Using Demand Response Programs to Provide Operating Reserves in Wholesale Power Markets: A Case Study of the ERCOT Market

Jay Zarnikau
Frontier Associates
Appearing in USAEE Dialogue, August 2006

Electricity markets across North America are taking steps to integrate customer load management capabilities into electricity market dispatch operations as ancillary services. There are a number of challenges inherent in introducing demand-side resources into markets originally established for generation resources. Should demand response resources adhere to the same protocols that apply to generation? Can demand-side resources truly be viewed as substitutes for generation resources? This paper explores these issues by reviewing the successful integration of demand-side resources in the Electric Reliability Council of Texas (ERCOT) ancillary services markets, where over 1,800 MW of interruptible loads are qualified to provide responsive reserves and other types of reserves.

Introduction

There is widespread agreement among policy makers on the value of demand response in electricity markets. As noted by FERC: "Demand response is essential in competitive markets, to assure the efficient interaction of supply and demand, as a check on supplier and locational market power, and as an opportunity for choice by wholesale and end-use customers."[1] The United States Congress has affirmed the importance of expanding demand response opportunities as a matter of national policy. Section 1252(f) of the recently enacted Energy Policy Act of 2005 reads:

> It is the policy of the United States that time-based pricing and other forms of demand response, whereby electricity customers are provided with electricity price signals and the ability to benefit by responding to them, shall be encouraged, the deployment of such technology and devices that enable electricity customers to participate in such pricing and demand response systems shall be facilitated, and unnecessary barriers to demand response participation in energy, capacity and ancillary service markets shall be eliminated.

The National Association of Regulatory Utility Commissioners (NARUC) approved a resolution calling for regulatory commissions to accommodate demand-side resources and "remove any unnecessary barriers to customer responses to such wholesale market price signals."[2]

---

[1] Federal Energy Regulatory Commission. *Working Paper on Standardized Transmission Service and Wholesale Electric Market Design*, March 15, 2002.
[2] NARUC, *Resolution Regarding Equal Consideration of Demand and Supply Responses in Electricity Markets*, July 2000.

In many markets, steps have been taken to formally recognize the value that demand response in the form of voluntary curtailments in response to high prices can have in markets for energy generation. Real-time pricing, critical peak pricing, and other innovative pricing strategies have similar goals. Many restructured markets in North America have established special emergency curtailment programs.

There has been less experience, but growing interest, in using interruptible or curtail able loads as an operating reserve – that is, a resource that is available to assist system operators in quickly restoring the balance between supply and demand on an electrical system in response to short-term unexpected fluctuations in supply or demand.

The wholesale market of the Western Electricity Coordinating Council (WECC) allows interruptible loads to be used for non-spinning reserves, and pilot programs are underway to test the feasibility of using interruptible loads as spinning reserves.[3] In May 2006, the PJM market expanded opportunities for interruptible loads to provide operating reserves,[4] and the New York Independent System Operator (ISO) has committed to allowing customers to offer some ancillary services soon.[5] Further, the New England ISO is implementing a pilot to establish the feasibility of using dispatch able loads as ancillary services.[6]

While other restructured wholesale markets in North America are just beginning to employ interruptible loads as a source of operating reserves, the Electric Reliability Council of Texas (ERCOT) market now has about five years of experience with this approach. ERCOT began using interruptible loads as an operating reserve shortly after in introduction of retail choice in early 2002, and the integration of interruptible loads into markets for ancillary services preceded any consideration of emergency curtailment programs. Over 1,800 MW of industrial interruptible load is qualified to provide various operating reserves. Typically, half of ERCOT's requirements for Responsive Reserves are satisfied through interruptible or curtailable loads which are qualified to provide an ancillary service, or *Loads acting as Resources* ("LaaRs"). This paper recounts the technical and economic challenges that were faced as LaaRs were integrated in competitive markets for ancillary services.

Background

About 85% of the electricity needs in the largest electricity-consuming state in the U.S. are satisfied through the intra-state ERCOT market. This electricity market has undergone gradual restructuring in recent years to introduce greater competition in the wholesale and retail segments of the industry and to relax regulatory oversight. Senate Bill 373, enacted in 1995, required the Public Utility Commission of Texas (PUCT) to establish rules to foster wholesale competition and create an Independent System Operator (ISO) to ensure non-discriminatory transmission access, an equitable interconnection process for new generation capacity, and customer protection. In the summer of 1997, ERCOT became the first ISO in the U.S. Further

[3] Mike Koszalka, "Load Control as Reserves in the West," AESP Brown Bag Seminar presentation, April 18, 2006.
[4] Bernie Neenan, Peter Cappers, and Jeremy Anderson, "Demand Response in Ancillary Services Markets," AESP Brown Bag Seminar presentation, April 18, 2006.
[5] Ibid.
[6] Ibid.

reforms occurred as a result of Senate Bill 7 (enacted in 1999), which allowed customers of most of the state's investor-owned utilities to choose among various retail electric providers for a retail supply of electricity beginning on January 1, 2002. SB 7 also prompted the establishment of formal markets for ancillary services.

Texas is home to a large number of industrial facilities involved in chemical production, petrochemicals, air separation, pulp and paper manufacturing, and steel production which can withstand short interruptions in their electricity supply with only modest economic loss. Traditionally, these facilities were served through interruptible tariffs, which provided an electrical supply to the facility at a lower level of reliability in return for a discounted price. Prior to the introduction of retail choice in 2002, the larger utility providers offered many levels of interruptible service. For example, Houston Lighting Power Company offered service which could be instantaneously interrupted by an under-frequency relay in response to a frequency excursion (IS-B), service which could be interrupted by the utility with a 10 minute notice period (IS-10) and service which could be interrupted with 30 minutes of notice (IS-30). TXU Electric offered instantaneous interruptible service (Rider I-II) and a service which could be interrupted following some notice (Rider I-NI).

Before restructuring, the ERCOT utilities reported over 3,100 MW of interruptible load.[7] These interruptible loads had two sources of potential value to the vertically-integrated utilities. Up to one-quarter of the utility's responsibility for providing *spinning reserves* to the market could be offset by interruptible loads equipped with under-frequency relays which were set to interrupt the load's purchases from the grid whenever frequency dropped below 59.7 Hz. Further, the utility was not required to procure planning reserves to meet the demand represented by interruptible loads. Total projected demand could be reduced by an estimate of the interruptible load likely to be on the utility's system during a peak period in order to establish an estimate of firm demand, which would be used in a reserve margin calculation. While some utilities (e.g., Houston Lighting and Power Company) considered all interruptible load to be "non-firm," other utilities (e.g., TXU Electric) employed a formula which used the tariff's notice period to establish the degree to which the load was non-firm for the purpose of calculating reserve margin adjustments.

To pave the way for retail choice, all tariffs (including all interruptible rates) offered by the state's investor-owned utilities serving in ERCOT were terminated on December 31, 2001. Overnight, the market lost a planning reserve resource of nearly 3,000 MW and a very significant spinning reserves ancillary service resource.[8]

[7] Project No. 22209: PUCT Market Oversight Division, *2000 Annual Update of Generating Electric Utility Data*, December 2000, p. 5.

[8] A small amount of interruptible load remained in the service areas of municipal utility systems and rural electric cooperatives which were not affected by retail competition.

The PUCT recognized the need to retain the ability to use interruptible loads. In its Order conditionally approving ERCOT's proposed Protocols in 2001, the PUCT ordered ERCOT to "Develop additional measures and refine existing measures, to enable load resources a greater opportunity to participate in the ERCOT markets."[9] A variety of steps were taken to implement this mandate.

Designing Ancillary Services Markets to Facilitate Demand-Side Participation

The design of ERCOT's wholesale market permits LaaRs to compete "head-to-head" against generation resources to provide ancillary services, such as Responsive Reserve Service (provided by interruptible loads with under-frequency relays) and Non-Spinning Reserve Service (which can be interrupted by the ISO with 30 minutes of notice).[10] The features and some of the requirements associated with these ancillary services are summarized in Table 1. Load-Serving Entities can self-arrange up to 50% of their Responsive Reserves obligations with LaaRs. LaaRs that are not self-arranged can be offered into ERCOT's day-ahead market for ancillary services. LaaRs selected to provide this service in the market receive the market-clearing price. The average monthly values for the market-clearing price of Responsive Reserves are indicated in Figure 1. The load levels of LaaRs providing Responsive Reserves are monitored by ERCOT every 2 or 3 seconds.

One of the basic principles in the design of the ERCOT wholesale market was that demand-side resources (e.g., LaaRs) should be permitted an opportunity to provide the same ancillary services as generation resources wherever technically feasible. Yet, limits were placed on the amounts of Response Reserves provided by LaaRs. Initially, this limit was 25% of ERCOT's requirements for this ancillary service (i.e., 575 MW each hour, given ERCOT's normal requirement of 2,300 MW). Later this was raised to 50% of ERCOT's need for Responsive Reserves (normally, 1,150 MW per hour). These limits were established to ensure that sufficient generation resources (with governors) were on-line to assist in the stabilization of frequency and to reduce the possibility of "over-shoot" situations (where too much interruptible load might trip off at the same time and raise frequency to an unacceptably high level).[11] The presence of this cap results in situations where a higher-priced generation resource is chosen to provide responsive reserves, despite the availability of a lower-priced demand-side resource.

[9] PUCT, *Final Order in Docket No. 23220: Petition of the Electric Reliability Council of Texas for Approval of the ERCOT Protocols*.

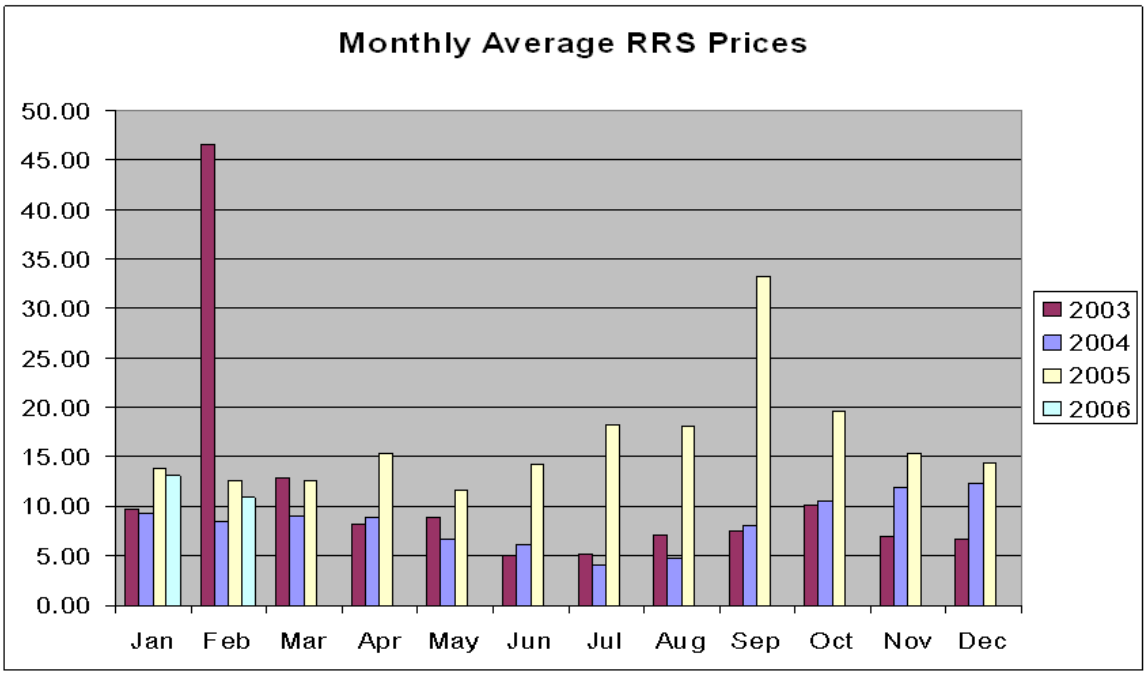
[10] LaaRs are permitted to provide Regulation on a pilot basis. In theory, LaaRs can also provide replacement capacity, although the systems necessary to permit interruptible loads to provide these services have not yet been fully implemented.

[11] Analyzed in detail in ERCOT Staff for the Reliability and Operations Subcommittee, "Utilizing High-Set Load Shedding Schemes to Provide Response Reserve Services," November 2002.

Table 1: Participation in Ancillary Services Markets by LaaRs

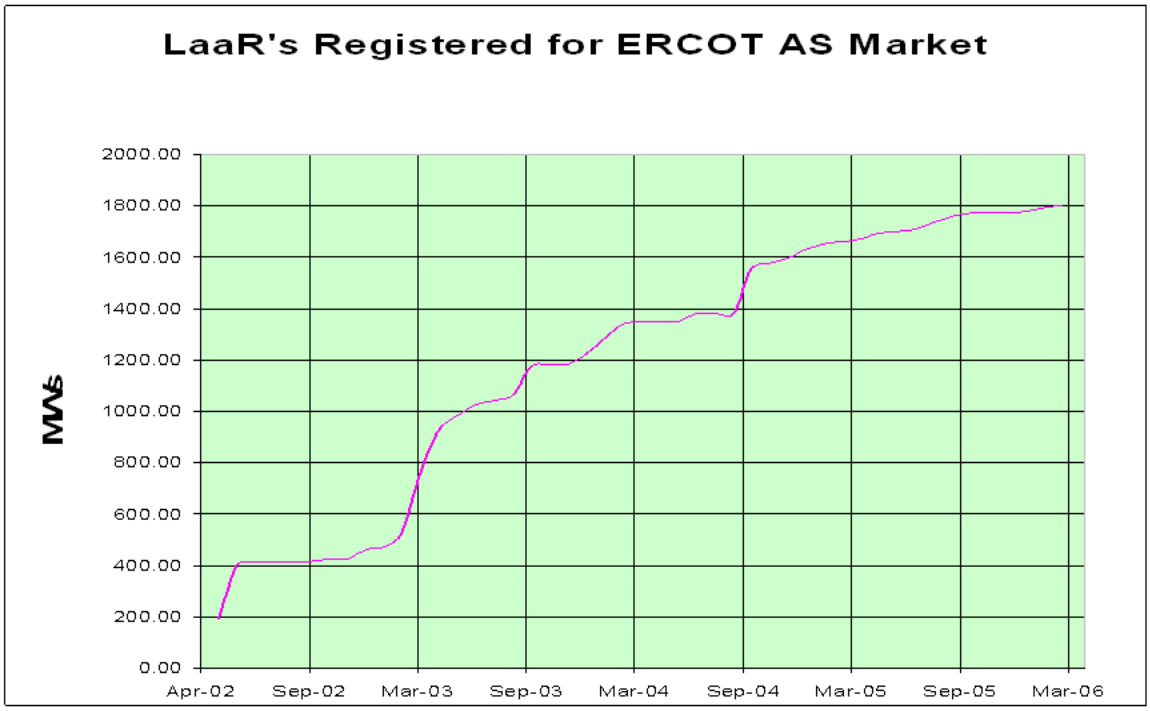
Resource	Metering	Procured by	Participation Basis for Payment	Ancillary Services Markets Payment Determination	Time to Curtail/ Interrupt
Responsive Reserves	Telemetry	Retail Electric Provider or ISO	Being available to be interrupted	Market clearing price in Responsive Reserve Market	Instantaneously, when frequency drops below 59.7 Hz; or within 10 minutes of a deployment request by the ISO
Non-Spinning Reserves	Telemetry	Retail Electric Provider or ISO	Being available to be interrupted	Market clearing price in Non-Spinning Reserve Market	30 minutes

Figure 1



Source: ERCOT staff.

Figure 2



Source: ERCOT staff.

The Successful Integration of LaaRs into Ancillary Services Markets

The integration of LaaRs into the market for Responsive Reserves has been quite successful. The amounts of load qualified to provide this ancillary service has grown over time, as indicated in Figure 2. Currently 94 LaaRs (working with 10 scheduling entities) are qualified to provide ancillary services for a total capacity of 1,826 MW. Thus the amount of load qualified to provide Responsive Reserves is well in excess of the constraint on LaaR participation in this market. LaaRs have been instrumental in preserving reliability, and typically one to four interruptions occur per year in response to under-frequency events. It is suspected that competition posed by these LaaRs has served to reduce the market price of Responsive Reserves. While there has been a good "mix" of LaaRs by size, it is noteworthy that about one-half of the total quantity of LaaRs is provided by five very large industrial loads, as indicated in Table 2.

LaaR response to a "manual" curtailment request (as opposed to an under-frequency relay action) on April 17, 2006, however, has caused some concern over LaaR performance. Although considerably more LaaR load was curtailed than requested by the ISO in response to a weather-related emergency which prompted rolling blackouts, the response of some LaaRs took longer than the 10 minutes permitted. As a result, some LaaRs may be disqualified from providing Responsive Reserves in the future.

Table 2: Categorization of LaaRs by Size

LaaR Capacity Range	Number of LaaRs	Total Capacity (MW)
1 to 10 MWs	66	283
11 to 50 MWs	20	388
51 to 100 MW	3	185
Greater than 100 MW	5	970

Source: ERCOT staff.

Until recently, low market prices combined with a higher frequency of interruptions have made the provision of Non-Spinning Reserves unattractive to LaaRs. However, as caps on the provision of Responsive Reserves by LaaRs have been reached and as the market price of Non-Spinning Reserves has soared, there appears to now be some interest among LaaRs in providing this service.

Remaining Challenges

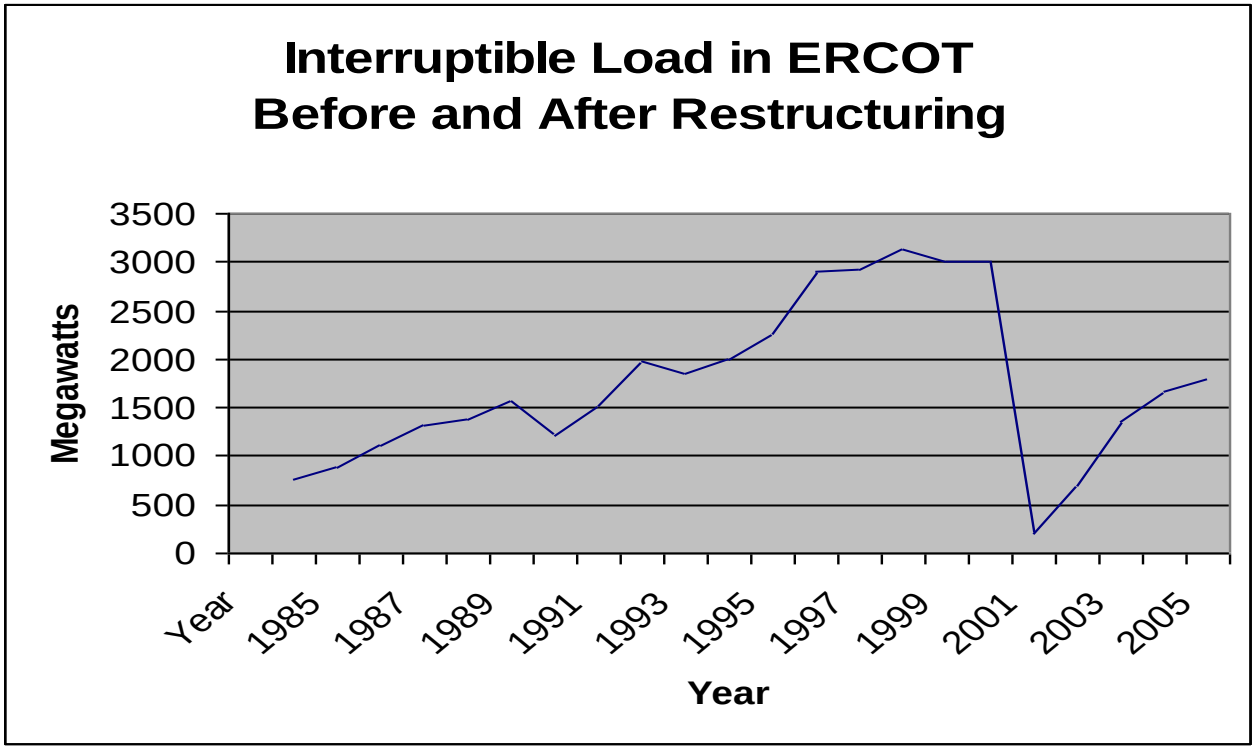
While ERCOT is ahead of North America's other restructured wholesale markets with the integration of demand-side resources into markets for ancillary services, there remains a long list of remaining challenges.

Accommodating the Full Resource Potential

While ERCOT's successful LaaR initiative constitutes the largest demand-side resource program in North America in terms of the quantity of demand that can be interrupted by an ISO in response to low frequency or system emergencies, there remains an enormous untapped potential demand-side resource in ERCOT. Indeed, the amount of load that is today qualified as LaaRs is only about one-half of the amount of load served through interruptible tariffs prior to restructuring, as indicated in Figure 3. Not only is there a "waiting list" of more than 500 MW of qualified LaaRs that cannot provide Responsive Reserves due to the cap mentioned earlier, but there are many potentially (and formerly) interruptible loads that cannot qualify as Responsive Reserves due to the unpredictable nature of their load pattern, their small size, or other reasons.

It was initially hoped that some of the formerly-interruptible load would migrate into ERCOT's Balancing up Load (BUL) program. BULs can provide offers into the market for balancing energy. However, the financial incentive for loads to participate in this market is tied to the market price of Non-Spinning Reserves, which tended to be low until 2005. There has been no participation in this program to date.

Figure 3



Sources: Data for 1985-1993 is from PUCT 1996 Statewide Electrical Energy Plan for Texas, June 1996, and represents interruptible loads plus a small contribution from various load cycling programs. Interruptible load data for 1994-1999 is from Project 22209 Annual Update of Generating Electric Utility Data, 2000. Data for 2000 and 2001 is from ERCOT Capacity, Demand and Reserve reports for those years. Data for 2002-2006 came from "Load Participation in ERCOT Ancillary Services Markets", April 18, 2006, AESP Brown Bag Seminar by Steve Krein, ERCOT staff.

Some large industrials are curtailing their electricity purchases in response to wholesale prices outside of any formal ERCOT market or program (and not in response to ISO requests or frequency drops). [12] Yet, many of these loads would prefer to formally provide their interruption to ERCOT, if an attractive program or market opportunity were available. In the coming months, the PUCT will consider proposals for new demand response programs. Some of these proposals are similar to the emergency curtailment programs established in other markets, while others would create new ancillary services to better reflect the types of operational benefits that can be provided by loads.

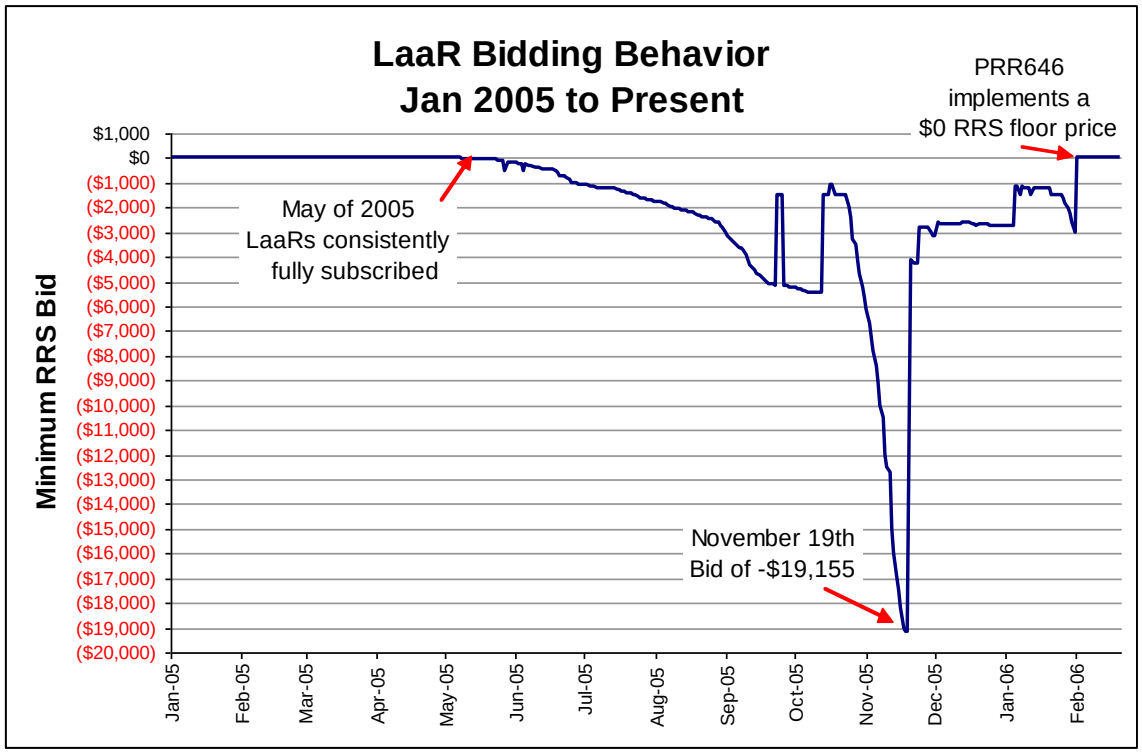
[12] Jay Zarnikau, Greg Landreth, Ian Hallett, and Subal Kumbhakar, "Industrial Energy Consumer Response to Wholesale Prices in the Restructured Texas Electricity Market," Draft, February 2005. Available at: http://www.frontierassoc.com/links.shtml.

The 'Negative Bidding' Problem

LaaRs are typically price-takers, since generators tend to set the market price for Responsive Reserves. As competition among LaaRs has intensified, many LaaRs began offering their interruption capability at increasingly-negative prices in hopes of being selecting, with the assumption that a higher-price generation resource would set the market-clearing price which all selected resources receive.

Concerns arose regarding what would happen if the market actually cleared at a price of -$19,155 per MW! All of the selected resources would then have to *pay* the market and the costs could be substantial. This potential credit risk led to the implementation of a temporary floor price that prohibited negative bids. The current prevailing proposal is to continue to prohibit negative bids by LaaRs until ERCOT adopts a nodal structure. At that time, separate markets may be established for LaaRs and generators providing Responsive Reserves.

Figure 4



Source: ERCOT staff.

Resource Planning

ERCOT has established general criteria to assist in determining the treatment of various demand-side resources into reserve margin calculations.[13] The PUCT's decision to pursue an "energy only" resource adequacy mechanism involving an increase in the wholesale price caps and reliance on prices to signal scarcity places greater emphasis on demand response to meet planning goals. Additional demand response programs will be explored shortly.

Transition to a Nodal Market Structure

The transition of ERCOT from its present zonal structure to a nodal market structure introduces further challenges. While the introduction of a day-ahead energy market may open up new demand response opportunities for loads that can accurately predict their energy requirements for each 5-minute or 15-minute period of the following day, loads without such foresight may be adversely affected by the introduction of new penalties (the "RUC Capacity Short Charge") to encourage the demand served by a scheduling entity to adhere to scheduled or forecasted load levels. While this change may make life easier for system operators, it will discourage loads from responding to price signals in real time unless price-responsive loads are shielded from such penalties. Also, the elimination of any advance notification prior to the calculation of binding prices will thwart demand response. Further complications arise from the use of zonal prices to settle energy purchases, while zonal prices may be used to establish the value of a resource in the market.

In Addition

This is by no means an exhaustive list of the challenges ahead.[14] It has also proven extremely difficult to introduce load management programs (e.g., direct load control) involving smaller energy consumers. The deployment of advanced metering equipment to enable demand response has been hampered over concerns that competitive metering will lead to stranded metering costs.

In Summary

Among North America's power markets, the restructured Electric Reliability Council of Texas (ERCOT) electricity market has probably been the most successful in promoting demand response in ancillary services markets. With a large base of industrial load that can be interrupted with limited economic cost and which was historically served through interrupted tariffs, the restructured ERCOT market has succeed in integrating large amounts of interruptible load into its market for Responsive Reserves. Yet a large number of technical and policy challenges must be resolved before Texas to fully benefit from its demand-side resource potential.

---

[13] ERCOT Generation Adequacy Task Force, "Report of the Generation Adequacy Task Force to the ERCOT Technical Advisory Committee: The ERCOT Reserve Margin Calculation," May 2005.

[14] See also the comments of Energy Data Source LP and other parties in PUCT Project No. 31972: Rulemaking Concerning Resource Adequacy and Market Power in the Electric Reliability Council of Texas Region.