Journal of Industrial and Systems Engineering
Vol. 1, No. 2, pp 97-115
Summer 2007

JISE

Strategies to Overcome Network Congestion in Infrastructure Systems

Jason W. Black, Richard C. Larson*

Center for Engineering Systems Fundamentals; Engineering Systems Division
Massachusetts Institute of Technology; Cambridge, Massachusetts 02139 USA

ABSTRACT

Networked Infrastructure systems deliver services and/or products from point to point along the network. Demand for the services provided by such systems is typically cyclic, creating inefficiencies in capacity utilization. Congestion pricing provides incentives to shift demand from peak time periods to lower demand periods. This effectively increases the capacity of the system without the need for additional capital investment. This paper investigates the potential for congestion pricing to reduce necessary infrastructure investments in the United States. Several types of congestion pricing schemes are presented, along with existing implementations across multiple infrastructure systems. We find over $20 billion in potential annual savings in electricity and road systems alone in the United States from implementing congestion pricing schemes.

Keywords: Networked infrastructure systems, Congestion pricing, Capacity utilization.

1. INTRODUCTION

Networked Infrastructure systems deliver services and/or products from point to point along the network. They include transportation networks (e.g., rails, highways, airports, sea ports), telecommunication networks (by frequency-bounded airwaves or cables), and utilities (e.g., electric power, water, gas, oil, sewage). Each is a fixed capacity system having marked time-of-day, day-of-week and season-of-year patterns of demand. Usually, the statistics of demand, including hourly patterns (i.e., means and variances) are well known and often correlated with outside factors such as weather (short term) and the general economy (longer term).

An infrastructure system is typically difficult and expensive to design and construct. Once built, it can have a mean lifetime from 20 years (telecommunications) to over 100 years (water). As population and the economy grow, increasingly large demands are being placed on infrastructure systems. Eventually they must be upgraded due to lack of adequate capacity and/or the need for improved technology. However, that moment can be delayed, often for long periods, by the use of congestion pricing to reduce peak demand. Congestion pricing provides incentives to shift demand from peak time periods to lower demand periods. We call this, "shaving the peaks and filling in the valleys." Such shifted demand effectively increases the capacity of the system without the need for additional investment.

* Corresponding Author

Current examples of congestion pricing schemes include: time of day congestion pricing for autos in Singapore and London; for-profit 'toll-ways' adjacent to freeways; time of day pricing for electricity; time of day pricing for long distance telephone calls; revenue management in airlines to balance out travel demands over the course of a week and over the year; and auction type bidding for some infrastructure services, with higher prices paid for congestion periods.

In this paper we investigate congestion pricing across critical infrastructures in terms of the potential benefits of forgone investment achieved by reducing peak demand. We first analyze various methodologies for implementing congestion type pricing, and then present several existing implementations of congestion pricing. We then look at the political and economic impediments to widespread adoption of such pricing schemes. Finally, we suggest areas of future research to develop congestion pricing strategies that provide efficiency gains and are politically acceptable and amenable to implementation across infrastructure domains.

With regard to systems engineering more generally, we believe this area to be a fertile one for research and development in the years to come. Our own approach is now called “Engineering Systems,” born at MIT in the Engineering Systems Division. A recurring theme of this broadened approach in contrast to more traditional systems engineering is this: Engineering Systems examines problems at the Venn diagram intersection of (traditional) engineering, management and social sciences. The Engineering Systems approach is especially important for large, complex socio-technical systems, such as critical infrastructures. In many instances in system design and analysis, either management or social sciences may dominate traditional narrowly technical engineering issues.

2. BACKGROUND

Most service industries have predictable cyclic patterns of demand for their services, with demand peaks and demand valleys. The cycles usually have multiple frequency components: daily, weekly and seasonal. To satisfy the peaks, capacity needs to be expanded to meet peak demands. At other times, the infrastructure supporting the service may sit idle. Demand management has the potential to greatly increase capacity utilization and to forestall additional capacity investments in service industries.

Some industries, like airlines, are now quite advanced in their use of demand management whereas others, like public utilities, are far behind the state of the art. Demand management was invented by the airlines in the early 1980’s. At that time, it was known as *yield management*, which suggested managing the yield of customers on various flights using ‘market segmentation’ and differential pricing. Currently, this process of managing yield is also called *revenue management* or *dynamic pricing*. We will use these terms interchangeably. Whatever one calls it, the science of dynamic pricing has advanced markedly in recent years, the mathematics and science winning many scholarly awards, and numerous firms adding substantially to their bottom lines using these concepts. In addition to airlines, other industry success stories have come from diverse places like the French National Railway, rental car companies, hotels, vacation cruise liners, cellular telephone firms and retailing.

Large-scale infrastructure systems provide services and thus are also likely to exhibit marked time-of-day, day-of-week and season-of-year demand patterns. They typically exhibit regular peaks that are significantly greater than the average levels of demand. This condition necessitates the construction of an infrastructure system with sufficient capacity to meet the peak demand, leading to underutilization of the system during non-peak time periods. For systems with little or no storage capability, the capacity must exceed the forecast peak demand to ensure continuous service.

As the peak demand for a particular system approaches the available capacity, the system will become congested. There are several management options available to deal with the congestion:

1) Capacity Expansion
2) Capacity upgrades
3) Substitution
4) Rationing (Discriminatory or non- Discriminatory)
5) Loss or Degradation of Service
6) Demand Management Congestion Pricing

Capacity expansion is simply investment in additional capacity, such as extra runways, highway lanes, power plants, wires, or pipelines. Capacity expansion is the most direct means of alleviating congestion, but it may not be the most cost effective. In some cases, capacity expansion may not be possible. For example, the frequency spectrum is fixed and airports and roads in urban areas often face geographical constraints that preclude expansion.

Capacity upgrades typically consist of new technologies to increase the efficiency of existing infrastructure. These include faster routers for the Internet, intelligent traffic signals for roads, improved air traffic control systems to reduce landing and takeoff delays, and methods for frequency sharing in communication systems. Capacity upgrades are dependent upon the development of new technologies, and are often limited by compatibility with legacy systems. In addition, although capacity upgrades can alleviate congestion, they do not deal with the issue of differential utilization during peak and off-peak time periods.

Substitution involves the development of alternatives for the congested service. Public transit is regarded as a direct substitute for driving automobiles. Other examples include train travel for air travel, and wireless for wired communications. Substitution can be limited by consumer preferences since perfect substitutes rarely exist.

Rationing is a means to allocate scarce resources in response to congestion. Rationing involves limiting the consumption of the good/service so that the system is able to function without demand overwhelming the capacity. Rationing can be implemented in a discriminatory or non-discriminatory manner. Discriminatory rationing favors one group over another based on selected attributes. It may be used to ensure that disadvantaged groups have access to limited resources, for example handicap parking spaces. Non-discriminatory rationing is often instituted without control of the allocation. Telephone busy signals, rolling blackouts, and packed subway cars are examples of non-discriminatory rationing. Rationing allows those who continue to receive the service to do so at the normal level of service quality. Rationing enables a portion of the demand to be served without the need for additional investment.

Loss or degradation of service, as opposed to rationing, involves a reduction in the level of service quality or the complete loss of service. Examples include uncontrolled blackouts, gridlocked roadways, air traffic control delays, and slower than normal Internet connections. Service loss or degradation is the result of insufficient or nonexistent action to manage congestion.

Demand management typically consists of incentives to reduce overall demand and can include subsidies for demand side efficiency investments or for switching to alternative systems and tariffs to reduce demand. Congestion pricing can be considered a form of demand management focused specifically on reducing temporal peak demand using economic incentives. Congestion pricing may be location based since demand patterns are typically spatially as well as temporally distributed. All further references to time-based pricing schemes generally apply to location-based pricing as

well. A combination of time-based and spatially based-pricing would typically be necessary to obtain maximum theoretical system level efficiency.

Congestion pricing can cause demand to shift to off-peak periods, to substitutes, or to be reduced overall. The additional costs of ensuring sufficient capacity to serve the peak demand can be eliminated by shifting demand from the peak periods/locations to off peak periods/locations. In addition to reducing the overall system capacity requirement, smoothing demand patterns will increase the overall utilization of existing capacity, potentially leading to even greater cost savings, for instance in the more efficient scheduling of support personnel.

2.1. Types of Congestion Pricing

Congestion pricing can be implemented in multiple ways. The theoretically most efficient method of congestion pricing from a purely economic standpoint is to dynamically set the congestion price at the level of the system marginal cost of the demand increment. More simple methods are typically implemented in practice. These include a single peak price that is higher than the off peak price, where the peak and off peak time periods and/or locations are well defined and static, and a variety of schemes that are in between fully dynamic and static pricing.

Fixed time of use pricing is simple, easy to implement, and mitigates the risk to consumers by eliminating price volatility. The vast majority of implementations of congestion pricing have been a form of fixed time-of-usage based prices. Time of use prices work better when demand patterns are regular since they can more closely approximate dynamic prices in this case. The disadvantage of time-of-use prices is that they do not change with demand or system conditions and therefore are less efficient than truly dynamic prices.

Dynamic prices provide efficient signals for demand and investment response, but can be difficult to implement and volatile. Dynamic pricing needs real time (or near real time) metering, communications, and control to enable calculation of prices, instantaneous communication of those prices, and demand response.

Critical peak pricing (CPP) is a hybrid scheme that uses a basic Time-of-Use (TOU) scheme with a limited number of critical hours (typically corresponding to system peak hours) designated with a premium charged above the standard price for consumption during these hours. CPP offers the potential for improvement over standard TOU rates by differentiating between peak hours and inducing greater response during the hours of highest demand. By limiting the number of critical peak hours, CPP effectively limits the maximum potential reduction in overall system peak load.

CPP requires a greater degree of metering than TOU since typical TOU schemes need only on peak and off peak consumption measurements versus CPP, which must be capable of hourly measurements of demand. CPP also needs additional communications equipment in order to notify customers of the hours that are designated as critical. The lead-time for notification is typically specified in the contract for CPP. Longer lead times allow consumers to plan ahead for any reductions in consumption, thereby reducing the costs. The tradeoff is that the need to forecast the critical periods ahead of time may lead to inefficient designation – the service provider may falsely anticipate a critical period due to uncertainties in demand.

Contracts for differences (CFD) are a mixture of dynamic and flat rate (or fixed fee) pricing schemes. They provide real time pricing signals to consumers, while at the same time providing a hedge against price volatility. CFD contracts imply basically a two-part tariff –

1) A flat rate for the baseline consumption
2) Real time prices for deviations from average.

CFD involves setting a baseline demand curve for which a fixed or flat fee is charged/agreed to *a priori*. Deviations from the baseline are priced dynamically. This has the advantage of providing a hedge for the consumer for their baseline consumption, while maintaining incentives for efficiency. As long as they consume their baseline or average levels of power they will be guaranteed a flat rate. By charging dynamic prices for deviations from baseline, however, the incentives for reducing demand during peak price periods are passed fully to the consumer. CFD has the advantage of protecting the consumer from spot price fluctuations while maintaining the incentives for peak reductions by allowing the consumers to effectively sell back their contracted power at times of peak prices. The opportunity cost the consumer incurs for consuming at the baseline level is equal to the actual cost they would face under dynamic prices.

Table 1. Comparison of incentives and risk allocations for congestion pricing schemes.

	Consumers		Service Providers	
	Incentives	Risk (short term)	Incentives	Risk
Flat Rate	Efficiency only – no incentives to shift from peak to off peak	Fully hedged	Max. incentive to smooth out demand curves	Fully exposed (except for forward or bilateral contracts)
Time of Use (Differential Pricing)	Provides some incentives to shift from peak to off peak -may lead to shoulder peaks- does not differentiate between a high priced day and a low priced day on the spot market – no real time adjustment incentives	Fully hedged	Incentives to flatten demand within periods Between periods depends on the differentials between pricing periods as well as actual spot prices	Fully exposed (except for forward or bilateral contracts)
Critical Peak Pricing	Same as TOU except provides incentives to shift/reduce peak demand during critical hours	Fully hedged except for the "critical hours"	Fewer incentives to decrease peak demand	Hedged for critical hours
Real Time Pricing	Full incentives to reduce peak consumption/shift from peak to off peak.	Fully exposed (although they may enter forward contracts to partially or fully hedge)	Incentives to reduce peak demand only	Fully hedged
Contract for Differences	Incentives to reduce deviations. Full incentives to reduce peak demand if payback is 100%. May reduce incentives to install technology for peak reduction since risk is removed. Also, may inflate baselines (moral hazard).	Fully hedged for baseline consumption fully exposed for deviations	Tighten deviation bands, underestimate baselines	Fully exposed for baseline demand fully hedged for deviations

Several issues surround the setting and updating of baselines:

1. How are they set initially? A 'moral hazard problem' may exist if consumers know in advance how the baseline will be set.
2. Are they based on average baselines (creates adverse selection) or individually determined?
3. How often are they updated?
4. Are the baselines adjusted for weather, seasons, holidays, etc?
5. What are the sizes of the deviation bands?

This section has presented a range of options for implementing congestion pricing. Table 1 below provides a comparison of the incentives for response and the allocation of risk provided by each of these pricing schemes.

3. POTENTIAL FOR SAVINGS FROM CONGESTION PRICING

This section presents gross estimates of the potential savings available by implementing congestion pricing for several infrastructure systems. These estimates are constrained by the available data for each system and the scope of this paper. They are intended to give a general idea of the magnitude of potential savings from avoided investment and reduced congestion.

3.1. Electricity

Shifting demand from the daily peak time periods[†] to off peak periods has significant potential to offset investment in generation and transmission capacity, as well as capturing several secondary benefits. An upper bound of the potential savings can be obtained by calculating the potential benefits of perfectly leveling demand across all hours of the year. This assumes leveling daily, weekly, and seasonal demand patterns, which would need large-scale storage capability to implement. Electricity demand has a 1.6% annual growth rate (EIA, 2006). Therefore, annual investment savings can be determined by calculating the avoided costs of expanding system generation and transmission capacity to keep pace with this growth rate.

In 2004, non-coincident peak demand for the US was 704,459 MW, while total demand was 3,970,555 thousand MWh (EIA, 2005). The perfectly smoothed demand would therefore be 453,000 MW continuously. This results in a maximum theoretical reduction in peak demand of 251,000 MW (36%).

Table 2. U.S. Electricity Demand Statistics

Peak Demand	704,459 MW
Total Demand	3,970,555,000 MWH
Smooth Demand	453,000 MW
Peak Reduction	251,000 MW (36%)
Annual growth Rate	1.6%
Annual Peak Growth	**11,000 MW**

Generation Investments

Generation capacity costs range from approximately $500 per KW for natural gas turbines to $2000 per KW for nuclear power plants. In order to maintain sufficient system capacity with a 1.6%

[†] Due to the nature of electricity (flows cannot be controlled) location-based shifting of load is not possible (e.g. you cannot move your house, factory, or power plant easily). Only long term investment decisions – such as building generation plants or relocating industrial facilities can affect location-based congestion.

annual demand growth rate, an additional 11,000MW of generation capacity would need to be added every year (based on peak demand). Since a mixture of generation technologies will necessarily be constructed, a relatively conservative figure of $750 per KW is used to estimate avoided costs for generation investment. Each 11,000MW of generation capacity would cost $8 Billion at this rate.

Transmission Investments

Since transmission capacity requirements are based upon the adequacy of transmission for serving peak demand, reducing the peak demand has a significant potential to reduce transmission capacity requirements. The costs of transmission expansion are approximately $2 billion for every 10K MW of additional transmission line capacity. (Ofori-Atta, 2004) By reducing peak demand as in the generation case above, $2 Billion in savings can be achieved.
Therefore, the Annual savings from flattening peak (1.6% capacity reduction):

$2 Billion Transmission
$8 Billion Generation

With a perfectly flat demand pattern, a 1.6% growth rate can be sustained for almost 28 years without the need to expand capacity over current levels. This would result in total avoided costs for generation and transmission of nearly $100 Billion (Present Value using a 10% discount rate)

Table 3. Avoided infrastructure costs for every 1% reduction in peak electricity demand ($)

Generation	5,300,000,000
Transmission	1,400,000,000
Total	6,700,000,000

In addition to direct investment substitution, the total benefits of peak shaving will include reductions in spot prices (including congestion costs) and reductions in reserves and ancillary services.

Location based Congestion Costs

Location based congestion in the electricity system occurs when specific transmission lines approach their capacity to transmit power. In order to maintain system stability, generation plants are typically re-dispatched (i.e. their power output is adjusted). Doing so ultimately increases the costs of electricity since the output of less efficient plants must be increased and that of more efficient plants reduced. In some markets, these costs are internalized in the overall electricity price, while in others there are explicit location based congestion charges (known as locational marginal pricing (LMP)). Demand response can reduce such location based congestion costs, and eliminate or at least defer the need for transmission expansion. Annual Congestion Costs in the PJM system alone are estimated at between $400 to $500 million per year (PJM, 2003 State of the Market).

If congestion costs are proportional to peak demand, however, it can be expected that system congestion costs will be reduced by at least 10% as a result of an 11% reduction in peak demand. This will result in a savings of $40 to $50 million annually in the PJM system alone.

The main beneficiaries of reductions in congestion costs, of course, are those who live in congested areas. In this case, the system benefits will not accrue evenly across all customers.

Loss Reductions

Up to 10% of electricity generated is lost as it travels over the transmission and distribution system. Losses increase as more power is delivered across the lines. By reducing the peak loads, the amount of losses can be reduced as well.
Losses increase exponentially with electrical current across transmission (and distribution) lines, therefore a reduction in power delivered (i.e., load) leads to an exponential reduction in losses, and reducing peak load has the largest impact on reducing losses. The larger the differential between the peak and the off peak loads is prior to shifting, the larger the reduction in total system losses. For systems with smoother load profiles, the reductions in losses will be smaller.

Ancillary services

Demand response also has significant potential to reduce the need for ancillary services. By smoothing the overall system load and shifting reactive power demand away from system peak loading, thermal storage will reduce the need for ancillary services such as VAR compensation, frequency control, and reserves. In addition to reducing the need for ancillary services from load shifting, demand resources can be utilized directly for VAR compensation, frequency control, and short-term reserves.

For reactive power, losses are more significant. Reactive power losses behave in the same manner as real power losses, except that they are proportional to line impedance rather than resistance. Since impedance is much greater than resistance, the reactive power losses are not only much larger than real power losses, but the subsequent reductions in reactive power losses during peak time periods also have a greater magnitude and potential impact.

Around 7% of total real power generation is lost through real power losses (U.S. Climate Change Technology Program, 2003). Losses during peak load periods can be expected to exceed this level due to the exponential relationship between current and losses. Reducing peak load by 20% has the potential to reduce peak losses by over one third. This would allow for additional reductions in peak generation of at least 2% and possibly up to 4 or 5%. In New England, for example, there are currently 11,826 MVARs of reactive power capacity at an annual cost of $12.4 million. New England pays a base rate of $1050/MVAR-yr. (ISO-NE, 2003)

The preceding section presented several areas for which peak shaving provides significant benefits. Congestion pricing programs in electricity can provide benefits through reductions in physical capacity, in requirements for ancillary services, and in losses and congestion.

Current Examples

There has been limited implementation of congestion pricing programs in the U.S. electricity sector. In California, consumers participating in Pacific Gas and Electric's time of use program have reduced their peak electricity consumption by 18%. (GAO**,** 2004) An 18% reduction in peak consumption is consistent with estimates developed by Black (2005) and can be achieved with little sacrifice from a consumer point of view. The number of consumers participating in time of use programs in California, however, remains small.

In addition, Southern California Edison has an air conditioning cycling program (SCE, 2004), and California plans to implement a critical day pricing program and is investing $1.8 billion to upgrade metering and communication equipment to implement the program. Under this plan, 15 days will be

designated as critical during each summer and the price of electricity will be dramatically higher for those days, in exchange for lower prices on non-critical days. (Smith, 2006)

There are several other time-of-use programs throughout the U.S. and a few critical peak-pricing programs (e.g. Gulf Power). In addition, there are direct demand response programs that typically offer payments to large consumers to reduce their demand in response to signals from the utility or system operator. (GAO, 2005) The government of New South Wales in Australia has made electricity demand management a requirement for licensing for electricity distributors. (*Demand Management for Electricity Distributors*, 2004)

In order to implement congestion pricing significantly, enabling investments in metering and communications technologies are necessary. Current metering technologies do not allow for monitoring of consumption during specific time periods nor is there a means to signal dynamic changes in prices. Existing control technologies, such as programmable thermostats, could be used by consumers to respond to congestion pricing schemes on a large scale, but would also require additional investment.

The best and most concise statement we have seen in support of electricity demand management is from The Warren Centre for Advanced Engineering in Sidney University, Australia:

> **Demand management answers growing electricity needs**
> Demand management is one of a number of ways in which suppliers of a resource can meet their customers' needs by either shifting or reducing demand peaks. Currently, a relatively large percentage (say 15%) of the assets required to deliver electricity to the consumer are used on a relatively small number of peak days (say 3, or less than 1%).
> . . .
> A proven performer internationally for over a decade, demand management addresses the causes rather than the symptoms of excessive energy needs. Demand management uses a range of strategies to modify the level and timing of energy demand. Within the demand management toolkit are energy efficient appliances and buildings, distributed generation, standby generation, interruptible contracts, improved network efficiency and more accurate pricing. For customers, the benefits of this alternative to more generation and network expansion include lower energy bills, better energy services, the improved utilisation of resources and fewer environmental costs.[‡]

3.2. Airports

Air traffic demand patterns have seasonal, weekly, and hourly components. Hourly demand patterns exhibit peaks in both the morning and evening periods. Seasonal and weekly patterns correspond to workweek and vacation/school schedules.

There are several potential bottlenecks in the air transportation system that can be alleviated through capacity investment. The most obvious means of eliminating congestion is through the construction of new runways. Alternative means include improving the air traffic control system, and adding gates, taxiways, or planes to the system.

Congestion pricing policies for air traffic can be implemented at the consumer level, the flight level, or both. Airlines already have sophisticated, dynamic pricing schemes for managing consumer

[‡] http://www.warren.usyd.edu.au/bulletin/NO38/ed38art7.htm

level demand and maximizing capacity utilization. A similar system of congestion pricing for planes would enable the air transportation system to operate more efficiently.

Peak demand in air transport has both a time and location component. The hub and spoke system employed by many major airlines leads to several highly congested hubs with coordinated flights to maximize connections and minimize consumer layover times.

In addition, there is a strong seasonal component to commercial air travel. Leisure travelers may be willing to adjust their flight schedules within a day or week, but it is unlikely they are able to dramatically alter the seasonal pattern to in their travel schedules. Business travelers have much less elastic demand (as evidenced by the significantly higher fares they are charged by airlines) and are unlikely to adjust their demand between days, and less willing to do so even within a day. Due to the lack of substitutes available for long distance flights, however, consumers must accommodate to the available flight schedules. For shorter flights, it is possible for rail, buses, or autos to be substituted for flying. The immediate post 9/11 decrease in air travel demand illustrates that there is a degree of elasticity which can make congestion pricing effective at smoothing demand peaks, however.

In order to estimate the total system capacity for the US air traffic system (commercial aviation only), we choose the day of the year with the highest number of departures. June is typically the peak month for air travel. For June 2005, the maximum number of flights in a day was 20,899. The corresponding maximum number of flights in an hour was 1375. (BTS, 2005)

Assuming that congestion pricing policies could smooth demand within a day between the hours of 6 a.m. and 8 p.m. results in a 9% reduction in peak demand. Smoothing flights across these hours on the busiest day would result in 1264 flights per hour, or a reduction of approximately 100 flights from the peak.

According to the FAA (2004), a single runway (ignoring other constraints such as gate, taxiways, etc.) can handle 200,000 operations (takeoffs and landings) per year. Assuming and equal number of takeoffs and landings, this is approximately 15 takeoffs per hour for an 18 hour day.
A reduction of 100 flights is then equivalent to adding 6 additional full capacity runways to the system.

Using the data from Delta airlines (Boatright, 2001), at approximately $500 million per additional runway, this is equivalent to savings of $3 Billion, not including costs of lost time, expended fuel, wages, and additional emissions caused by delays.

This smoothing would involve distribution of flights only in time. Since not all airports are operating at full capacity when the system is fully loaded, even greater reductions are possible by redistributing flights spatially. This could be accomplished by altering the current hub and spoke systems while minimizing disruptions/eliminations of service by simply rerouting some flights. In order to dramatically smooth out airport demand peaks across the entire system, it is likely necessary to eliminate or at least reduce the hub and spoke system.

Weather related reductions in system capacity are a significant contributor to delays in the air transport system. Some of these delays can be reduced by additional investment, such as the runway project at Logan Airport in Boston. This runway would increase capacity during certain weather conditions, but would not increase the overall capacity of the airport. Congestion pricing, however, may have limited applicability to weather related delays.

Applying congestion pricing to weather related congestion is complicated by the inherent uncertainty associated with weather events and the general lack of short-term substitutability of flights. Congestion pricing could determine the priority of flights during weather events, allowing the most valuable (in terms of willingness to pay) flights to receive preference for take off or landing slots. Weather-related dynamic congestion pricing, however, would not likely be able to significantly alter flight schedules ahead of time to prevent delays/congestion from occurring during weather events.

In addition to reductions in capacity investment, additional savings can be obtained by reducing delays associated with congestion. Winston (1991) estimated that approximately $6 billion[§] could be saved by congestion pricing and 17 billion[**] by congestion pricing combined with optimal investment.

Attempts to implement congestion related pricing have been made in Boston, New York, and London. In each case, the congestion pricing was fairly static and based on aircraft size and/or time of day. The programs were successful in terms of reducing flight demand and congestion. The majority of demand reductions were obtained through significant increases in the relative costs for smaller aircraft. Legal challenges from those most affected by the congestion, however, forced each jurisdiction to eventually abandon or significantly reduce the programs. (Schank, 2005)

3.3. Sea Ports

Investment in harbors and channels more than doubled from 1991 to 2000 to $252 million (GAO, 2002). Smoothing demand for harbors could alleviate some of this cost. Some cargo (such as foodstuffs) is much more time dependent than other cargo (e.g. clothing). The less time critical cargo would then theoretically be more elastic in terms of paying congestion prices. This is offset, however, by the opportunity costs to ships of delays. The combination of (relatively) low levels of investment and long time scales makes it likely that the benefits of congestion pricing applied to shipping in terms of reducing investment costs will be small.

Altering sea traffic is more difficult due to the time constants involved in transit, loading and unloading, and diverting traffic from one port to another. In addition, the interdependency with land based transit (roads and railroads) and security/customs checks makes diversion of ships to underutilized harbors more complicated.

3.4. Roads

Over $25 billion was spent to increase highway capacity in the U.S. in 2000. Yet, according to Schrank (2006), this was not sufficient to meet even 50% of the growth in vehicle miles traveled. Traffic increased an average of 3.5% per year from 1983 to 2003 in the U.S. Since traffic has grown at a faster pace than highway capacity, congestion has continuously increased. The average travel delay for peak period drivers has grown from 16 hours per year to 47 over this time, creating over $60 billion in annual congestion costs. (Schrank, 2006)

In order to deal with the growing problem, there are 4 basic options:

1. Congestion pricing
2. Capacity expansion
3. Shifting demand to public transit
4. Do nothing – live with congestion delays (Downs, 2004).

[§] Converted to year 2003 dollar-equivalent.
[**] Converted to year 2003 dollar-equivalent.

Due to the difficulties in expanding capacity in urban areas and the limitations of public transit (both physical and due to preferences), congestion pricing has the most potential to alleviate congestion. Congestion pricing faces political obstacles such as concerns over fairness and unwillingness by consumers to pay for a service that has been "free" for some time.
Benefits of congestion pricing for roadways include:

- Avoided investment
- Reduction in delays
- Reduction in fuel costs
- Reductions in Maintenance
- Improvements in Emergency services.

Due to the localized nature of traffic congestion, there is no overall national demand pattern from which to estimate overall effects of a generic congestion pricing policy. The TRB (1994) estimates that peak period pricing could save between $5 and $11 billion in investments and time.

There are several methods to implement congestion pricing for highways. They include manual toll-booths (which can create bottlenecks themselves), automatic toll booths (e.g. fast lane), zone licensing, or electronic pricing with automated identification (photos or smart cards). (Hau, 1992)

London and Singapore have both instituted congestion pricing programs and have seen significant results. In London, drivers pay a fee to enter the central London zone during the hours from 7 a.m. until 6:30 p.m. The fee is static and does not vary by hour or road conditions. (Litman, 2006) In the first year of implementation, traffic in the central London zone was reduced by approximately 15% (see Figure 1 below). (Transport for London, 2005)

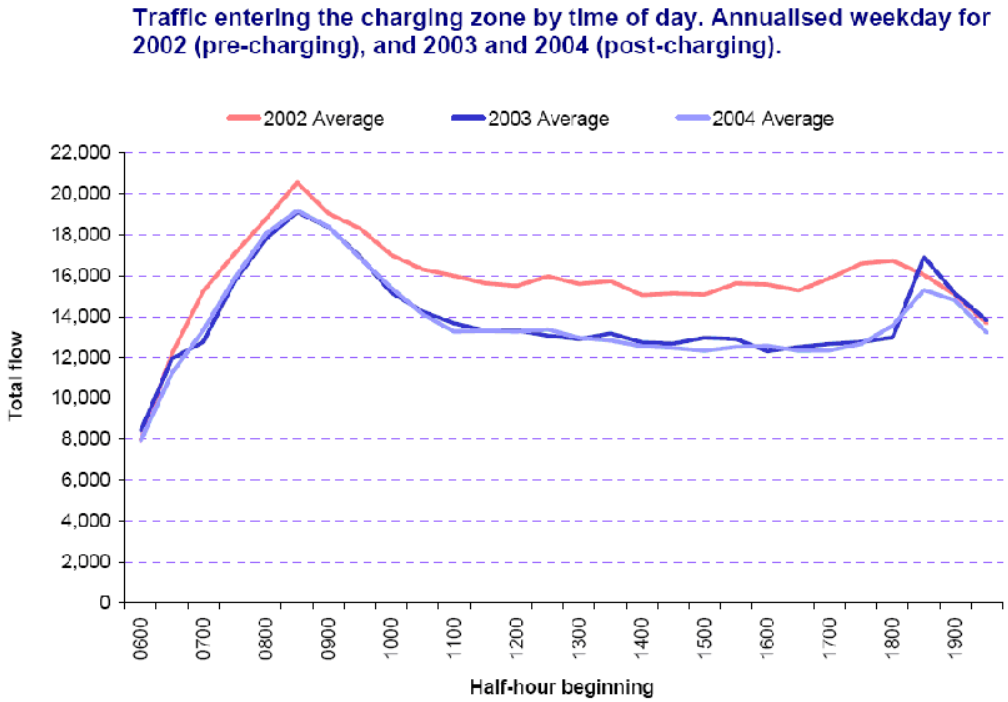

Figure 1

Source: Transport for London, 2005.

As shown in the figure above, even though the program significantly reduced traffic in the central zone, the majority of the reduction is in the intermediate hours. The evening peak actually increases and the morning peak reduction is less than the overall reduction. This is the limitation of having a single fee for all of the commuting hours. The overall system capacity requirement was not reduced as much as it would be with differential pricing for the two peak hours.

Singapore introduced zonal pricing in 1975 and as increased the sophistication of the system over time. Singapore now uses electronic charging with 30-minute time increments and locational differentiation. Also, the transitions from one charging level to another are graduated to prevent sudden, dramatic changes that can lead to adverse consumer behavior (such as speeding to avoid a higher priced time period). There are also different prices for cars, motorcycles, and light and heavy trucks. Congestion prices are reviewed and updated quarterly. The congestion pricing system has enabled average highway and arterial road speeds to be maintained. (LTA, 2006)

3.5. Communications (wire and wireless)

Between 1980 and 2001, there was approximately $1.2 trillion invested in the US telecommunications sector (Phoenix Center, 2006). Averaging investment over this time period allows for a smoothing of the telecom investment "bubble" and gives a value of approximately $56.5 billion per year. This investment encompasses the entire telecom industry. The convergence of ITC services makes this a reasonable figure to use in estimating the amount of investment needed to keep pace with annual demand growth.

Stuck (2003) estimates a 7.25% annual growth rate in the telecom sector, across all services. Using congestion pricing across telecom services would therefore result in a potential savings of approximately $7.8 billion for each percentage reduction in overall peak demand[††]. A 10% reduction in peak demand could therefore eliminate up to $80 billion in investment.

Telephone service has an established history of time based congestion pricing, but not with dynamic congestion pricing. Implementation of dynamic congestion pricing for telephone service is technically feasible and straightforward. The nature of the Internet (independent servers and routers, which are co-dependent for transferring data), however, makes implementation of congestion pricing complex under current operating methodologies and jurisdictions. Current protocols do not lend themselves to implementation of priority pricing. Although individual ISPs are able to implement their own congestion-based prices, these may or may not synchronize with overall system conditions. The potential for significant gains through dynamic or priority pricing exists, but is largely dependent upon the development of standards/protocols to enable such pricing schemes.

3.6. Pipelines

Pipelines represent another category of networked infrastructure systems. The ability to store oil, gas, water, and other resources that utilize pipelines, however, tends to limit the potential for congestion pricing, since storage may be used to alleviate spikes in demand without the need for additional pipeline capacity. In cases where storage costs are significant and storage is not used as a hedging mechanism for fluctuations in the price of the basic commodity, congestion pricing may be a viable means to reduce storage/pipeline investments. Total annual investment in pipeline systems (including equipment and storage) is approximately $2 billion for natural gas (EIA, 2005) and $3

[††] Assuming that investment is equal to demand growth.

billion for water (BTS, 2006). Many of these expansions are for service to new areas, and not to directly relieve congestion.

4. THREE R's: ROBUSTNESS, RESILIENCE AND REDUNDANCY

Critical infrastructure systems provide life-supporting services to customers. They must be able to continue to do that under a wide range of circumstances. Our 'intellectual device' to estimate present value, discounted future savings by flattening the demand over the course of 24 hours is clearly just that – a device that gives an upper bound to possible savings. No one expects that demands could ever be smoothed to total flatness. Customers with their preferences and life style constraints would not tolerate it. In addition, if such flat demand patterns were to become a reality, these infrastructure systems would not be able to function at full capacity 100 per cent of the time.

There are needs for scheduled and unscheduled maintenance, repair and system upgrading. That means that we must anticipate and plan for 'down time' in parts of the system.

There are outside perturbations and even shocks to the systems that can cause them to operate in extreme modes, something they can do only for short periods of time. An example is the set of supply chain systems shocked by the terrorist events of September 11, 2001, when for instance massive queues of trucks lined up at Canadian border points attempting entry into the USA, but enhanced security delayed their passage[‡‡]. Hurricanes, tornados, floods, intense heat and long droughts - all sorts of extreme weather can interfere with planned operations of infrastructure systems.

Thus these systems must be designed with robustness and resilience in mind. Robustness means strong and durable, able to withstand shocks. Resilience means able to recover quickly from unexpected conditions or setbacks, bendable but not breakable. Redundancy – meaning multiple parallel system components -- is one system design strategy that can lead to improved robustness and resilience.

Despite some components of system redundancy, the nation's electric power system has shown repeated lack of resilience over the past 50 years. Massive cascading blackouts on the East coast are examples: the first major one in 1965 (November 9)[§§] and the most recent one in 2003 (August 14)[***]. But even this past summer (2006), there were numerous smaller system breakdowns, due to unseasonably hot temperatures in New York, severe storms in St. Louis, etc. So, any attention to demand management in electrical power systems must simultaneously address issues of resilience and robustness. A semi-automated demand management system to support 'graceful degradations of electrical service,' could in theory go a long way toward increasing both robustness and resiliency.

Care must be taken not to over-optimize an infrastructure system. Many airlines use sophisticated mathematical optimization tools of operations research to schedule aircraft and crews, to minimize costs while satisfying myriad constraints. Unfortunately many of the mathematical depictions of airline operations used in these optimizations assume that the sun is shining over the entire country,

---

[‡‡] See for example Sheffi, Y. ***Supply Chain Management Under the Threat of International Terrorism***, **International Journal of Logistics Management**, Vol. 12, No. 1, pp 1 – 11, 2002, and Y. Sheffi, **The Resilient Enterprise: Overcoming Vulnerability for Competitive Advantage,** MIT Press, Cambridge, MA, 2005.

[§§] Sometimes called **The Great Blackout of the Northeast.** See http://www.historyplace.com/specials/calendar/november.htm#5.

[***] Sometimes called **The Great Blackout of 2003.** See http://en.wikipedia.org/wiki/2003_North_America_blackout.

that no crewmember is sick and that no plane requires unscheduled maintenance. Rarely do these conditions hold, meaning that optimized plans that assume a deterministic system operating under idealized conditions become brittle plans, with little resiliency. Manual scrambling is then needed to adjust system operation to the realities of weather, crew sickness and airplane repair. At least one private parcel delivery air carrier plans for contingencies, in essence flying two redundant empty airplanes over the USA at critical times to be dispatched to trouble spots to pick up parcels that otherwise would not be delivered on time – due to one or more of the problematic conditions cited here.

The maximum jolt to the US airlines occurred on September 11, 2001, when all planes were indefinitely grounded, many sent to airports for from intended destinations. This left the airlines in a condition far from usual stochastic perturbations due to the factors discussed above, and in a state never anticipated. CALEB Technologies, working with Continental Airlines, developed the CrewSolver decision support system to generate globally optimal or near optimal crew recovery solutions. Continental used this system as soon as clearance was given by the U.S. Federal Government to resume flying. As a result, Continental was first in the air back at full planned schedule. The airline claims to have saved $40 million just by this first application of the system.

Their recovery time was minimized by the CrewSolver system, a system that has now been adopted by other airlines[†††]. This type of system shows that service systems that use heavily demand management tools and techniques and are thus tightly optimized, can still be resilient in the presence of minor and major perturbations, but care and attention must be devoted to the topic.

5. BARRIERS TO IMPLEMENTATION

As this paper has shown, there are large potential benefits from the implementation of congestion pricing to several infrastructure systems. There are, however, several barriers to implementation of congestion pricing. They include:

- Distributional Effects
- Cost Allocation
- Status Quo Bias
- Lack of information

Congestion pricing programs typically will have a greater impact upon certain consumer groups, such as small aircraft (airports), commercial customers (electricity), trucking companies (roads). These distributional effects create a strong interest in the disaffected groups to oppose congestion pricing. The majority, on the other hand, is not motivated to action in support of congestion pricing since the benefits are uncertain and diffuse. As seen above, airport congestion pricing programs have been rolled back in several jurisdictions for exactly this reason (Schank, 2005).

In addition to political interests, it is often difficult to calculate the actual costs of congestion in order to set the "right" congestion price. The theoretically optimal congestion price is equal to the marginal costs of congestion created/incurred by the last increment of demand. Even if this cost could be calculated, dynamically implementing it would be difficult and create fairness issues.

[†††] Yu, Gang, et. al., A New Era for Crew Recovery at Continental Airlines, Interfaces, Vol. 33, No. 1, Jan.-Feb., 2003, pp. 5-22. On line at http://pubsonline.informs.org/feature/Edelman/1526-551X-2003-33-01-0005R.pdf

Consumers also are conditioned to free services (roads) or stable prices (electricity) and therefore proposals for implementation of congestion pricing are typically viewed as attempts to raise prices and/or create volatility. Status quo bias amongst consumers and regulators creates a reluctance to implement congestion pricing until after a crisis has been identified. This reluctance is only relieved then if the congestion pricing can be seen to directly solve the crisis. Regulators especially tend to be risk averse since they will rarely lose their jobs for continuing previous policies or following the methods practiced by a majority of other regulators. Innovative programs such as congestion pricing can be risky for regulators, especially since they will disadvantage certain constituency groups.

A lack of information often leads to opposition to congestion pricing programs as well. In California, for example, a consumer advocate criticized the variable pricing program that offers a discount for the vast majority of days and charges a premium only on the hottest days of the summer. The advocate complained that consumers without air conditioning would not be able to reduce their consumption on the hot days when prices are raised (Smith, 2006). The advocate failed to understand that customers without air conditioning would benefit the most from the variable pricing. Consumers without air conditioning would enjoy the benefits of discounted electricity for the vast majority of the time and already consume much less during the high priced hours.

Consumers with air conditioners use the most electricity on hot days, since their air conditioners are running more often, and these consumers will pay the most under variable pricing. With fixed pricing, consumers without air conditioning are subsidizing those who do have air conditioning since all consumption is treated the same and averaged over time. A lack of understanding/information as to the true benefits and costs of congestion pricing programs can create opposition from the very consumers who would benefit the most.

Interdependency of system (internal and external)

Low levels of interdependency can allow for local/regional experimentation with congestion pricing policies with little opposition from (or impact on) other regions. High interdependency makes coordination across regions more necessary due to externalities. Actions in one interdependent region can have great affects on other regions.

Electricity and ICT – most interdependent. Local actions can have far reaching effects instantaneously. Also, many other systems are dependent upon both the electric system (e.g. traffic lights, light rail, pumps for pipelines, ATC systems) and ICT for operation.

Air Traffic – Highly interdependent (internal). Due to hub and spoke system and the existence of a few major airports with the majority of all traffic, actions in one airport will affect several others.

Pipelines – somewhat interdependent. Individual branches of a pipeline are independent. Pipelines may require electricity to operate.

Rail – some interdependence – central nodes can affect other nodes. Regional traffic may be independent.

Roads – mostly independent. Local actions to relieve traffic congestion will have little or no effect on other cities/areas.

Downs (2004) claims that the only viable option is to allow highway congestion to continue since congestion pricing is politically unacceptable, expansion is too costly and would take up too much urban space, and public transit is only able to relieve a small portion of the problem even if greatly expanded.

6. CONCLUSIONS

Congestion pricing has the potential to provide large benefits across multiple infrastructure systems. The highest potential for benefit is in electricity and roads simply due to the magnitude of spending on these infrastructure systems. The few examples of actual implementation have demonstrated success (see: Singapore, London – roads; California – Electricity; London - Airports), but have not led to large-scale implementation. Political barriers, including distribution of costs, often inhibit implementation of congestion pricing. Rent sharing mechanisms can overcome political barriers by reducing opposition by disadvantaged groups. Status quo bias is a large impediment and making examples of existing successful programs can build regulatory will/confidence to implement congestion pricing on a larger scale.

Table 4. Potential for congestion pricing to reduce infrastructure investment costs (US).

Domain	**Peak Cycle time**	**Min Demand response Time**	**Method of response**	Potential Savings
Electricity	Daily	instantaneous	Shift, forego	$10B per year
Roads	Daily	~ hourly	Substitute, shift	$5-11B[###] per year
Airports	Daily, Weekly	Hours to days	Substitute, forego, shift	$17B
Ports	Daily	Weekly or more	shift	Not Estimated

Acknowledgement

We gratefully thank Mr. Cordell Hull for awarding a research grant to MIT to support this research.

REFERENCES

[1] Black, Jason W. (2005), Integrating Demand into the U.S. Electric Power System: Technical, Economic, and Regulatory Frameworks for Responsive Load, Ph.D. Thesis, MIT.

[2] Boatright, John, Vice President - Delta Air Lines, Properties and Facilities, Aviation Gridlock: Airport Capacity Infrastructure, How Do We Expand Airfields? www.aviationweek.com/aw/generic/story_generic.jsp, April 11, 2001.

[3] Downs, Anthony, Traffic: Why It's Getting Worse, What Government Can Do, Policy Brief #128, The Brookings Institute, www.**brookings**.edu/comm/**policybriefs**/pb**128**.htm, January 2004.

[4] Gupta A., Stahl Dale O., Whinston Andrew B. (1999) , The Economics of Network Management, Communications *of the ACM* 42; No. 9, 57-63.

[5] Gupta, D., Jukic, B., Whinston, A. (1999), Impact of Congestion Pricing on Network Infrastructure Investment, *Internet Service Quality Economics,* Cambridge, Massachusetts.

[###] NRC 242

[6] Hau, Timothy D., Congestion Charging Mechanisms for Roads: An Evaluation of Current Practice, Transport Division, Infrastructure and Urban Development Department, The World Bank WPS 1071, December 1992.

[7] ISO New England Quarterly Market Report, Q2, 2003.

[8] Land Transport Authority (LTA), Government of Singapore. http://www.lta.gov.sg/. Last Accessed July 24, 2006.

[9] Litman, Todd, London Congestion Pricing: Implications for Other Cities, *Victoria Transport Policy Institute*, 10 January 2006.

[10] Ofori-Atta, K., Roseman, E., Saha, B., Stuart, S., et al., Profiting From Transmission Investment**,** *Public Utilities Fortnightly*, Vol.142, 10; 72-78, Oct 2004**.**

[11] Phoenix Center for Advanced Legal & Economic Public Policy Studies, The Truth about Telecommunications Investment, Phoenix Center Policy Bulletin No. 4, 24 June 2003, www.Phoenix-Center.Org , Last Accessed July 2006.

[12] PJM Press Release, Regional Transmission Expansion Plan Approved, June 7, 2001 www.pjm.com, Last Accessed July 2006.

[13] PJM, 2003 State of the Market Report, PJM, www.pjm.com, 2003.

[14] Schank, Joshua L. (2005), Solving Airside Airport Congestion:Why Peak Runway;Pricing Is Not Working, *Journal of Air Transport Management* 11; 417–425.

[15] Schrank, David, Lomax, Tim, The 2005 Urban Mobility Report, Texas Transportation Institute, Texas A&M University, http://mobility.tamu.edu, May 2005, Last Accessed June 6, 2006.

[16] Smith, Rebecca, PG&E to Spend up to $1.8 Billion on Upgrading Gas, Electric Meters, *The Wall Street* Journal, July 20, 2006.

[17] Southern California Edison (SCE), Schedule D-APS (Base ACCP - Residential) Tariff. www.sce.com, March 2005.

[18] Stuck, B., Weingarten, M. (2003), Telecom Demand: A Macroeconomic Analysis, Business Communications Review, May; 12-14.

[19] Transport for London 2005, *Congestion Charging: Third Annual Monitoring Report*, April 2005, http://www.tfl.gov.uk/tfl/cclondon/pdfs/ThirdAnnualReportFinal.pdf, Last Accessed June 2006.

[20] TRB. *Special Report 242: Curbing Gridlock: Peak-Period Fees to Relieve Traffic Congestion.* Vols. I and II. National Research Council, Washington, D.C. 1994.

[21] U. S. Bureau of Transportation Statistics, Airline On-Time Performance Data, http://www.transtats.bts.gov/Fields.asp?Table_ID=236, June 2005.

[22] U.S. Climate Change Technology Program, *Technology Options 2003*, Section 1.3.2, November 2003, http://climatetechnology.gov/library/2003/tech-options/index.htm, Last Accessed Sep. 2006.

[23] U. S. Department of Energy, Utilities and Sustainability, New South Wales Government, Australia, *Demand Management for Electricity Distributors*, NSW Code of Practice, Sept. 2004.

[24] U. S. Energy Information Administration, Annual Energy Outlook 2006 with Projections to 2030, December 2005.

[25] U. S. Energy Information Administration, Noncoincident Peak Load, Actual and Projected by North American Electric Reliability Council Region, *Table 3.1, in Electric Power Annual with data for 2004*, November 2005, *http://www.eia.doe.gov/cneaf/electricity/epa/epat3p1.html*, Last Accessed June 2006.

[26] U. S. Energy Information Administration, Summary Statistics for the United States , *Electric Power Annual with data for 2004*, November 2005, http://www.eia.doe.gov/cneaf/electricity/epa/epates.html, Last Accessed June 2006.

[27] U. S. Energy Information Administration, Noncoincident Peak Load, Actual and Projected by North, *Electric Power Annual with data for 2004 Report Released: November 2005, Table 3.1*, *http://www.eia.doe.gov/cneaf/electricity/epa/epat3p1.html*, 2005.

[28] U. S. Energy Information Administration, Summary Statistics for the United States , Electric Power Annual with data for 2004 Report Released: November 2005, http://www.eia.doe.gov/cneaf/electricity/epa/epates.html, 2005.

[29] U.S. Federal Aviation Administration, U.S. Department Of Transportation, National Plan Of Integrated Airport Systems (2005-2009), Sep 2004.http://Www.Faa.Gov/Arp/Planning/Npias/, 2004.

[30] U.S. Government Accountability Office, Electricity Markets: Consumers Could Benefit from Demand Programs, but Challenges Remain, August 2004.