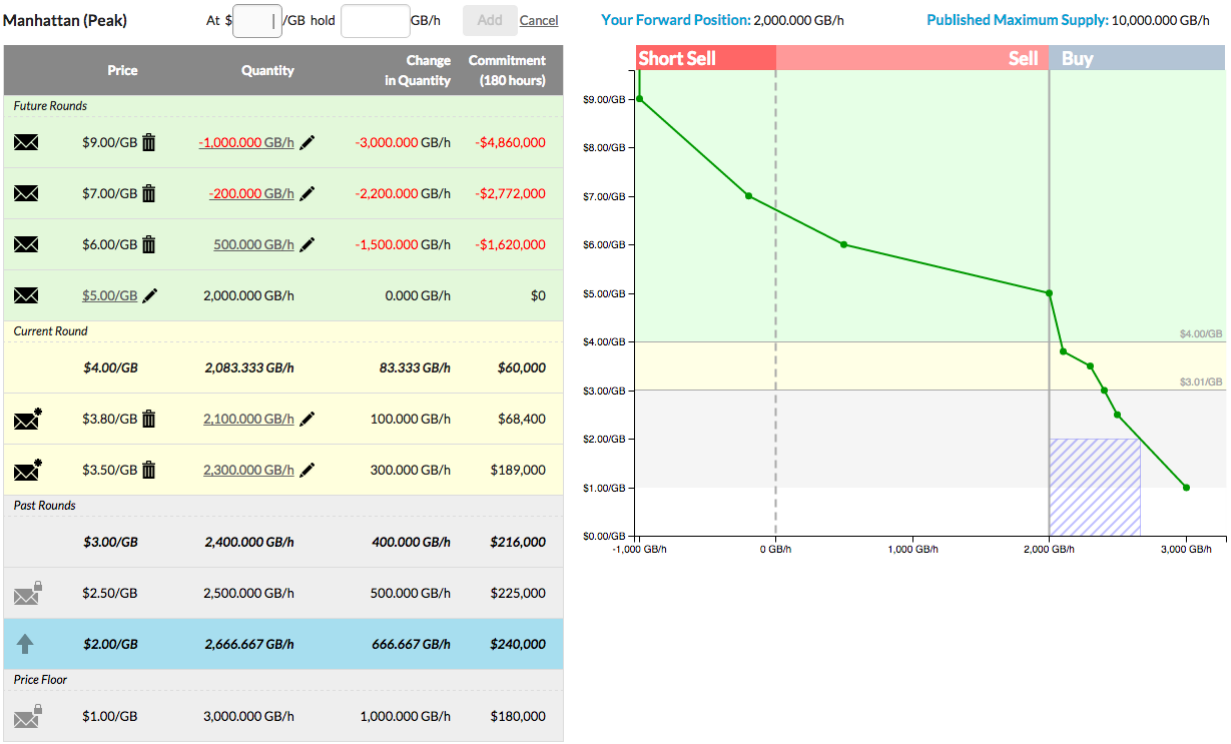
Figure 10: Timeline representation of yearly, monthly, and hourly auctions

Figure 10 shows the timeline of the auctions in greater detail. There are 1 + 12 = 13 forward auctions each year and 365 × 24 hourly auctions. All auctions are single-price (uniform-price) auctions. For each product, the auctioneer collects the demand bids and supply offers as strictly monotone piecewise-linear functions. The auctioneer then identifies the unique clearing price for each product where supply and demand balance.

Yearly and monthly auctions involve greater volume. In electricity markets 80 to 90 percent of volume transacts in forward markets. We anticipate a similar split in mobile communications markets. When auctioning many related products in infrequently conducted forward auctions, bidders find it helpful to learn about market prices and likely winnings during the auction, while they can still make adjustments to their bids. This learning about prices and winnings is called *outcome discovery*. To allow outcome discovery in these auctions, the auctioneer collects bids with a simultaneous ascending clock auction. In each round the auctioneer asks bidders to express a piecewise-linear demand curve for a range of prices.

Figure 11 gives an example of a service provider bidding for the Manhattan monthly product (peak hours). Past rounds are shaded in gray, the current price range is in yellow, and future prices (proxy bids) are in light green. The demand curve is expressed with a sequence of price-quantity pairs (left panel). The right panel traces out the resulting demand curve (green line). Bids in the grayed-out range cannot be changed; all other bids can be adjusted and new bids entered consistent with the requirement that the demand curve is downward sloping. Offers to sell are expressed in the same way as negative demand. Thus, a single screen can be used to express both buying and selling. In this example, the service provider has a positive position of 2,000 GB/h from the yearly auction. This is why for quantities less than 2,000 GB/h, the bidder is selling in the monthly market to reduce its net position. At prices above $6.66, the bidder sells its entire forward position from the yearly auction, and indeed engages in a short sale. At these very

high prices, the bidder is anticipating that it will be able to buy all it needs and resolve the short position in the hourly market.



	Price	Quantity	Change in Quantity	Commitment (180 hours)
Future Rounds				
	$9.00/GB	-1,000.000 GB/h	-3,000.000 GB/h	-$4,860,000
	$7.00/GB	-200.000 GB/h	-2,200.000 GB/h	-$2,772,000
	$6.00/GB	500.000 GB/h	-1,500.000 GB/h	-$1,620,000
	$5.00/GB	2,000.000 GB/h	0.000 GB/h	$0
Current Round				
	$4.00/GB	***2,083.333 GB/h***	***83.333 GB/h***	***$60,000***
	$3.80/GB	2,100.000 GB/h	100.000 GB/h	$68,400
	$3.50/GB	2,300.000 GB/h	300.000 GB/h	$189,000
Past Rounds				
	$3.00/GB	***2,400.000 GB/h***	***400.000 GB/h***	***$216,000***
	$2.50/GB	2,500.000 GB/h	500.000 GB/h	$225,000
	$2.00/GB	***2,666.667 GB/h***	***666.667 GB/h***	***$240,000***
Price Floor				
	$1.00/GB	3,000.000 GB/h	1,000.000 GB/h	$180,000

Figure 11: Sample demand curve for a bidder for the Manhattan (peak) monthly product

The ISO determines the supply curve that is offered in the forward auctions. The supply curve is set to be consistent with the service providers' forward demands. Much like an effective speculator, this approach has the benefit of stabilizing prices across the three markets. To motivate participation, the ISO announces the supply curve in advance of the auction. An example is shown in Figure 12.

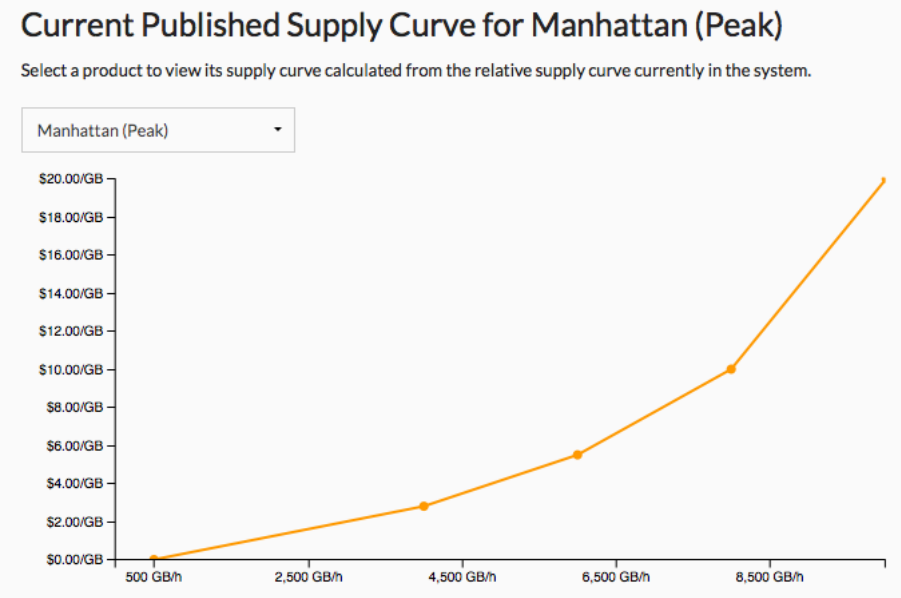


Figure 12: Sample supply curve for the Manhattan (peak) monthly product

At the end of each round, the auction system reports the demand up to the end of round price. This is shown in Figure 13.

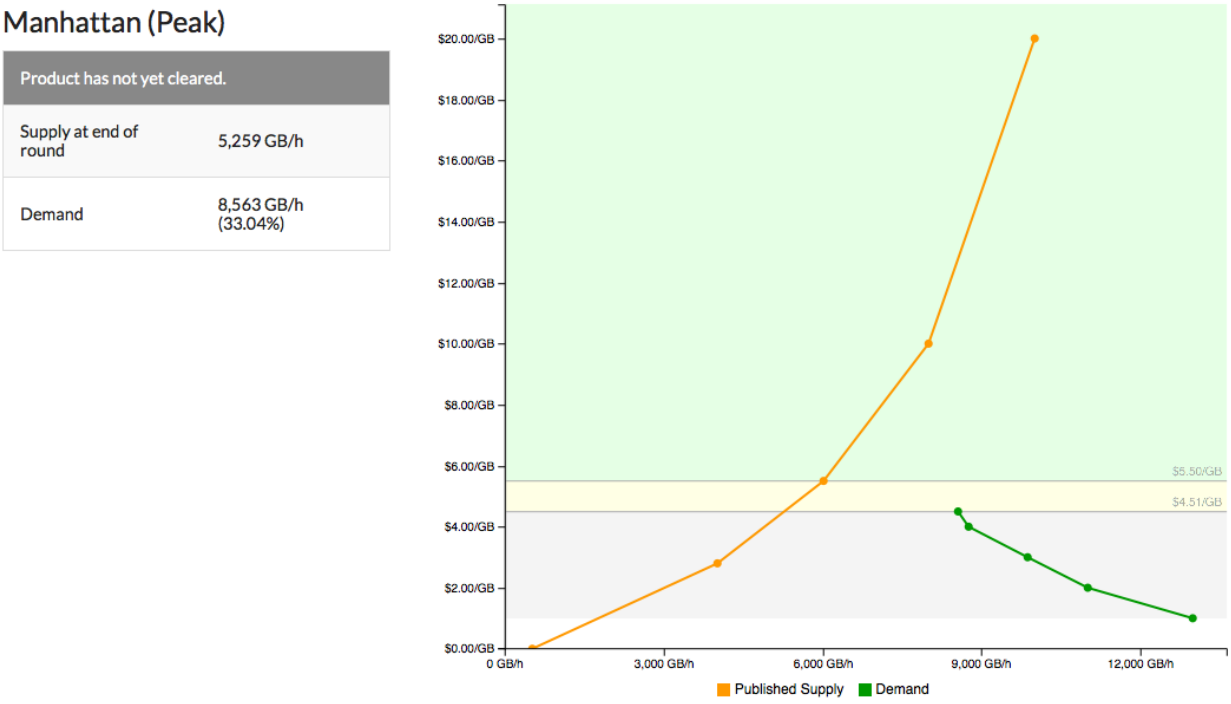


Figure 13: Bidder's screen showing supply and demand up to the end of round price ($4.50/GB)

The auctioneer sees much more detailed information. Figure 14 gives an example of the auctioneer's screen, showing both the full aggregate demand curve, including demands that are not yet binding (the

portion in the non-grey price range. The auctioneer can also see individual demands. This information is helpful as the auctioneer guides the pace of the auction.

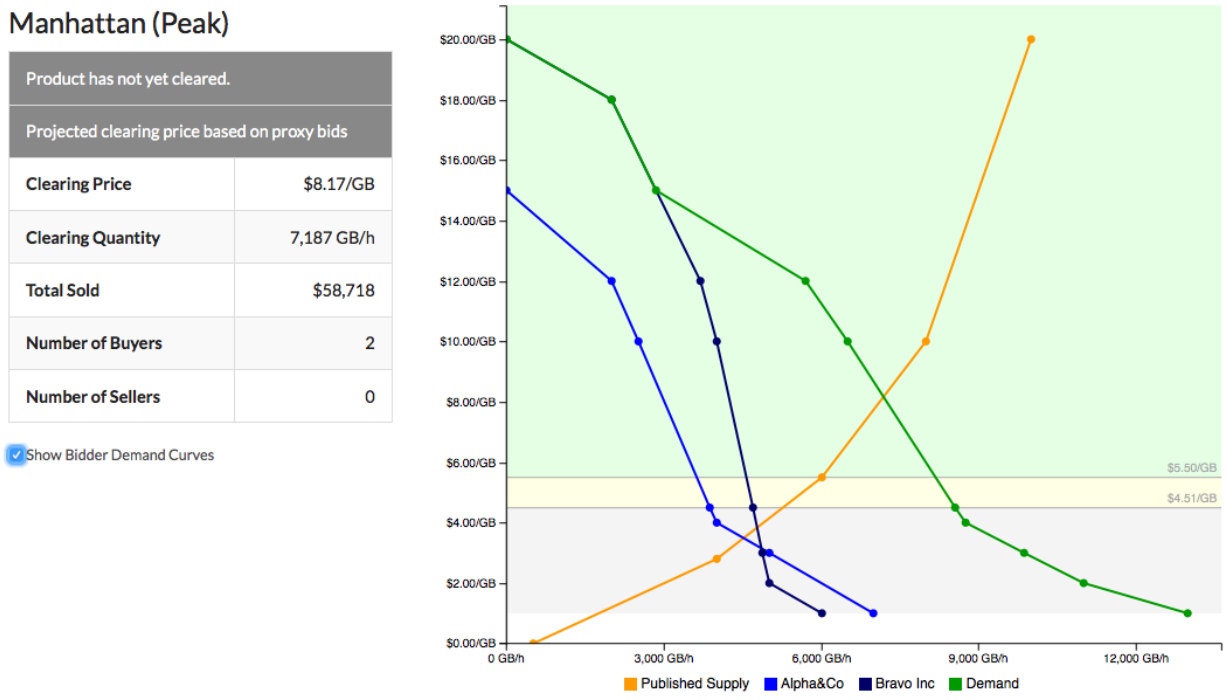


Figure 14: Auctioneer's screen showing published supply (amber), aggregate demand (green) and individual demands

There are two differences in the hourly auctions. First, because of the smaller volume and the much greater frequency, outcome discovery is less important. Indeed the prior hourly auction provides excellent information about prices and likely winnings in the current auction. As a result the auction is conducted in a single round (sealed-bid). Second, since this is the last opportunity to trade, all remaining supply including any short positions is offered for sale.

Price floor

The market rules include a price floor, greater than or equal to zero. Supply is not offered at prices below the price floor. The price floor is intended as a means to assure some usage-based revenue accrues in areas with surplus capacity. This strengthens incentives for network investment in low-demand areas. Consistent with the open access principle, the price floor ideally is a nominal amount at or near zero. However, due to the lumpiness in investment, especially in low-demand areas, there is an economic justification for setting the price floor at long-run marginal cost.

The market clears either at the price floor in the event of surplus (supply greater than demand) or at the marginal value of demand where excess demand is zero.

Real-time balancing of supply and demand

The hourly auctions are conducted one hour ahead to give service providers an opportunity to control realized demand to match purchases. During the hour, the service provider can take actions to limit deviations between its realized demand and its net auction purchase. This matching of realized and purchased demand is desirable as it forms the basis for efficient trade and pricing.

The extent of service provider control is defined in part through the contract forms the service provider has with retail users. This is very much a part of existing electricity markets, where different contracts allow more or less price-based intervention in the demand of its users. Good examples from electricity are the control of heating and cooling units based on prices. These programs have been in existence for more than 20 years and are greatly expanding. Demand response is now a major element of wholesale electricity markets. Demand response can play large role in mobile communications both at the device and network levels. For example, devices could have contract and user dependent preferences that cause changes in demand in response to price signals.

The real-time market should provide incentives that motivate service providers to estimate as best they can their hourly demand at the location and then bid that quantity as a function of price in the real-time market (the hourly auction at each hourly area). A system of penalties for deviations from hourly plans is a common method for inducing bidders to balance supply and demand in real time.

As an example, we propose the following simple settlement rule for deviations in the hourly (real-time) market. Define:

p_{hk} = price in hour *h* in hourly area *k* that balances as-bid demand with estimated supply

q_{ihk} = total quantity bought by bidder *i* in hour *h* in hourly area *k* (includes yearly, monthly, and hourly net purchases)

Q_{ihk} = actual quantity consumed by bidder *i* in hour *h* in hourly area *k*

$D_{ihk} = Q_{ihk} - q_{ihk}$ = deviation between actual quantity consumed and quantity bought

Tolerance = percentage tolerance band (e.g., 5%); no penalty if deviation is within Tolerance

Penalty Factor = a factor that is applied to square deviations above Tolerance

The adjustment for deviations in the real-time market is $Adjustment_{ihk} = p_{hk} \times D_{ihk} + Penalty_{ihk}$, where

$$Penalty_{ihk} = \begin{cases} 0 & if \ \dfrac{|D_{ihk}|}{q_{ihk}} \leq Tolerance \\ Penalty\ Factor \times p_{hk} \times D_{ihk}^2 & if \ \dfrac{|D_{ihk}|}{q_{ihk}} > Tolerance \end{cases}$$

Thus, the adjustment for deviations simply accounts for the deviation in the usually way provided the deviation in percentage terms no more than the Tolerance; that is, the service provider pays the hourly

price times the deviation. However, when the deviation is larger in percentage terms than the Tolerance, then there is a second term equal to the price times the Penalty Factor times the squared deviation. It is this penalty for large deviations that motivates the service provider to estimate its hourly physical demand accurately and then to take steps to during the hour to attempt to get its actual demand within the tolerance. Since due to the law of large numbers, it should be easier for a large service provider to better estimate its aggregate hourly demand, it may make sense for the Tolerance to depend on the service provider's size, with higher Tolerance allowed for smaller service providers. Smaller service providers also have less incentive and ability to manipulate the market price through strategic bidding.

Secondary market for forward products

The yearly and monthly auctions give service providers a good opportunity to manage risks and take positions consistent with plans. Secondary market trading on an exchange can give participants even greater flexibility to make adjustments as plans change. One challenge of the secondary market is the large number of products due to the time and locational structure. This may hamper liquidity, especially for the less valuable products. In the simplest design, the secondary products are for the remainder of the relevant period. For example, a yearly product trading toward the end of March would cover April through December.

Coexistence of open access and existing markets

The open access network complements existing networks. Nonetheless, dominant incumbents at least initially are apt to resist open access as a result of the heightened competition it brings. However, smaller incumbents may benefit immediately as it provides a means to achieve coverage at lower cost and to monetize surplus network resources. In the longer run, all network operators and service providers are apt to benefit from the improved use of resources. Competition and innovation will lead to a vibrant and expanding mobile communications industry. Consumers and efficient providers will benefit from open access.

Promising opportunities for open access markets

The open access market will add value in all countries. Here we illustrate the promising possibilities with a brief discussion of important avenues for open access wireless in Mexico, the European Union, the United States, and Singapore.

Mexico: enhancing competition

Mexico has set aside 90 MHz of 700 MHz spectrum to jump-start an open access market. The low-band spectrum is perfectly suited to the task as it allows an economic buildout with deep coverage in buildings, rural areas, and difficult terrain. The band is globally harmonized and fully supports the latest LTE standard. It also offers sufficient bandwidth for excellent capacity, especially at its start. Additional spectrum can be added in urban areas either from incumbent carriers or in spectrum auctions such as the AWS and 2.5 GHz auctions already planned for 2016 and 2017.

The primary intent of the initiative is to enhance competition. Mexico has excess concentration in mobile communications; the dominant incumbent (Telcel) has 71 percent market share and the top-2 (Telcel +

Movistar) have 92 percent. The open access market is especially desirable in such a setting, where the gains from low-cost entry will be dramatic. The market will quickly become highly competitive and innovative. Large segments of the population who were unable to afford mobile communications will gain mobile access to the Internet. Connectivity will allow the Internet of Things to arrive sooner and with greater impact in Mexico.

European Union: merger remedy

As the mobile communications industry matures, there is constant pressure for consolidation. Mergers from four carriers to three carriers have been proposed in many countries. In some cases, such as the proposed merger of AT&T and T-Mobile in the US, the regulators rejected the merger arguing that the identified operational efficiencies were insufficient to offset the loss in competition. Similarly, the proposed merger of Sprint and T-Mobile in the US was set aside for the same reason.

By contrast mergers in the European Union that have brought the number of national competitors from four to three have recently been accepted in Austria (2012), Germany and Ireland (2014). In each case, the European Commission required a remedy to address competition concerns. In each case, the merger remedy involved, in part, requiring the merged entity to set aside a portion of network capacity that would be made available to service providers (MVNOs).

In the first case of Austria, the remedy involved the merged entity offering 30 percent of its network capacity at prices set administratively. One problem with this approach is the merged entity has an incentive to set the capacity price too high, so as to discourage MVNO entry and weaken any downward price pressure should MVNOs enter. Although one MVNO did decide to enter, entry did not occur for two years after the merger and the share of network capacity sold under the remedy is a tiny fraction of the 30 percent offered. Indeed, the most successful MVNO entrant in Austria has a market share of 2.5 percent and purchases its network capacity from outside of the merger remedy.

To address this weakness, the subsequent MVNO capacity remedies in Germany and Ireland involved the long-term sale of capacity. This approach encourages the MVNO to offer its capacity to consumers. However, the approach in Germany and Ireland still suffers from inflexibility. The MVNOs are established and priced in advance. Prospective MVNOs are forced to guess about customer volume far in advance, rather than adjust capacity purchases in response to market fundamentals. This shifts risk to prospective MVNOs and harms the efficiency of the capacity assignment.

The open access remedy begins in a similar fashion—the merged entity commits to set aside a portion of network capacity to service providers (MVNOs) *in an open access market*. The open access remedy allows the outcome, both assignment and pricing, to be determined competitively and evolve with market fundamentals.

There are several advantages of the open access remedy compared with earlier MVNO capacity remedies.

- Competitive forces determine the identity and number of service providers.
- The open access market determines both the assignment of capacity and the wholesale prices to maximize gains from trade at every time and location. Prices reflect the marginal value of demand

at the particular time and location. In sharp contrast, prices in the MVNO capacity remedy are set in private negotiations between the merged entity and potential service providers—a negotiation in which the merged entity benefits from both market power and asymmetric information.

- The open access market lets each service provider purchase a quantity of capacity that is consistent with its underlying demand at each time and location. Both assignments and prices adjust as the market evolves, yet service providers can manage risks with forward purchases on a yearly and monthly basis. In contrast, the MVNO capacity remedy requires the service provider to purchase a large amount of capacity initially and without the flexibility to re-optimize purchases on an ongoing basis.
- The open access market is an enduring arrangement. It adds an important means to encourage competition among service providers and encourages the efficient use of scarce network capacity.
- The open access market has an organizational and governance structure that will adapt and improve as the market evolves.
- The merged entity receives the competitive market value of the open access capacity and can take advantage of this feature to offer additional capacity on the open access market as desired to optimize the value of the network.
- Finally, the merged entity has an ongoing incentive to encourage the success of services providers buying from the open access market, so as to increase wholesale revenues. This incentive mitigates the merged entity's incentive to foreclose competition from service providers so as to improve retail revenues.

One important element of competition we have yet to discuss is competition among mobile networks. This competition is essential in providing incentives to improve the network. A four-to-three merger may weaken investment incentives from the loss in network competition. Network competition is important as it can impact quality of service to consumers. The potential loss of network competition can be addressed within and outside the remedy. Within the remedy, the merged entity can set aside some quantity of spectrum that is auctioned to prospective network operators, who would build network capacity and sell that capacity in the open access market. Outside the remedy, the regulator can set aside spectrum for open access use and limit the spectrum award to non-dominant incumbents.

A related remedy has been used since 2001 in response to electricity and gas mergers in Europe. The first and most significant instance was the merger of France's EDF with EnBW, a German energy company. The remedy involved the quarterly auction of a significant share of EDF generating capacity in the form of option contracts, called "virtual power plants" (VPP), with durations between three months and three years. The key idea was to have the same impact from a competitive viewpoint as physical divestiture without losing any of the efficiency gains of the merger. The approach was proposed by EDF and adopted by the regulator. Later the VPP approach was replicated in many other European countries for both electricity and gas products in France, Germany, Denmark, and the Netherlands. The EDF VPP auctions continued for 12 years, were highly successful, and indeed EDF continued the auctions beyond the agreed term (Ausubel and Cramton 2010). The approach created much needed liquidity in medium-term energy products, and this fostered competition.

To summarize, the open access remedy is based on the fraction of network capacity that is offered via open access and the duration of the commitment. In particular, prices are determined by market forces at each time and location, rather than set somewhat arbitrarily in advance and on a long-term basis. The open access remedy brings the essential flexibility in pricing and allocation that is needed in a rapidly changing environment. This allows carriers and MVNOs to access network capacity at competitive prices and manage risks through forward purchase. MVNOs can enter knowing that network capacity will be available at competitive prices for an extended period. Merging carriers can enjoy operating efficiencies while consumers enjoy improved competition, and the entire market benefits from the more efficient allocation and pricing of network resources.

United States: public safety and universal service

In the US, merger remedy may also be a motivation for an open access wireless network. Regulators in the US would likely be far less negative on a Sprint/T-Mobile merger, the two smaller national carriers, if the merger came with the open access remedy.

However, the most immediate area of opportunity is spectrum that is set aside for public safety use, 20 MHz of 700 MHz spectrum (the FirstNet spectrum). Historically public safety spectrum has been underutilized for two reasons: (1) public safety has lacked the funds to build a state-of-the-art network and (2) public safety use is characterized by high-intensity use during emergencies, and emergencies are rare. A public-private partnership between public safety and a private system operator, which builds and manages the network, is ideally suited to solve both problems. Public safety retains priority use of the spectrum in emergencies, but the system operator manages real-time and longer-term markets for spectrum, ensuring that the capacity is used efficiently and providing funds for building and operating the network. Since capacity on the network would be available to all parties on an equal basis (aside from the rare emergency uses by public safety), the network provides an easy entry point for companies that have innovative ideas for using the spectrum but are currently excluded because of the difficulties of negotiating contracts with incumbent carriers. Negotiations are difficult because incumbent carriers are protective of their networks. This is not surprising given that the carriers have spent many tens of billions of dollars building their networks.

With time, this same public-private model can be expanded to other government spectrum. Forward thinking agencies are already acting on this model. For example, in a conference in early 2015, Stuart Timmerman, the Director of the Defense Information Spectrum Agency (DISA), said that DISA is looking at "commoditizing our radio spectrum and making it available in a dynamic marketplace. Who would have believed we would even think such a thing just two years ago?" The gains in terms of more efficient use of the spectrum, expanded opportunities for innovation, and improved competition in the wireless services market are large.

Singapore: new entrant set aside

Three carriers currently compete in the Singapore mobile market. The regulator (IDA), seeking to stimulate competition further, has set aside 60 MHz of spectrum—40 MHz below 1 GHz and 20 MHz at 2.3 GHz—for a new entrant in an auction to happen in 2016. One possibility is that the new entrant will operate an open access wireless network, dramatically enhancing retail competition. The Singapore

market is especially well-suited for such an innovation. First, its existing fixed broadband market operates under the open access model, which is delivering best-in-world Internet connection speeds at competitive prices. Second, Singapore is a small City State, which allows a rapid economic build out in 2017. And third, the dense population means that efficient use of the spectrum through congestion pricing is apt to have large gains.

Independent System Operator

One of the key entities in the open access market is the Independent System Operator (ISO). The term ISO is used in the US electricity markets to refer to the centralized authority managing the wholesale electricity market (FERC 2002). We use the term here in a similar manner as the entity that manages the forward and real-time open access markets. The ISO acts as a trusted party, ensuring transparency, fairness and efficiency. The ISO is independent of both the demand and supply side of the open access network. The ISO is guided by a mission, such as: "To serve the public by operating a reliable and efficient market for open access mobile communications."

The idea of some central authority managing access to resources in the communications world has garnered support in different spheres in the last number of years. The concept of the TV white spaces database to manage access to shared spectrum in the TV bands, and the Spectrum Access System (SAS) used to manage different flavors of access to shared Federal spectrum, are two examples in which some centralized authority acts as gatekeeper. There are also many references to entities such as coordinated band managers in the dynamic spectrum access literature (Zhao and Sadler 2007). The concept of an ISO managing access to network capacity will therefore not be an alien concept.

An ISO serves a specific geographical area. In the electricity market in the US for example there are several ISOs. Some cover a single state, such as ERCOT in Texas, others cover many states, such as ISO-NE, MISO, and PJM. For most countries, a single ISO makes sense for mobile communications.

Service providers rely on the ISO to make sure that users receive the mobile communications contracted for in the forward and hourly markets. Network operators require a reliable means of offering network services. Both sides of the market require proper financial settlement.

A primary function of the ISO is to serve as an exchange where available resources of one or more network operators can be bought and sold. The yearly auction enables service providers to acquire a base level of capacity consistent with their estimated throughput demands. Then in the monthly and hourly auctions the service providers can buy or sell capacity to better match more recent estimates of demand.

The complete list of the ISO duties are as follows:

1. Qualifies market participants and establishes any limits on each participant's bidding activities.
2. Reveals the yearly, monthly and hourly supply curves for the open access network.
3. Conducts the yearly auction.
4. Conducts the monthly auctions.
5. Conducts the real-time auctions.
6. Operates the open access market.

7. Settles all transactions on a monthly basis consistent with market rules and supply and demand realizations.
8. Provides information on market performance to market participants and the market monitor.
9. Improves the market as problems are identified.

Many of the ISO's tasks are highly automated. System development and maintenance is a critical function. The ISO operates with a high level of transparency. Market improvements are discussed and developed with participation of stakeholders.

Governance

Good governance of any public-private partnership is an essential driver of long-run success. Fortunately, there are examples of best practice from related markets. One of the blueprints for governance that we find especially attractive is the one applied in Texas' electricity market. The Texas electricity market has succeeded for twenty years at encouraging both wholesale and retail competition in a reliable electricity market.[5]

There are several layers of governance. Mobile communications is a regulated industry, so a government agency or commission would have primary jurisdiction over the ISO's activities. This would be the communications regulator of the jurisdiction, for example the Federal Communications Commission in the US or the Instituto Federal de Telecommunicaciones in Mexico. The ISO is governed by a board of directors made up of independent members as well as representatives from both sides of the market. The board appoints the ISO's officers to manage day-to-day operations consistent with the ISO's mission. A technical advisory committee makes policy recommendations to the board of directors. The committee has broad stakeholder representation and is assisted by workgroups focused on technical matters. Finally, an independent market monitor reports regularly to the board. The market monitor is independent of ISO management. The job of the market monitor is to observe the market, identify problems, and suggest improvements. Importantly, the market monitor is not a judge or decision-maker; the market monitor is an independent expert voice. This allows the market monitor to react to problems more quickly than the regulator.

Development of the market over time

As with all complex, innovative markets, designing the market is not a one-time task, but a process. Indeed, it is best to adopt a market design that not only works well on day-one but is structured to improve as we learn more and technology evolves. This process approach is highlighted by the name of a key underlying technology: long-term evolution (LTE). The same process approach is needed in the wireless market design. Drivers for this process approach include the ISO management team, constantly seeking to improve the market, the market monitor—looking for problems and suggesting solutions—and the board, making sure that the ISO and market monitor are doing a good job.

[5] Disclosure: Peter Cramton is an Independent Director on the Board of the Electric Reliability Council of Texas (ERCOT).

We can anticipate some aspects of the market evolution. Handset and network technology will continue to improve as we move from 4G to 5G and beyond. Devices will become more powerful and flexible and there will be an ever expanding set of devices as the Internet of Things takes off. On the network side, there will be a greater ability to control demand in ever shrinking time intervals, both by service providers and the ISO. This will allow supply and demand to be balanced in shorter time intervals and smaller geographic areas. Electricity markets experienced this same transition. Initial pricing was hourly with large pricing zones. Most electricity market in the US have since moved to nodal pricing in five-minute time intervals. The open access market is apt to follow a similar path. New technologies will allow even greater granularity in the real-time market.

Spectrum auctions and competition policy

Governments will continue to use spectrum auctions to allocate spectrum rights on a long-term basis. One critical element of these auctions is the competition policy adopted to assure robust competition in mobile communications. Current policy puts the regulator in a difficult position in every spectrum auction. Is there sufficient competition? How best can the regulator enhance competition? Cramton et al. (2011) discuss these challenging issues. An open access wireless market provides a powerful means for the regulator to address competition issues. First, the open access market fully addresses competition in the retail market. Second, the open access market provides detailed pricing data that provides both regulators and investors information about where network development is most required.

Should the regulator identify a competition problem at the carrier level, then spectrum can be reserved for the open access network. This may be desirable to guarantee that the open access network has sufficient network throughput to provide its competitive benefits.

Reserving spectrum for open access wireless still allows the public to receive the value of the spectrum resource. However, rather than taking the form of a lump-sum payment to the Treasury, the value is received as a return over time stemming from competitive congestion prices. As an example, suppose a country conducts a spectrum auction for a band of spectrum with some blocks reserved for open access. The pricing of the non-reserved blocks can be used to establish the investment in network capacity coming from the spectrum. Suppose it is $4 billion and that there is an additional investment of $6 billion for network infrastructure. In this case, network revenues would be split 40% to spectrum owners (the public) and 60% to infrastructure investors. The public still enjoys the returns from the spectrum, but these returns are determined from the ongoing sale of network capacity in the open access market.

Conclusion

Today's markets for mobile communications suffer from an outdated regulatory model that creates an oligopoly of network operators. Entry is nearly impossible as a potential service provider would need to acquire spectrum at spectrum auctions held once or twice a decade and then build a dedicated network. In the US, the cost of such an investment would be many tens of billions of dollars. As a result, successful new entry does not occur and indeed over the last fifteen years there has been consolidation in mobile communications, further reducing competition.

Mobile communications need not suffer from a lack of competition. Restructuring the market to allow for open access is all that is required. Mexico already has proposed such an approach. The open access network is a public-private partnership that is quite similar to the restructured electricity markets that have been in operation for twenty years. Service providers compete for network resources in a wholesale market, and then sell those resources as mobile communications to end users in a retail market.

An open access market is also analogous to the Internet ecosystem where entry is easy. The end result is a high level of competition and innovation, which not only benefits consumers, but the entire communications industry. Further benefits come from the complementary growth and innovation in other sectors stemming from competitive mobile communications.

Early markets for mobile communications required the traditional oligopoly structure as a result of limitations of devices and networks. However, these limitations are no longer present. Today flexible communications devices can be produced by the billions without sacrificing cost, weight or battery-life. Further, LTE's network architecture is much better suited to disaggregating and reassigning network resources than earlier generations of cellular technology. LTE's inherent qualities can now be leveraged to create time and locational markets for network resources.

The introduction of an open access wireless market is inevitable. Its benefits are simply too great to be ignored. The only question is which countries will take the regulatory steps to embrace open access. The economic gains to such first-movers will be staggering, as mobile communications transition to a competitive and highly innovative structure.

References

Allaz, Blaise and Jean-Luc Vila (1993), "Cournot Competition, Forward Markets and Efficiency," *Journal of Economic Theory*, 59: 1-16.

Ausubel, Lawrence M. and Peter Cramton (2004), "Auctioning Many Divisible Goods," *Journal of the European Economic Association*, 2, 480-493, April-May.

Ausubel, Lawrence M. and Peter Cramton (2010), "Virtual Power Plant Auctions," *Utilities Policy*, 18, 201-208, 2010

Ausubel, Lawrence M., Peter Cramton, Marek Pycia, Marzena Rostek, and Marek Weretka (2014), "Demand Reduction and Inefficiency in Multi-Unit Auctions," *Review of Economic Studies*, 81:4, 1366-1400.

Berry, Randall, Michael Honig, and Rakesh Vohra (2010), "Spectrum markets: Motivation, challenges, implications," *IEEE Communication Magazine*, 48, November, 146-155.

Budish, Eric, Peter Cramton, and John Shim (2015), "The High-Frequency Trading Arms Race: Frequent Batch Auctions as a Market Design Response," *Quarterly Journal of Economics*, 130:4, 1547–1621, November.

Cisco (2015), "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2014–2019," White Paper, February.

Cramton, Peter (2013), "Spectrum Auction Design," *Review of Industrial Organization*, 42:2, 161-190, March.

Cramton, Peter (2015), "Market Design in Energy and Communications," *OEDC Policy Brief*, April.

Cramton, Peter, Evan Kwerel, Gregory Rosston, and Andrzej Skrzypacz (2011), "Using Spectrum Auctions to Enhance Competition in Wireless Services," *Journal of Law and Economics*, 54, 2011.

Federal Energy Regulatory Commission (2002), "Working Paper on Standardized Transmission Service and Wholesale Electric Market Design," March.

Joskow, Paul and Jean Tirole (2006), "Retail Electricity Competition," *Rand Journal of Economics*, 37(4), 799-815.

Joskow, Paul and Jean Tirole (2007), "Reliability and Competitive Electricity Markets," *Rand Journal of Economics*, 38(1), 60-84.

Klemperer, Paul (2004), *Auctions: Theory and Practice*, Princeton University Press.

MacKie-Mason, Jeffrey K. and Hal R. Varian (1994), "Pricing the Internet," in B. Kahin and J. Keller, eds., *Public Access to the Internet*, Englewood Cliffs, NJ: Prentice-Hall.

MacKie-Mason, Jeffrey K. and Hal R. Varian (1995), "Pricing Congestible Network Resources," *IEEE Journal on Selected Areas of Communications*, 13:7, 1141-1149.

Milgrom, Paul (2004), *Putting Auction Theory to Work*, Cambridge: Cambridge University Press.

Noam, Eli (1998), "Spectrum Auctions: Yesterday's Heresy, Today's Orthodoxy, Tomorrow's Anachronism. Taking the Next Step to Open Spectrum Access, *Journal of Law and Economics*, 41:S2, 765-790.

O'Connor, Phillip R. and Erin M. O'Connell-Diaz (2015), "Evolution of the Revolution: The Sustained Success of Retail Electricity Competition," *COMPETE Discussion Paper*.

PCAST (2012), "Realizing the Full Potential Of Government-Held Spectrum to Spur Economic Growth," Report to the President, July.

Stoft, Steven (2002), *Power System Economics*, IEEE Press.

Zhao, Qing and B.M. Sadler (2007), "A Survey of Dynamic Spectrum Access," *Signal Processing Magazine*, IEEE, 24:3, 79-89.