Markets in Road Use: Eliminating Congestion through Scheduling, Routing, and Real-Time Road Pricing

Peter Cramton and R. Richard Geddes[1]

1 March 2017

[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, runway access, and other products. Rick Geddes is Professor in the Department of Policy Analysis and Management and Director of the Program in Infrastructure Policy at Cornell University; his research studies private infrastructure investment through public-private partnerships, corporate governance, and regulation. We have benefitted from helpful discussions with Ken Fields, Declan Ganley, Hector Lopez, Bob Poole, Yip Chun Seng, and Leong Wai Yan. We are grateful to Arpent for funding this research. The views expressed are our own.

Abstract

Traffic congestion is a global problem with annual costs approaching $1 trillion. The cost of traffic congestion across the combined British, French, German and American economies was estimated at $200 billion, or about 0.8 percent of GDP, in 2013. In Los Angeles alone, traffic jams cost $23 billion each year. The health and environmental costs are severe in urban centers worldwide.

With the right policies those high social costs can be avoided. Advances in mobile communications and computer technology now make it possible to efficiently schedule, route, and price the use of roads. Efficient real-time pricing of road use can eliminate traffic congestion, enhance safety, improve the environment, and increase vehicle throughput. It also raises reliable, much-needed revenue to modernize decaying infrastructure while improving the allocation of transportation investment. We describe the design of a market for road use and transportation that is based on efficient scheduling, routing, and pricing. Under our design, road use is priced dynamically by marginal demand during constrained times and locations. In unconstrained times and locations, a nominal fee is paid for road use to recover costs, as in other utilities. Transport is scheduled based on forward prices and then routed in real time based on real-time road-use prices.

Efficient pricing of network capacity is not new. Indeed, wholesale electricity markets have been dynamically priced for over a decade. Communications markets are adopting dynamic pricing today. Efficient pricing of road use, however, has only recently become feasible. Advances in mobile communications make it possible to identify and communicate the location of a vehicle to within one cubic meter—allowing precise measurement of road use. User preferences can be communicated both in advance to determine scheduled transport and in real time to optimize routes based on the latest information. Computer advances also facilitate efficient scheduling and pricing of road use. Consumer apps help road users translate detailed price information into preferred transport plans. Computers also allow an independent system operator to better model demand and adjust prices to eliminate congestion and maximize the total value of road infrastructure. An independent market monitor, distinct from the operator, observes the market, identifies problems, and suggests solutions. A board governs the market subject to regulatory oversight.

The market objective is to maximize the value of road infrastructure via scheduling, routing, and real-time pricing of its use. The optimization of road use eliminates congestion, making our roads safer, faster, cleaner and more enjoyable to use. The road-use market thus maximizes the value of our existing transport infrastructure while simultaneously providing essential funding for the roads network as well as valuable price information to evaluate road enhancements. The market is highly complementary with and indeed promotes rapid innovation in the transport sector.

Introduction

Traffic congestion is a pervasive and growing worldwide problem. Global congestion costs are estimated at about $1 trillion. In Europe congestion costs are about 1 percent of GDP ($200 billion). The cost of traffic congestion across the combined British, French, German and American economies was estimated at $200 billion, about 0.8 percent of GDP, in 2013 (The Economist 2014). In Los Angeles alone, traffic jams cost $23 billion each year. Two-thirds of those costs stemmed from wasted time and fuel, with the remainder due to increased costs to businesses.

For the United States, traffic congestion wasted 6.9 billion hours of motorists' travel time and almost 3.1 billion gallons of fuel (Schrank, Eisele, Lomax and Bak 2015) in 2014. Moreover, congestion's social costs are growing over time. The U.S. congestion "invoice" for added costs in terms of fuel and time grew from $42 billion in 1982 to about $160 billion in 2014 (in 2014 dollars)—almost a three-fold increase—in the 471 urban areas studied by the Texas Transportation Institute (Schrank, Eisele, Lomax and Bak 2015). Similar increases are occurring worldwide as car ownership rises rapidly with development. Roads are thus failing to perform their core task of safely facilitating vehicle passage while having minimal environmental impact, and the problem is getting worse.

Fortunately, a major transformation in transportation performance will occur over the next twenty years, if not sooner. Complementary technological advances—especially in computers and communications—will greatly reduce congestion, thus improving mobility. We explain below how to harness existing technological developments which, when combined with advanced markets in road use, can fully eradicate traffic congestion along with its attendant social harms.

Those improvements in transportation performance can occur now. Today's technology allows for monitoring a vehicle's road use and charging directly for that use. Real-time charging for road use allows accurate pricing of congestion. With the right prices to guide behavior, traffic congestion can be eliminated. Accurate road-use pricing not only ensures that the right vehicles are in the right lanes at the right times, but dramatically increases the number of vehicles that can flow through the available lanes (i.e. “throughput”) during peak periods. Traffic jams are thus avoided.

Proper road-use pricing has a number of additional benefits. It improves environmental quality since vehicles operate more efficiently, spend less time idling, while pollution costs are recognized explicitly via road prices. Safety improves due to more consistent, predictable traffic flow. Perhaps as importantly, real-time road prices provide essential information to identify what investments are needed while generating the funds necessary for that investment.

Wholesale electricity markets offer a useful analogy.[2] Those markets have operated on a similar basis for over a decade (Cramton 2017). Electricity resources are optimally scheduled and priced one day ahead, after which a real-time market allows for necessary adjustments throughout the day. In transportation, crude versions of the pricing we envision are observed in certain express lanes, in Uber’s surge pricing, and in airlines’ pricing of seats, among other examples. Our approach goes far beyond these early versions, however, building on the proven success of electricity markets. Under our approach, transport is scheduled and priced in advance with real-time routing and pricing to reflect inevitable changes in demand.

In addition to mitigating traffic congestion, dynamic road-use pricing creates a non-distortionary revenue source for infrastructure operation and maintenance. The resulting

[2] A second analogy is in wireless communications (Cramton and Doyle 2016), but time-and-locational markets in communications are only beginning to be implemented.

congestion prices also provide essential information to direct scarce investment resources toward projects where those dollars are most highly valued by road users and away from lower-valued uses. Dynamic pricing thus creates the crucial link between customers' value of a facility and investment flows. Stated differently, congestion pricing provides objective market signals regarding where additional investment should be directed that are based on users' willingness to pay for that investment. Our proposal thus also addresses one of the most challenging problems facing transportation policy today: the perceived misdirection of scarce public dollars caused by politicization of spending.

Price information alone is not sufficient to ensure that the best investment choices are made, however. Good governance and oversight of the planning and investment process is also critical. Yet price information remains essential for expert planners guided by an objective of maximizing social welfare. This approach has worked well in restructured electricity markets, especially the successful market design used in most of North America (Cramton 2017), which prices energy every 5 minutes at every location. In electricity, regulatory capture by special interests is reduced or largely avoided via good governance and a data-driven analytical planning process.

Although these are the core benefits, congestion pricing is appealing for several additional reasons. Those include: (i) generating sustainable, long-term transportation system funding; (ii) divorcing charges for the use of road space from fuel type used, which makes road charges independent of rapidly evolving engine technology; (iii) adopting the basic horizontal-fairness principle that the motorists using a road should pay for them, which enhances social equity; (iv) allowing scarce road space to be allocated to motorists who value it most highly at that particular time of day; and (v) encouraging commuters to explore the travel alternatives of their choice during peak times by providing current road-use prices. Although partial, this list suggests that the social benefits created by dynamic road-use pricing are substantial. Moreover, these benefits are apt to be shared broadly. All share in the health and environmental benefits of eliminating congestion, and nearly all benefit from improved throughput.

Commentators have been aware of road pricing's many benefits for decades. Writing in the early 1950s, Milton Friedman and Daniel J. Boorstin stated that:

> At first glance, it seems hardly possible that this apparently trivial problem of how to charge people for the highway services they use is a key to the whole problem of how to plan and pay for better highways; yet it is just that. This fact cannot be too strongly emphasized. It is a key not only for a system that would involve operation of roads by private enterprise but equally for the present system of public operation. Should a particular road be built? How should it be built? How should it be financed? Should an existing road be maintained, improved, or allowed to deteriorate? If we could charge directly for the service of the road, we could answer those questions—whether under private or public ownership—in the same way that we now decide how many automobiles should be manufactured, what kind of automobiles should be manufactured, how their production should be financed, whether a particular model should be discontinued, and so on. (Friedman and Boorstin 1996, 223)

At the time Friedman and Boorstin were writing, tolls were paid entirely in cash. Widespread tolling implied stopping to pay tolls, thus slowing travel. Massive leaps in electronic and computer technology since then make new charging approaches feasible.

Despite broad academic agreement, use of direct road-use fees in the United States has been limited. Although Oregon debuted a system-wide user-fee program (Morris 2015), road pricing is most often limited to new lanes, such as high-occupancy toll (HOT) lanes or on conversions from high-occupancy vehicle (HOV) to HOT lanes. This has left existing transportation facilities—often older roadways in need of fresh investment—out of the funding streams generated by that pricing. Moreover, only a few small sections of the U.S. highway system are dynamically priced.

Looking abroad, many cities around the world have adopted some form of congestion pricing. Examples include London, Oslo, Stockholm, and Trondheim. However, its application is typically limited to a cordon around the city, or to a small set of roads within. Moreover, prices are typically not responsive in real time. Our proposal represents a significant leap forward to what can be viewed as the most sophisticated end-state for road pricing. We propose a comprehensive system of direct, variable road-use charges. This creates a market for transport that maximizes the value of the transport network through efficient scheduling, routing and pricing of road use.

By scheduling of road usage, we mean an advance plan of when road segments are used and by whom. In some cases, scheduling can occur well in advance of actual usage. In other cases, it may occur one or more hours ahead. Scheduling contrasts with routing in that routing results in a real-time usage plan. Scheduling, however, requires at least some demands to be expressed in advance of real time. Of course, schedules can change in response to unexpected events and new information. Our proposal can handle such changes.

Analysts have attributed the limited use of road-use fees in the United States to motorists' opposition to new rates and fees. Because motorists often think that roads are "already paid for" via gas taxes, dynamic pricing may initially be viewed as a new tax. This suggests that the rebating of state fuel taxes paid when motorists pay usage fees (the approached used in Oregon) is critical. Such issues are particularly important for existing transportation facilities, where user-fee revenues are usually utilized to improve road quality rather than to add additional capacity easily visible to motorists. Many commentators argue that displaying tangible improvements from road pricing greatly enhances motorists' acceptance of user fees (Poole 2014).

A key challenge in implementing a market for road use is coordinating the multiple jurisdictions (e.g., city, county, state, and federal) that may own various adjoining facilities. Although demanding, the large gains from coordination that our approach offers can motivate jurisdictions to work together, especially when regulators properly encourage coordination. Again, this has worked well in U.S. electricity markets where federal and state regulators have played an essential role in facilitating coordination among market participants.

Direct road-use pricing may also raise privacy concerns. Privacy is however readily addressed with a privacy policy that limits who sees what information about a vehicle's road use. These same issues are addressed in other industries, such as communications. Who and when calls are made are recorded and priced, but access to this information is strictly limited. Further, many users have today expressed a high willingness to provide locational information to service providers, such as Waze, Google and Uber.

Enforcement is an added concern. A mechanism to ensure that users pay for road use is critical. There are standard ways of addressing enforcement. For example, license plates could be randomly photographed and fines applied whenever the vehicle's road-use device is not operating. As an aid to the user, the vehicle could provide a warning whenever the road-use device is not engaged. We next discuss traffic congestion and dynamic pricing of road use.

Dynamic pricing and traffic congestion

A movement toward direct road pricing rather than fuel taxes to finance roads is an important step in improving transportation policy. However, its effect on traffic congestion will be limited unless the price charged varies with the cost that one motorist imposes on others when utilizing a particular facility at a particular time of day. Similar to water flowing through a pipe, or a wire carrying electrical waves, a road, bridge, highway, or tunnel has a physical capacity limit. A highway lane can only transport a certain maximum number of vehicles per hour when traffic is flowing freely. When too many motorists try to use a facility at the same time, the facility becomes congested, like a clogged water pipe. Traffic flow collapses, and the bridge, highway, or tunnel is unable to handle as many vehicles as its physical capacity allows.

When drivers decide to use their vehicles, they generate various costs. Those include fuel cost, depreciation of their vehicles, time spent driving, and wear and tear on transportation facilities, as well as the crowding out of other motorists using the facilities at the same time. Drivers bear many of those costs directly, and will thus take them into account in their decisions about when, how often, and how far to drive. They will, for example, directly bear fuel costs, vehicle depreciation, as well as time costs.

They will not, however, directly bear the costs associated with facility wear and tear or the crowding out of other motorists who also want to drive at that time, and will not take them into account in their driving choices. Their decisions will thus not be aligned with the true overall cost of using particular facilities at particular times of day. Facility wear-and-tear costs can easily be addressed through a nominal fee for road use.

That second component, however, is more complex, since motorist crowding has highly non-linear effects on travel time. That is, as a facility gets close to its physical limits, even a small increase in the number of vehicles leads to a large drop in average speed for all vehicles. If one additional motorist tries to use the road at those peak times, she imposes a large cost on all other motorists through slower travel times. For example, the motorist who decides to use a highway at 3:00 AM imposes no crowding or congestion costs on other motorists, since there is usually excess road space. One who instead uses the same highway at 8:00 AM, when many other motorists also want to use it, imposes substantial congestion costs.

Under the current approach, the individual motorist does not consider those large social crowding costs. The second key aspect of road pricing is therefore a variable charge, or congestion price, that reflects the costs of social crowding. The phenomenon described above suggests the effectiveness of such charges. If even a small number of motorists can, via congestion prices, be encouraged to drive at other times, to use alternative modes of travel, to carpool, to telecommute, or to adjust in any number of other ways, then traffic flow will rise disproportionately for the remaining motorists who choose to use the facility at that time. Reducing the number of drivers by as little as 5 percent at peak times may enable traffic to flow smoothly, allowing the same facility to handle many more vehicles. Variable road prices therefore have the effect of allocating scarce road space at peak times to those who value it most highly. The highest-valuing motorists will choose to use it at those times and pay the associated higher fee.

Available transportation capacity is currently allocated by queuing. Queuing is an especially wasteful allocation method in transport, because the queue degrades throughput. Queuing ignores the cost that one motorist imposes on others in trying to use a facility at peak times, as well as the differing values motorists place on the road's use at particular times.

Because congestion prices keep traffic moving smoothly, travel times also become more predictable when road space is accurately priced. This is critical to parents, for example, who need to pick their children up (or drop them off) from school or day care at specific times, as well as to companies relying on just-in-time inventory techniques. It also reduces the time wasted in planning for possible congestion, or in leaving a time cushion to allow for travel uncertainty. Furthermore, congestion prices help ensure that facilities are used more evenly throughout the day.

Current use of congestion pricing

Congestion pricing is not a novel concept. As noted, similar variable charges have been successfully utilized in many other industries. For example, airfares, cell phone rates, electricity rates, room rates at hotels and resorts, train fares, and some local transit systems use variable pricing.

Congestion pricing has also been used successfully on a small number of U.S. roads. It is currently used in Minneapolis on the I-394 MnPass Express Lanes, which are dynamically priced in real time. It is used on the I-15 FasTrak Lanes in San Diego, where prices are updated every six minutes, and on the I-25 express lanes in Denver. It is also used on the SR-91 Express Lanes in Orange County, California, where the price varies between $1.15 and $9.25 per trip and is posted prior to entry so motorists can choose between priced and non-priced lanes. Priced lanes are popular because they save substantial amounts of time. The Oregon pilot program mentioned earlier indicated that variable pricing could be incorporated into an overall pricing approach.

Many international examples of successful use of congestion pricing are also available. In 1975, Singapore became the first city to implement it successfully for urban traffic. Under this approach, called the Area Licensing Scheme, cars were charged an additional fee to enter the central business district between 7:30AM and 9:30AM. This form of congestion pricing is known as cordon or central area pricing. It was strikingly successful, resulting in a 73 percent decrease in the use of private cars, a doubling of bus usage, and a 30 percent increase in carpooling. In 1995, congestion pricing was extended to three of Singapore's major freeways. On one freeway, average speed during the morning peak increased from 31 to 67 kilometers per hour.

In addition to the demand-side benefit of helping to manage traffic flow, variable pricing creates another, supply-side, benefit: it provides information on how much motorists value the use of particular facilities and thus reveals the most valuable use of the marginal investment dollar. Prices reveal value, and the congestion price required to smooth traffic flows is a reflection of how much value motorists place on a particular facility at that time. Relatively high prices suggest that motorists place a high value on using that facility during peak times. The social returns to expanding the road, bridge, or tunnel will thus also be high, and added investment should be directed there. By providing an observable, objective indication of where system expansion should or should not take place, congestion prices also help depoliticize transportation investment, making it more efficient. A consensus has emerged that tolling (and, importantly, public-private partnerships) can provide critical information on where investment should take place and thus reduce political influences in transportation spending (Geddes 2011).

If road users are prepared to pay a price for road use greater than the costs of providing additional road space (including all costs, externalities etc.) then the additional road space should be built. As in any other economic activity, charges for the use of the new facility should be sufficient to finance its cost. In short, road pricing and congestion pricing would yield important benefits on both the demand and supply sides of the transport sector.

New technologies and dynamic road pricing

Technological developments in electronics and communications facilitate accurate pricing of the use of road capacity. Currently available technologies allow a road segment to be priced to the sub-meter level. Moreover, road use can be priced in real time based on current conditions of scarcity. Data on the current price of road capacity as established by user demands and measured externalities can be fed into a vehicle's on-board computer, giving directions to the driver in real time on what route to take. For autonomous vehicles, real-time road price data can direct the vehicle's route decisions without driver intervention. In principle, real-time adjustment by vehicles in response to highly granular road pricing data has the potential to eliminate all traffic congestion.

Such approaches have been successful in other contexts. The most analogous is the real-time pricing of wholesale electricity. In restructured markets in the United States, electricity prices at each time and location reflect the efficient congestion prices. Electricity supply and demand is scheduled to maximize gains from trade subject to all physical constraints. The electricity is produced at least cost and consumed by those valuing it the most. The analogous outcome for road usage is efficiently scheduled transportation. Roads would be utilized so as to provide maximum throughput, priced high enough to reduce demand to the efficient (uncongested) level.

A key principle of the electricity market is open access: the transmission grid is available to all on non-discriminatory terms. The open-access market substantially enhances competition and efficiency in electricity. Those same benefits could accrue in the pricing of road capacity. With open access, transmission capacity cannot be withheld—all available capacity must be made available for use. This has profound implications for pricing—spot-market pricing reflects congestion. The congestion price is zero at times and locations without congestion. However, on constrained lines the prices balance supply and demand and assure that all transmission constraints are satisfied. This is called *locational marginal pricing*. 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 that includes future network investments.

Since real-time congestion prices tend to be volatile, electricity market participants have a desire to manage risks. Forward auctions, conducted in advance of real time, allow participants to make plans and lock in prices consistent with their anticipated needs. The market also allows traders to arbitrage across related products—such as yearly, monthly, and spot products in the same area—to improve price signals and resource allocation. In modern electricity markets, the vast majority of energy trades in advance of the real-time market. The forward markets enable planning and hedging of risks, while the real-time market sends the precise price signal to efficiently manage congestion. With road use, more volume is apt to trade in real-time, since individual demands are difficult to predict. Forward markets may still play a useful role for those with more predictable demands.

Alternative market designs for road use

Congestion pricing for road use is apt to be introduced gradually as the technology evolves and opportunities for implementation are identified. In the simplest application, congestion pricing may be limited to key bottlenecks and new express lanes. Further, pricing may be limited to too narrow a range and perhaps adjust too slowly to fully relieve congestion. Nonetheless, such steps will reduce congestion and are likely to build support for more significant congestion management.

Additional steps would expand the number of roads with congestion management and relax the constraints on pricing. That will generate further improvements in congestion pricing and reduced congestion. However, the system would remain one in which a system operator is adjusting real-time prices to limit congestion, rather than a full optimization and scheduling of transport.

The final step forward involves both full optimization and scheduling of transport. The system operator would receive preferences from all vehicles for road use at alternative times, such as $10 at the most preferred time to travel from A to B and lower prices at less desirable times. The system operator then optimizes the expressed demands with the available supply to identify the assignment of road use that maximizes the value of road use. The optimization also identifies congestion prices that support the efficient assignment. Although this is a complex optimization, the problem is made easier by the limited number of bottlenecks and the large number of vehicles. The latter reduces the importance of integer constraints in the optimization problem making it more likely that efficient congestion prices exist. Another helpful factor is that aggregate traffic patterns tend to follow a predictable cycle.

One important detail is the timing of the optimization. The state of the system, such as levels of use and availability, and user preferences are constantly changing. The initial optimization needs to be done when preferences are relatively complete, not too far from real time, and then re-optimized as circumstances change, say in response to a lane loss.

The evolution of the road-use market from primitive to advanced is both desirable and inevitable. This evolution has been observed in other industries. The electricity market may again be the best example. Early markets of about twenty years ago did a good job of facilitating trade across locations, but initially ignored important factors such as congestion pricing. This shortcoming was resolved with the introduction of locational marginal pricing. Many other shortcomings have been addressed both in response to better market rules and improved technology. The result has been a steady and significant improvement in electricity markets over the last twenty years.

A key insight from experience with electricity markets is that it is important to adopt as good a market design as possible initially, but to also ensure that issues of governance and management are such that there are strong incentives for constant improvement of the market over time. In electricity markets, scheduling of generating resources plays an important role. This is because many resources are large and limited in how quickly they can respond—both time to start/stop and ability to ramp up/down. As a result, in electricity, the optimal scheduling of resources was introduced early to the market.

Efficient scheduling may be much less important in road use because of the large number of vehicles and the speed with which marginal vehicles can respond to price information. This suggests that a useful market simplification, at least initially, may be to focus entirely on efficient congestion prices and to initially ignore scheduling of transport. We do that in the initial market design discussed below. In the subsequent section we extend the market design to include forward purchase and scheduling. This allows users to purchase in advance the right to use particular road segments at future times.

Congestion pricing of road use

We next present a market model for road use based entirely on efficient congestion pricing. Operation of the transport system is centralized, although users' decisions regarding road use are decentralized. Transport is here not scheduled. Vehicles and drivers simply respond to congestion prices that are set to efficiently manage congestion.

The market is conducted by an independent system operator (ISO). The ISO's mission is to maximize the value of scarce existing road resources. The key instrument available to the ISO is the ability to set efficient congestion prices. The ISO models demand for road use and establishes prices at each congested segment and at each time in order to minimize congestion. Usage is monitored and charged to each user based on the marginal cost of segment use including congestion costs.

Demand modeling is critical to establishing efficient prices. Users do not directly express how they substitute across alternative travel times. They instead simply select their most preferred option given the current prices and expectations of future prices. This complicates the ISO's modeling, but the ISO will have abundant data on transport choices with which to adjust prices. When prices are set too low, congestion forms. When prices are set too high, the segment is underutilized. System uncertainties mean that the ISO cannot operate the road at 100 percent of capacity. Some capacity is reserved to handle momentary surges in demand or drops in supply.[3]

System modeling and management is a complex engineering problem. Nonetheless, the ISO likely can address this challenge well. The problem is well-studied in both economics and engineering and there have been large computational and algorithmic advances in recent years. Most importantly, existing mobile communications allows a vehicle's road use to be precisely monitored, even down to the lane of travel.

A second major challenge is consumer acceptance. Consumers must be able to easily use and understand the system and to realize large benefits. The status quo approach creates much frustration and delay. The reduction or elimination of delay and frustration is thus likely to be the source of consumer gains that will breed acceptance. But ease of use is also critical to consumer acceptance. Sophisticated apps are needed to translate a consumer's preferences and the price

[3] Drops in supply are likely to be a larger problem in the electricity and other sectors than in roads but remain a concern due to road construction, accidents, and other supply-side events.

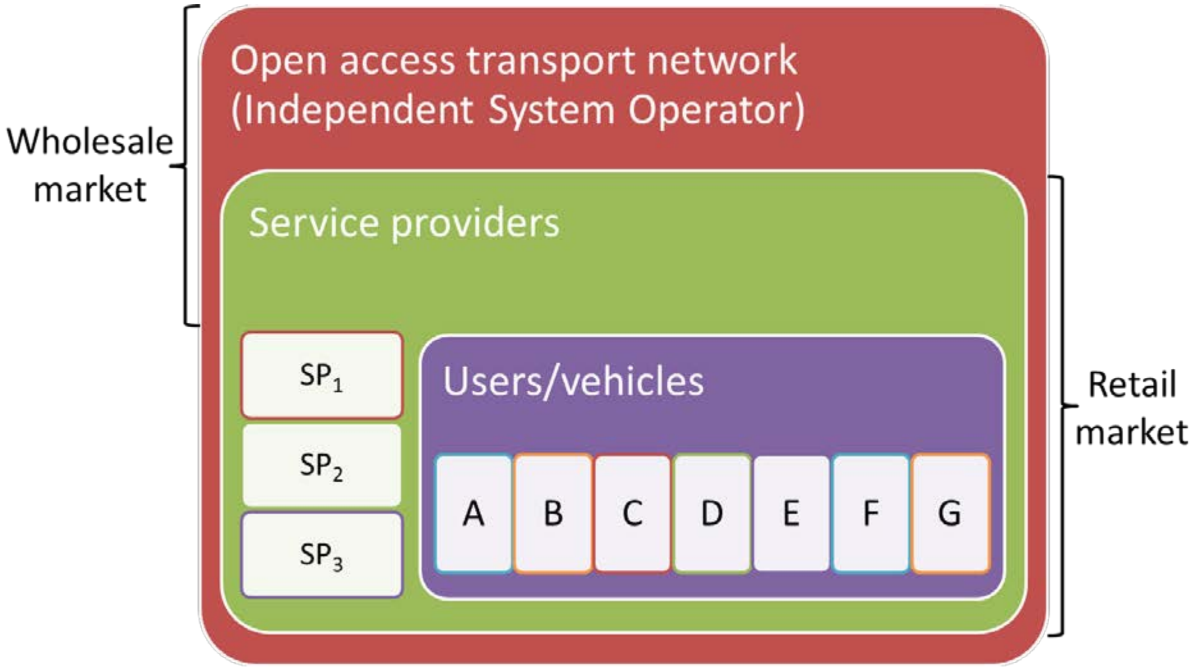
information into a recommended transport strategy or menu of choices. All of this is possible with existing technology. Indeed, most users already have experience with sophisticated apps to facilitate transport. Google Maps, Waze, Uber, and Lyft are examples. Adding road-use prices to those apps is straightforward. These apps already offer users a menu of choices. With road-use prices included, the menu would be expanded. We can expect rapid innovation in the apps that help users make better and easier transport decisions.

A market design focusing solely on congestion pricing is a good starting point. The market may eventually shift to a more complex system that both *prices and schedules* transport system-wide. However, congestion pricing alone is much simpler and likely to go a long way toward maximizing the value of road use.

A wholesale market for transport

We next discuss an alternative market design that is built on the wholesale/retail market model. This is the market model successfully used in electricity markets for over a decade (O'Connor, et al. 2015). The independent system operator has the same mission—to maximize the value of road use—but does so by operating a wholesale market in which service providers (e.g., Google, Uber, and Lyft) compete for road use in forward markets as well as in real time, aggregating the demands of individual users. Wholesale pricing is determined in frequent auctions. The market model is shown in Figure 1.

Figure 1: Wholesale/retail market model



Users provide the fundamental demand for road use. Service providers compete for users in the retail market. Providers that offer more attractive plans are likely to be more successful. Some large companies, such as United Parcel Service, would participate directly in the wholesale market.

An advantage of the wholesale market model is that entry at the service-provider level is relatively easy. This fosters competition and innovation, which is desirable given the important role played by service providers. In the wholesale market, service providers aggregate user demand. To do this, the service provider must develop a user app that enables users to easily and effectively express demand. The service provider also guides the user, both in scheduling future demand as well as routing during real time. Finally, the service provider establishes user plans and settles payment. We expect a great deal of innovation to occur in service provision.

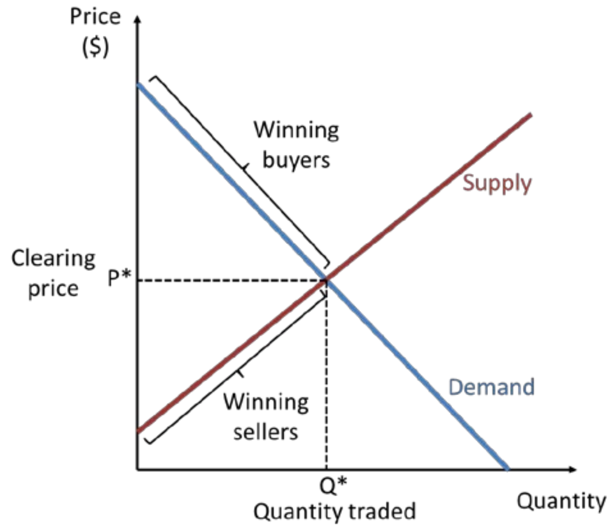
The wholesale market allows a relatively simple product design. The product is a slot on a congested road segment at a particular time. That is, the "commodity" being traded is the right to use a well-defined section of road for a particular time slot. Time is broken into discrete intervals, such as 10 minutes, to keep the number of products manageable.

There are three promising features of the market. First, the number of congested road segments is limited. The most obvious congested segments are bridges, tunnels, and other bottlenecks. Second, the congested segments are highly predictable. Rush hours are a good example. Third, demand does respond to price, in close to real time. The demand response takes one of four forms: (1) time shifters, who shift transport to a less congested time; (2) route shifters, who shift to a less congested route; (3) mode shifters, who decide to take a train, bus, or bike, rather than drive; and (4) curtailers, who decide to ride-share, work at home or otherwise reduce demand.

In one sense, responding to prices in transport is nothing new. Today one “price” is delay cost. Users do respond to this price, but the price is waste and is set incorrectly—the delay cost does not reflect the negative externality one user imposes on others. In the market model, congestion is eliminated—the real-time price is set at the marginal value of demand at the point of supply and demand balance.

The foundation of the transport market is the real-time market, pricing road use in real time, for example every ten minutes. The real-time market is a physical market, based on actual (i.e. physical) road use. One challenge of the real-time road-use market is that prices can be volatile as it becomes more difficult for demand to respond to events, such as a lane loss, in a short period of time. User apps can redirect traffic given preferences and prices, but the response is limited.

Figure 2: Single-price auction model



Each product is traded in a single-price auction, as shown in Figure 2. Bidders express demand schedules, which indicate the quantity demanded at each price. The ISO forms the aggregate demand curve and crosses it with supply to find the clearing price (P*) and quantity (Q*) where supply and demand balance. All trade takes place at the market-clearing price. Demand schedules are expressed via a series of price-quantity pairs to form a weakly decreasing demand curve. Identifying equilibrium prices and quantities is complicated, because user demand is expressed across multiple interrelated products: the value of transport from A to B at various times. Fortunately, we have seen tremendous advances in our ability to find value-maximizing assignments and prices in this setting. Wholesale electricity markets are a leading example. Large optimizations are solved in real time to determine the least-cost dispatch of energy in a network with thousands of nodes and thousands of constraints.

To mitigate risk and promote efficient scheduling, it is important for the market to provide multiple trading opportunities. We envision three forward markets: yearly, monthly, and daily. The ISO determines the supply offered in each forward market consistent with service providers' interest in taking forward positions. The ISO offers more supply at higher prices—an increasing supply curve.

Service providers then bid in the forward and real-time markets to maximize net value to users and manage risk. Typically, this involves purchasing some fraction of user demand in each of the markets and to make adjustments to positions as uncertainty about demand is resolved.

The forward markets are financial, not physical. Service providers take positions in the forward markets, while subsequent markets allow adjustment of positions as uncertainty is resolved. Speculators also participate in the forward markets, arbitraging across forward markets. Profitable speculators improve price efficiency in forward markets before covering positions in the real-time market.

In the yearly auction, service providers estimate demand for the coming year and bid in the auction based on this demand and expectation of future prices. In the monthly auction, service providers now have better information about demand and can make adjustments to their current positions. In the daily auction, further uncertainty about demand has been resolved. Service providers can further adjust positions. The real-time auction occurs shortly before physical consumption.

Having multiple opportunities to trade reduces risk to the service provider and facilitates planning. Forward markets also facilitate price discovery through price transparency. Finally, the forward trading mitigates market power in the real-time market by putting service providers in a more balanced position entering the real-time market.

The further forward markets (yearly and monthly) offer more aggregated products to promote liquidity. For example, it would be natural to have a weekday product and a weekend/holiday product for each congested road segment and 10-minute interval. Thus, with the 8AM weekday product, the bidder is purchasing some number of slots on each weekday of the month or year in the 8AM to 8:10AM interval for the specified road segment.

Governance

The independent system operator conducts the market for road use. “Independent” means that the ISO has no ownership interest in the road network and no interest in the congestion revenues collected. The ISO is charged by its board to operate the market to maximize the value of road use. The ISO’s chief instrument to achieve that efficiency goal is congestion pricing, setting prices for road use that mitigate congestion.

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. The market monitor is instead an observer who writes reports and makes recommendations. In electricity markets, this has allowed the market monitor to quickly identify problems and suggest solutions. 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 directors who are affiliated with a particular stakeholder group. The board also has a number of unaffiliated directors who are independent of the stakeholder groups, but bring essential subject-matter expertise. Unaffiliated directors are approved by the regulator.

Opportunities for early implementation

Dynamic pricing of road use can be implemented in any congested area. Both the United Kingdom and the United States are attractive markets for early implementation. Both countries have existing pilots and high adoption rates of the enabling technologies. Singapore stands out as another country for early implementation. Singapore is a city-state with a long history of innovation in transport. Also Singapore has a single transport regulator, the Land and Transport Authority (LTA), so there are no issues of coordination among jurisdictions.

Conclusion

Traffic congestion severely reduces the value users realize from our road infrastructure. Worldwide congestion costs are estimated at about $1 trillion per year. Fortunately, technological advances allow us to largely eliminate congestion through efficient congestion pricing. In the simplest approach, an independent system operator models demand and computes real-time prices for road use to maximize the value of road use. Prices on congested road segments are set to induce demand to eliminate the congestion and maximize throughput. User-friendly computer apps armed with this price information then guide consumers in making transport choices consistent with their preferences.

A more sophisticated market design is based on a wholesale market model, as we see in electricity markets. The advantage of a wholesale market is that it allows relatively easy entry as a service provider. The ensuing competition among service providers then promotes innovation. That innovation helps service providers to better understand user demands, translate user demands into bids in the wholesale market, and develop forward trading strategies to mitigate risk. A complete system of scheduling, routing, and congestion pricing may seem like a radical idea. It, however, has been successfully applied in electricity markets for over a decade, and is increasingly adopted in communications markets.

Modern mobile communications allow road use to be monitored and charged based on real-time scarcity. Doing so gets the most out of our existing transportation infrastructure and simultaneously provides essential funding of the roads as well as valuable price information to evaluate road enhancements. This is the inevitable future of roads.

References

Cramton, Peter (2017) "Electricity Market Design," Working Paper, University of Maryland.

Cramton, Peter and Linda Doyle (2016) "An Open Access Wireless Market," Working Paper, University of Maryland.

Friedman M. and D.J. Boorstin (1996) "How to Plan and Pay for the Safe and Adequate Highways We Need," in Roth, G., *Roads in a Market Economy*, Aldershot: Ashgate Publishing.

Geddes, R. Richard (2011). The Road to Renewal: Private Investment in U.S. Transportation Infrastructure (Washington, DC: AEI Press).

Morris, David Z. (2015) "Oregon Just Debuted This Radical Alternative to the Gasoline Tax." *Fortune*, July 17.

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*.

Poole, Robert W., Jr. (2014) "Value-Added Tolling: A Better Deal for American's Highway Users," Reason Foundation Policy Brief 116.

Schrank, David, Bill Eisele, Tim Lomax, and Jim Bak (2015) *2015 Urban Mobility Scorecard*. Texas Transportation Institute, August.

The Economist (2014) "The Cost of Traffic Jams," November 3 (London: The Economist magazine).