

New England Demand Response Initiative

DEMAND SIDE RESOURCES AND RELIABILITY

Framing Paper #2
Eric Hirst & Richard Cowart

March 20, 2002

ACRONYMS	2
I. INTRODUCTION	3
A. Overview	3
B. Emerging FERC Policy	3
C. Policy and Program Issues for NEDRI	4
II. RELIABILITY RULES AND PRACTICES - - THE POTENTIAL ROLE OF DEMAND-SIDE RESOURCES	6
(A) Reliability Fundamentals.....	6
(B) Demand-Side Resources for Reliability – Opportunities and Barriers.....	7
III. DEMAND-SIDE RESOURCES IN WHOLESALE ELECTRICITY MARKETS	13
(A) Energy Markets: Day-Ahead and Real-Time	14
(B) Ancillary Services.....	16
(C) Installed Capability	17
(D) Participation of Retail Loads in Reliability Service Markets.....	18
IV. DEMAND-SIDE RELIABILITY RESOURCES -- RECENT ISO EXPERIENCE	22
(A) Program Designs	23
(B) Summer 2001 Results:.....	24
(C) Program Experience Summary.....	25
V. POLICY AND PROGRAM ISSUES	26
Policy Options and Topics for Discussion.....	26
REFERENCES.....	31

ACRONYMS

ALM	Active Load Management
BUG	Back-up Generators
CBL	Customer Baseline
CCM	Congestion Cost Multiplier
CSP	Curtailment Service Provider
DAM	Day-Ahead Energy Market
DAPRP	Day Ahead Price Response Program
DRR	Demand Response Resources
FERC	Federal Energy Regulatory Commission
ICAP	Installed Capability
ISO	Independent System Operator
ISO-NE	Independent System Operator – New England
LBMP	Locationally-based Marginal Price
LSE	Load serving entity
MCP	Energy market clearing price
NEDRI	New England Demand Response Initiative
NERC	North American Electric Reliability Council
NYISO DADRP	New York ISO Day-Ahead Demand Response Program
NYISO EDRP	New York ISO Emergency Demand Response Program
PCLB	Day-ahead Price-capped Load Bidding
PJM (ISO)	Pennsylvania, New Jersey, Maryland ISO
PRL	Price-Responsive Load
RES	Competitive Retail Energy Supplier (or Energy Service Provider)
RTO	Regional Transmission Organization
SMD	Standard Market Design

I. INTRODUCTION

A. Overview

This paper focuses on the role of demand-side resources in providing network services (reliability and system benefits to the grid and its interconnected users) rather than purely economic gains to participating end-users. Of course, purely economic programs (e.g., demand-side backs in response to price, and real-time pricing at retail) also have reliability and power quality impacts, and “reliability programs” also have price impacts, but these two types of programs can be distinguished in two ways:

(1) **They differ in their principal purposes.** Framing Paper #1 focuses on the role of demand-side resources in establishing electricity market clearing prices, and in adjusting real-time prices through adjustments in load. This paper focuses on the role of demand-side resources in providing reliability services and benefits to the grid as a whole.

(2) **The decision-makers and decision-making processes are different.** Because reliability is in many respects a public good, provided to a great many interconnected users in a control area, reliability standards are imposed administratively, rather than a matter of choice by individual market participants. The costs of reliability services are imposed on market participants by rule, often through tariffs, uplift, or other collective means.

B. Emerging FERC Policy

In its recently-released “Standard Market Design”(SMD) Working Paper, FERC has directly addressed regulation services and operating reserves, including their potential provision by demand-side resources. FERC distinguishes the provision of these services from the operation of the basic price-based energy markets, while recognizing that there are very important linkages between them. In relevant part, the SMD paper states:

General Principles for Standard Market Design:

6. Market rules must be technology- and fuel-neutral. They must not unduly bias the choice between demand or supply sources nor provide competitive advantages or disadvantages to large or small demand or supply sources. *Demand resources and intermittent supply resources should be able to participate fully in energy, ancillary services and capacity markets.*

Regulation and Operating Reserves to Meet Reliability Requirements

Transmission providers must ensure that ancillary services, including regulation and operating reserves, are provided. Regulation provides moment-by-moment balancing of generation and load on the system. Operating reserves ensure reliable service by covering contingencies such as the failure of a supply source or a transmission line. Order No. 888 envisioned that these would be provided as a tariff service subject to a cost-based

rate. With the establishment of markets to provide balancing services, a more market-oriented approach is needed for regulation and operating reserves. (Other ancillary services, such as reactive power, would continue to be procured much as they are today.)....

General Features

1. The LSE has the responsibility to procure regulation and operating reserves or pay for the regulation and operating reserves procured by the transmission provider on its behalf.
2. Suppliers of regulation and operating reserves must meet specific operational requirements to provide these services.... *Demand must have the opportunity to supply operating reserves if it meets the necessary operational requirements (which should be designed to enable demand response participation).*

Scheduling and Bidding Rules

6. The transmission provider must procure regulation and operating reserves through a bid-based auction for all those who do not self-supply....
7. *Demand-side supply of operating reserves must have non-discriminatory bidding opportunities in the market.*
8. Regulation and operating reserve markets *must allow sellers to submit availability bids in addition to energy bids.* The availability bid allows the bidder to specify the minimum payment that it requires to be available to provide regulation and operating reserves.¹

The general outlines of the SMD working document have been endorsed by the full Commission, and it is likely to form the framework for a comprehensive rulemaking on Standard Market Design in the next few months. Thus, there is clear support from FERC for the effort to design reliability rules and markets that tap the value of demand-side options where appropriate.

C. Policy and Program Issues for NEDRI

As a general matter, it is useful for the reader to consider four different ways in which demand-side resources may make a significant contribution to reliability:

- (1) Routine system operations -- demand-side services can compete with supply-side resources in ancillary service markets;
- (2) Emergency operations -- system operators can purchase load reductions as well as emergency generation to balance the system under high load or in the event of contingencies;
- (3) Demand-side resources can address system adequacy needs on a longer-term basis by providing services in installed capacity markets; and
- (4) Demand-side resources can improve reliability of transmission and distribution

¹ Federal Energy Regulatory Commission, "Working Paper on Standardized Transmission Service and Wholesale Electric Market Design," March 15, 2002 (excerpts from pages 5-6, and 18-19, emphasis added).

systems by relieving congestion and reducing overloaded circuits and substations.

This paper addresses the first 3 of these topics; the role of demand management in mitigating or resolving transmission and distribution challenges will be taken up at a later date.²

This paper is organized into 5 sections. The next section discusses current North American reliability rules and shows how these requirements have historically discriminated against loads in the provision of reliability services. Section 3 explains the wholesale markets that today's independent system operators (ISOs) run and that tomorrow's regional transmission organizations (RTOs) will likely manage. These markets include the day-ahead and real-time energy markets (which are covered in much more detail in Framing Paper #1) as well as the markets for ancillary services, real-time balancing, and long-term installed capability, which have important reliability attributes discussed here. Section 4 briefly reviews the experience of the three Northeastern ISOs (New England, New York, and PJM) with regional demand-side reliability programs over the past two summers.

The final section, Section V, sets out a number of program and policy questions and options to consider for the use of retail load response to provide bulk-power reliability services. It poses the following questions:

- 1) How should ancillary service markets be organized in New England to realize the potential value of demand-side resources?
- 2) Should customers be permitted to contract with the ISO for reliability load response programs (directly or through a Curtailment Service Provider), or must those arrangements be made through LSEs?
- 3) How can regional DR reliability programs be coordinated with legacy interruptible contracts operated by LSEs?
- 4) How should demand resources be integrated into "installed capacity" markets?
- 5) Can reliability-oriented programs be designed to recognize the unique characteristics and needs of loads, just as they do now for generators?
- 6) What types of demand-side resources should be eligible to participate in "emergency" load response programs?

² The potential role of demand resources in addressing transmission and distribution reliability challenges was discussed by the NEDRI Stakeholders at their initial meeting, and there was strong support for addressing these topics in this process. There was a preference, however, to reserve these topics for a later discussion, and to combine the discussions of price-responsive load and reliability-oriented demand response in Framing Papers #1 and #2.

- (A) What is the role of and/or limits on use of diesel-fired back-up generators?
 (B) Will ISO rules permit participation and compensation to aggregations of small, widespread load responses resources?
- 7) What actions should state regulators take to enhance customer participation in reliability-focused DR programs?
- Should state regulators mandate interval meters, at least for larger customers?
 - Should PUCs redefine Standard Offer service, at least for larger customers, as a service with interval meters and real-time pricing?
 - Should interval meters be required and/or subsidized?
- 8) Should the ISO and state PUCs support development of Curtailment Service Providers to accelerate and deepen DR for reliability?
- 9) What target levels, if any, should New England adopt as a benchmark level of success for reliability-oriented DR programs?
- 10) What lessons should ISO-NE and Stakeholders in New England take from the experiences of New York and PJM with demand-side response for system reliability?

II. RELIABILITY RULES AND PRACTICES - - THE POTENTIAL ROLE OF DEMAND-SIDE RESOURCES

(A) Reliability Fundamentals

Bulk-power systems are fundamentally different from other large infrastructure systems, such as air-traffic control centers, natural-gas pipelines, and long-distance telephone networks. Electric systems have two unique characteristics:

- The need for continuous and near instantaneous balancing of generation and load, consistent with transmission-network constraints: this requires metering, computing, telecommunications, and control equipment to monitor loads, generation, and the transmission system, and to adjust generation output to match load.
- The transmission network is primarily passive, with few “control valves” or “booster pumps” to regulate electrical flows on individual lines: control actions are limited primarily to adjusting generation output and to opening and closing switches to add or remove transmission lines from service.

Because of these two characteristics, bulk-power system operators rely primarily on changes in generation output (MW movements up or down) to keep the system in balance and to comply with transmission limits. Operators go to great expense to avoid involuntary changes in load (“unplanned outages,” or blackouts) as a means of balancing or protecting the system. In

principle, *voluntary* changes in electricity consumption could serve as well as generator movements in meeting reliability requirements, but the use of customer loads for reliability purposes has been the exception rather than the rule.

The North American Electric Reliability Council (NERC) was organized by the electric utility industry in 1968, a part of the industry's response to the Northeast Blackout of 1965. NERC establishes reliability standards, guidelines, and criteria addressing many aspects of power system operation. NERC defines reliability as "the degree to which the performance of the elements of [the electrical] system results in power being delivered to consumers within accepted standards and in the amount desired." NERC's definition of reliability encompasses two concepts, *adequacy* and *security*. *Adequacy* is defined as "the ability of the system to supply the aggregate electric power and energy requirements of the consumers at all times." *Security* is "the ability of the system to withstand sudden disturbances." In plain language, adequacy standards require that there be sufficient generation and transmission resources available to meet projected needs plus reserves for contingencies. Security standards require actions by system operators to ensure that the system will remain intact even after outages or other equipment failures occur.

The traditional, vertically integrated utility managed short-term reliability by dispatching its own generating units as well as adjusting transformer settings and turning breakers on and off at its own transmission facilities. In competitive wholesale markets, system operators increasingly work for entities that own no generation and, in many cases, own no transmission. In such cases, the system operator must establish markets for the generation reliability services and negotiate contracts with transmission owners for other reliability services.

This change in industry structure and the associated emergence of wholesale energy and reliability markets create new opportunities for demand-side resources. If the market rules are technology neutral (i.e., they focus on the function to be performed and not on how that function is done), customer loads will be able to participate in these markets. Such participation will either enhance reliability or lower the costs of maintaining reliability for all customers (by deepening reliability markets) and will save money for participating customers.

(B) Demand-Side Resources for Reliability – Opportunities and Barriers

As noted above, reliability standards, rules, and traditions play an important role in power system reliability. While utility adherence to NERC standards is legally voluntary³, these standards are routinely followed, and have a great deal of influence throughout the industry, including operations at ISO-New England. Thus, it is important to consider the degree to which those rules invite or impede efforts to mobilize Demand Response Resources (DRR) to support reliability goals. As a general matter, many NERC standards could be interpreted to permit the use of demand-side resources for reliability purposes, but in other respects both the standards

³ In response to reliability challenges and crises in several regions in recent years, NERC and many industry stakeholders developed proposed federal legislation that would create a new reliability organization, the North American Electric Reliability Organization, and make reliability standards mandatory throughout the industry. Proposals to that effect are presently pending in Congress.

themselves and the traditions of interpretation within the industry will have to change to permit the robust use of DRR. The example below illustrates this challenge.

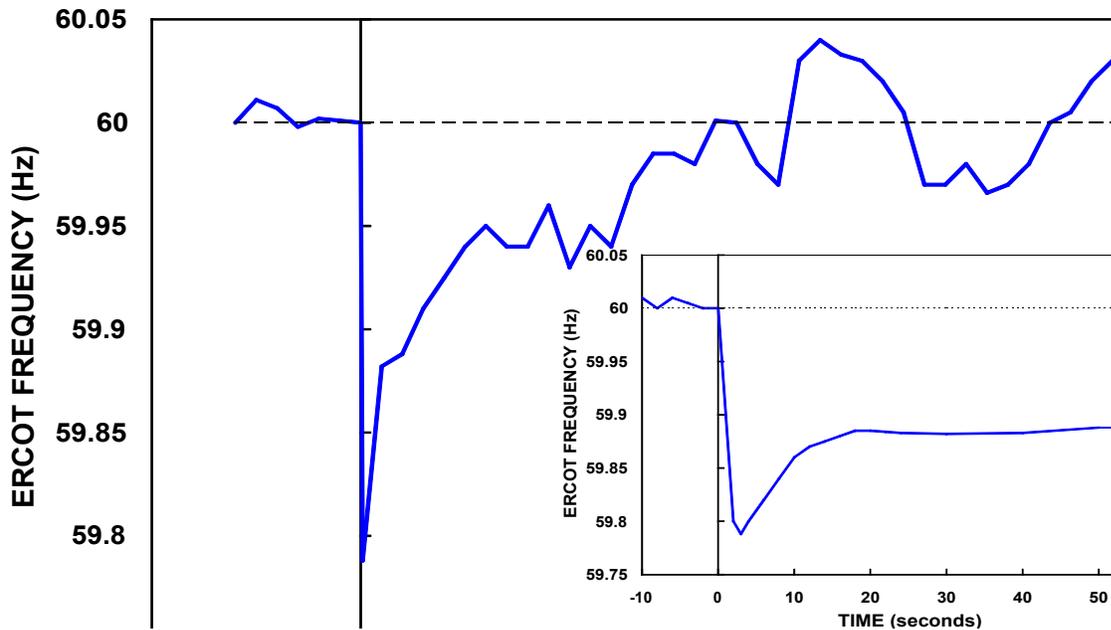


Figure 1: Interconnection frequency before and after the loss of a 653-MW generator. The inset shows frequency for the first minute after the outage, and the larger figure shows frequency for the first 20 minutes after the outage.

Responding to a major generation outage provides an example of how the electricity industry responds to its unique features. Figure 1 illustrates how the electric system operates when a major generating unit suddenly fails. Prior to the outage, system frequency is very close to its 60-Hz (cycles per second) reference value. Generally, within a second after the outage occurs, frequency drops, in this case to 59.79 Hz. The frequency decline is arrested primarily because many electrical loads (such as motors) vary with system frequency. If the frequency decline is large enough, the governors at those generators so equipped sense the frequency decline and open valves on the turbines, which rapidly increases generator output. This governor response accounts for the initial increase in frequency during the first several seconds after the outage occurs, as shown in the Figure 1 inset. At this point, the generating units that provide contingency reserves, in response to signals from the control center, begin to increase output. More fuel is added to the boiler, leading to a higher rate of steam production, which leads to higher power output. In this example, the system worked as it was intended to, and frequency was restored to its pre-contingency 60-Hz reference value within the required 10 minutes (at 8.5 minutes).⁴

⁴ In early 2000, NERC extended the allowable disturbance-recovery period from 10 to 15 minutes.

A reasonable question to ask, in reviewing this example and Figure 1, is whether this reliability problem could have been solved, at least in part, by a reduction in load. The answer, of course, is yes—in principle. We write “in principle” because in practice there are numerous barriers to the active engagement of customer load response to address a situation like that in the example.

(1) Demand-side resources as reserves

First, how might DRR help to address the reliability challenge in the illustration – the unplanned loss of a major generating unit? This is the purpose of system reserves. As a general matter, the total reserve requirement consists of three parts: spinning reserve, supplemental reserve, and replacement reserve (These are given different names in different systems. In New England, the NEPOOL Rules call for 10 minute spinning reserve, 10-minute non-spinning reserve, and 30-minute non-spinning reserve.) What types of resources might best provide those services?

Older steam units are very flexible at providing spinning reserve. Valves can be opened from a pressure vessel, and units can be operated for short periods above their normal maximum rating. This allows a quick response to a demand for spinning reserves. Since the units have relatively flat heat rate curves over a wide operating range, it is also possible to operate them significantly below their maximum capacity, with the ability to come up to full capacity very quickly. Because they are "hot" and synchronized to the grid, they meet the definition of "spinning reserves." However, they also tend to have relatively high pollution output, poor fuel efficiency when running at part or full output, and high fixed operating costs.

Modern combined cycle have very different operating characteristics. Combined cycle units typically cannot be operated below about 70% of maximum capacity and still have the ability to come up to full power quickly. Because of their low heat rates, they are very attractive to run as baseload resource, and the opportunity cost of using them for reserves is much greater than for steam units⁵. New CCCT units can have very low pollution output, relatively good fuel efficiency when running at full capacity, and very low fixed operating costs.

Simple cycle turbines also have different operating characteristics. They can be brought from cold standby to full power in 15 minutes or less.⁶ A "cold" unit, therefore, does not meet the conditions for "spinning reserve" but does meet the conditions for longer-term reserves.

Because of pollution control requirements, the higher labor costs associated with maintaining steam units, and other economic criteria, simple and combined cycle units are increasingly attractive as utility system resources, and steam units increasingly unattractive. This creates a fairly serious problem with respect to provision of spinning reserves: the preferred resources being added to the grid today do not "spin" as well as the older resources.

Demand-side resources offer a unique opportunity to meet this need. In order to allow a "cold" simple cycle CT or a longer-term load management program to provide required reserves, some

⁵ With a steam plant, the owner avoids 10,000 - 14,000 btu of fuel cost and loses 1 kwh of revenue; with a CCCT, the owner avoids 7,000 btu of fuel cost, and foregoes the same revenue.

⁶ GE Web Site for LM-6000.

way of bridging the gap between the experience of a contingency, and the 5 - 15 minute response time of a turbine. There are many end-uses which may not be particularly good candidates for frequent longer-term interruptions (say, 4-6 hours) which can be curtailed for this duration without significant adverse impact on consumers.

Examples include: Residential water heating, and space conditioning commercial space conditioning, a fraction of commercial and retail lighting, and certain low-visibility functions like municipal water system pumping.

On a typical utility grid, these loads can amount to 20 - 40% of total usage. Loads of this type could be interrupted for 5 - 5 minutes without a significant burden to consumers. By installing any combination of underfrequency control and/or radio or ripple control to these types of end-uses, it may be possible to provide spinning reserves without the economic cost or environmental impacts of fuel-burning resources. If a sufficient amount of cold quick-start resources or longer-term load management resources are also available, then the curtailment of these loads can be limited to the duration needed to bring a cold resource up to full output. In this manner, either supply or demand-side resources which cannot provide spinning reserve economically (e.g., combustion turbines or industrial load interruption contracts) can be used in conjunction with demand-side short-duration curtailment to provide both spinning and operating reserves.

(2) Barriers to DRR deployment

Provision of reserves in New England is costly in both financial and environmental terms, but the deployment of DRR options is limited. In part, this is simply a matter of tradition, but there are a number of other barriers to consider. Many of those barriers are noted in Framing Paper #1, and will not be repeated here. However, particular problems related to NERC standards and traditions for reliability services are noted briefly below.

NERC Policies: One embedded problem is that existing NERC Policies inappropriately favor generation resources over customer loads in the provision of reliability (ancillary) services. Consider, as an example, NERC's (2001c) Policy 1 — Generation Control and Performance.⁷ This policy deals with the generation:load balance required under normal operating conditions and under emergency (contingency) conditions. The Policy refers to three kinds of reserves used to respond to a major contingency (e.g., loss of a large generator or major transmission line): frequency response, spinning reserve, and supplemental reserve. The discussion of frequency response deals only with generating units with governors (e.g., governor droop, deadband, and limits). No mention is made of loads providing frequency response.

Policy 1 defines spinning reserve as “*unloaded generation* that is synchronized and ready to serve additional demand” (emphasis added). Clearly, this statement excludes customer loads from providing this valuable and expensive ancillary service. NERC's definition of nonspinning reserve, on the other hand, does allow for the use of loads to provide this service: “that operating

⁷ Note that this policy deals with *generation* and not *demand*.

reserve not connected to the system but capable of serving demand within a specified time, or interruptible load that can be removed from the system in a specified time.” Unfortunately, NERC’s definition of interruptible load is rather narrow: “demand that can be interrupted by direct action of the supplying system’s system operator in accordance with contractual provisions.”

Metering and telemetry requirements. In 1993, NERC (1993) issued a reference document that explains how customers loads can substitute for some reliability services traditionally provided by generators, especially operating reserves. This paper distinguishes between indirect (passive) and direct (active) load management, with the latter involving the ability of the system operator to take action to disconnect the load. The extent to which NERC considers changes in customer loads a reliability resource depends on the ability of the system operator to control that load change, either directly through automatic controls (e.g., to turn off electric water heaters) or indirectly (e.g., through a phone call to a facilities manager). In addition, this document suggests that the system operator must know what the load is both before and after the exercise of control actions.

This statement on the information that must be made available to the system operator appears to require that (1) each controlled load be individually metered at, say, the 1-minute level and (2) the load’s electricity consumption be telemetered to the control center in real time. Such requirements – for individual real-time meters and for direct control telemetry to the systems operations center --make small-scale, distributed DRR options virtually impossible to provide. An alternative would permit aggregation of enough loads so that the system operator need see only the combined response of all the loads. Such aggregation is especially important for residential loads that might provide contingency reserves because it might be too expensive to install interval meters and the associated communications systems on individual appliances. A single residential water heater might provide only about 0.005 MW of load reduction. However, an aggregation of 10,000 water heaters could provide 50 MW of load relief, enough to be visible to the system operator without any special metering or communications equipment.

Economic consequences of the exclusions. Distinctions between the treatment of supply-side and demand-side resources in NERC standards and industry practice have the effect of raising the cost of reliability services, and undermining the potential for DRR to provide them. NERC’s Policy 1 currently requires that “at least 50% of operating reserve shall be Spinning Reserve.” Spinning reserve is much more expensive to provide than non-spinning reserve, and it is paid higher prices⁸ in power markets. (See Figure 2). If loads were permitted to supply spinning reserve, they would have more incentive to participate in markets for ancillary services

⁸ In New York, the price of spinning reserve was, on average, 50% higher than the price of nonspinning reserve (\$3.0 vs \$2.0/MW-hr) over the 20-month period from April 2000 through November 2001 (Fig. 2). In ERCOT, the price of responsive (spinning) reserve averaged \$6.4/MW-hr from September through November 2001, compared with only \$1.4/MW-hr for nonspinning reserve. Finally, the prices for spinning and nonspinning reserves in California for 1999 averaged \$6.5 and \$3.6/MW-hr, respectively. Clearly, spinning reserve is a more valuable, and therefore more expensive service. Why should the demand side be precluded from participating in these more lucrative markets? More important, why should certain resources be prohibited from performing these valuable reliability functions?

and the prices for spinning reserves would decline. This change would enhance reliability by providing more resources for contingency reserves and would save money for electricity consumers by lowering the costs for these reserves.

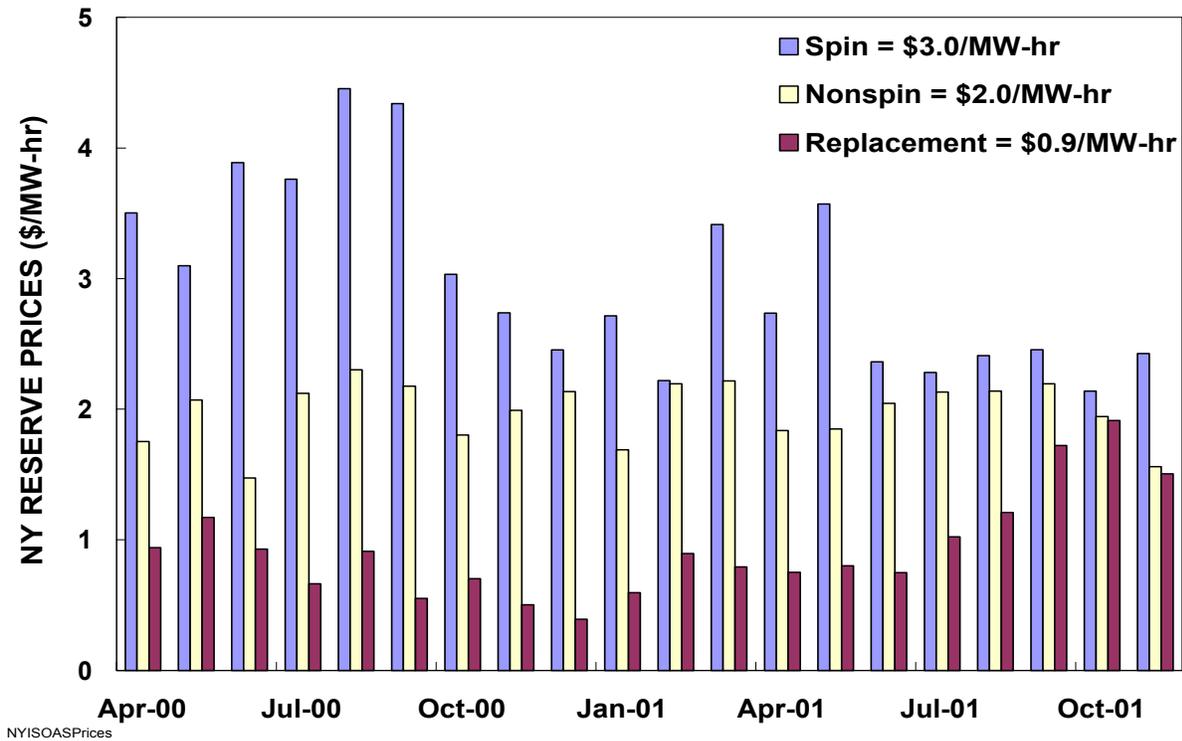


Figure 2: Prices for operating reserves in New York from April 2000 through November 2001.

Fortunately, NERC has recognized these limitations in its current operating policies, and is now considering amendments that would open up opportunities for DRR. A recent draft revision to Policy 1 would define spinning reserve as “the Resource Capacity in excess of current and anticipated demand that is synchronized to the grid and deployable” including “controllable load resources.” Note that this definition is technology neutral; it does not specify whether the required capacity comes from generation or customer loads, a welcome modification. (NERC 2001b)

III. DEMAND-SIDE RESOURCES IN WHOLESALE ELECTRICITY MARKETS

Wholesale electricity markets typically include long-term markets for transmission rights (either financial or physical) and installed generation capability as well as short-term markets for energy, ancillary services, and congestion. Framing Paper #1 addresses price-responsive DRR in the principal energy markets and includes a table of the principal markets for energy and capacity. This section focuses on a subset of those markets – on demand-side options to provide reliability services. The principal reliability services are set out in Table 1 below, and discussed in the text following.

Table 1. Wholesale markets for ancillary services and transmission congestion relief

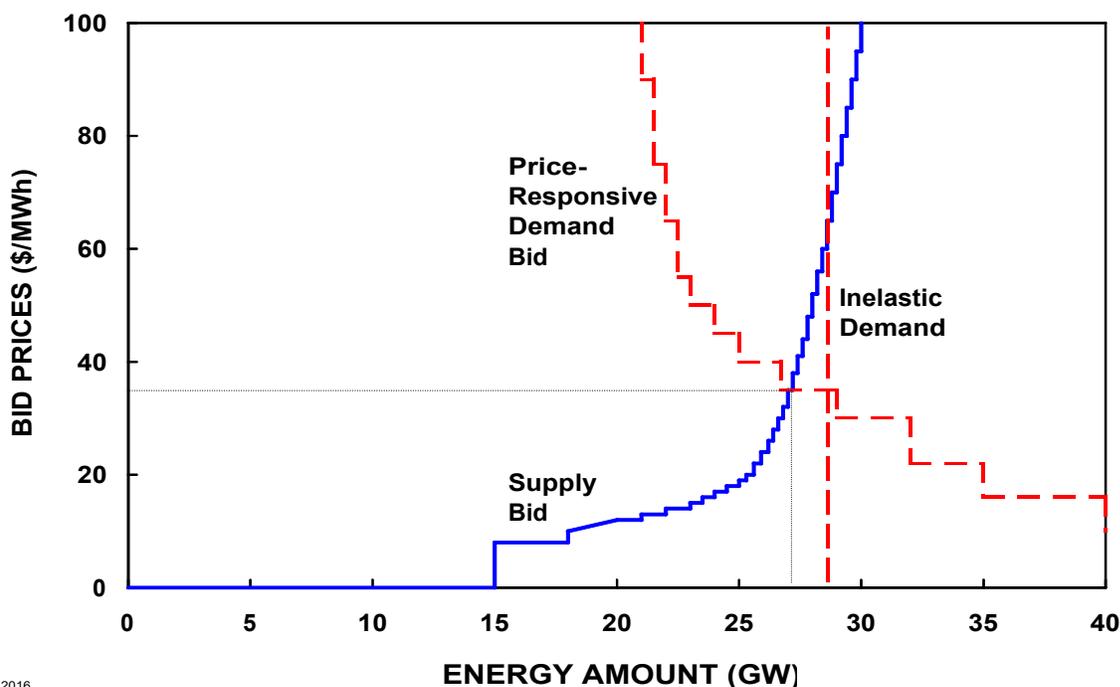
Market	Description	Demand Response Resource
Day-ahead ancillary services	Potential suppliers submit capacity (\$/MW-hr) and energy (\$/MWh) bids to supply service on an hourly basis. These resources are required to meet NERC security requirements in real time	DRR providers could submit bids for many of these services
Regulation^a	Operating facilities with automatic generation control, that can respond rapidly to system-operator requests for up and down movements; used to track the minute-to-minute fluctuations in system load and to correct for unintended fluctuations in generator output to comply with NERC's Control Performance Standard (CPS)	Difficult (but not impossible) to provide through DRR
10-Minute Spinning Reserve or "Spinning Reserve" ^a	Generators online, synchronized to the grid, that can increase output immediately in response to a major generator or transmission outage and can reach full output within 15 minutes to comply with NERC's Disturbance Control Standard (DCS).	Rapidly-adjustable loads with automatic controls can provide; direct telemetry for larger (e.g. pumping) loads, radio or ripple for water heating, air conditioning, etc.
10-Minute Non-Spinning Reserve or "Supplemental Reserve" ^a	Same as spinning reserve, but need not respond <i>immediately</i> , therefore units can be offline but still must be capable of reaching full output within the required time	Same as above
30-Minute Non-Spinning Reserve or "Replacement Reserve"	Same as supplemental reserve, but with a 30- or 60-minute response time, used to restore spinning and supplemental reserves to their precontingency status. Dispatched when reserve shortfalls are forecast or imminent	Larger industrial and commercial loads on contract, requiring lead times to adjust. Customers paid either price floors (e.g., \$500/MWh) or market clearing price.
Installed Capacity (ICAP)	LSE required to procure capacity call options equal to their load-serving obligations; generators selling this service are typically obliged to bid that resource amount into the market each day and to be available when called. Difficulties in defining product and ensuring availability.	Load could provide the same resource; "capacity payments" for availability would enhance value of DRR options generally.

^a Terms generally used in New England or set out in the NEPOOL Rules are in **bold print**. The three ancillary service terms in quotation marks are generic terms used by FERC its Order-888 open-access transmission requirements. (FERC, 1996)

(A) Energy Markets: Day-Ahead and Real-Time

Although most electricity is bought and sold under long-term bilateral contracts, perhaps 5 to 10% will be traded either day-ahead or in real-time. These short-term trades could be a consequence of changed circumstances (e.g., a competitive retail provider signed up more customers than it anticipated, a generator completed its planned maintenance outage faster than it expected to, a generator suddenly trips offline, or the weather is different from what was expected) or they could be part of a company's risk-management strategy (maintaining a portfolio of long- and short-term options).

Most such markets today operate with virtually no price-responsive demand. That is, the system operator treats demand as a fixed quantity each hour and optimizes across the generator bids only. Figure 3 illustrates the two situations. The solid curve that slopes up to the right represents the combined bids from all the generators, showing the typical increase in bid prices associated with increased output. The vertical dashed line reflects the situation that typically occurs today, in which customer demand is specified each hour independent of the price of electricity. The dashed curve that slopes up to the left represents the combined bids from all retail loads that would occur in a fully functioning wholesale day-ahead electricity market. In this case, retail customers or their retail-service providers would specify how much electricity they would use at various prices. In both demand cases, the intersection of the supply and demand curves determines the amount of electricity to be produced and consumed each hour (ignoring losses) as well as the market-clearing price for that hour.



2016

Figure 3: Hypothetical generation-supply curve and two demand curves. The vertical demand curve reflects the typical situation in today’s wholesale markets, in which retail demand is fixed (independent of electricity price). The dashed line represents the situation in which retail customers bid in different amounts of load as a function of the price.

Details of the operation of the day-ahead and real-time energy markets are addressed in Framing Paper #1, on economic price-responsive load. We address the topic in connection with reliability to make three straightforward observations:

- (1) Although the addition of price-responsive demand to the day-ahead market has strong positive implications for economic efficiency (including a reduction in the ability of generators to exercise market power), it also has reliability benefits. In efficient markets, prices are high when reliability is threatened. To the extent demand is

- reduced during high-priced periods, reliability is improved because additional supplies are available to meet any reliability contingencies that might occur. The hypothetical example in Figure 3 shows a reduction in demand from 29 GW with inelastic demand to 27 GW with price-responsive demand, a release of 2 GW of generating capacity that can be called upon in an emergency.
- (2) In addition to freeing up capacity, suitably located demand reductions bid into the day-ahead market can also relieve potential congestion problems on the transmission system.
 - (3) The demand-supply relationships shown in Figure 3 applies to the markets for reliability services just as it does for energy. When demand-side resources can compete with traditional supply-side resources to provide reliability services to the wholesale market, the additional available resources can improve diversity and lower system costs.

(B) Ancillary Services

In addition to operating day-ahead and “day-of” or “real-time” energy markets, RTOs also run markets for certain real-power ancillary services (Table 1).⁹ These services are required to respond to the two unique characteristics of bulk-power electric systems noted above, the need to maintain generation:load balance in near-real-time and the need to redispatch generation (or load) to manage power flows through individual transmission facilities.

Regulation. NERC’s Policy 1 on “Generation Control and Performance” specifies two standards that control areas must meet to maintain reliability in real time. The Control Performance Standard (CPS) covers normal operations and the Disturbance Control Standard (DCS) covers recovery from major generator or transmission outages. The regulation ancillary service is the primary resource system operators use to meet the CPS requirements. In principle, customer loads could provide the service as well as generators. Because provision of this service requires a change in output (or consumption) on a minute-to-minute basis and, therefore, requires special automatic generation control equipment at the generator (or customer facility), it seems unlikely that retail loads will want to provide this service. Therefore, we do not discuss regulation further.

Reserves. The three reserve services listed in Table 1 are all intended to help control-area operators meet the DCS. Briefly, DCS requires that the system recover from a major outage within 15 minutes. The three reserve services provide responses of different quality. **Spinning reserve** is the most valuable, and therefore generally the most expensive, service because it requires the generator to be on line and synchronized to the grid. Because such generators are online, they can begin responding to a contingency immediately; that is, their governors sense the drop in Interconnection frequency associated with the outage and begin to increase output

⁹ Ancillary services are those functions performed by the equipment and people that generate, control, and transmit electricity in support of the basic services of generating capacity, energy supply, and power delivery (Hirst and Kirby 1998).

within a second (Fig. 1). **Supplemental reserve**, which could include generators that are already online, is less valuable because it does not necessarily provide an *immediate* response to an outage. **Replacement reserve** is less valuable still because it need not respond fully until 30 or 60 minutes after being called upon. Replacement reserves are used to permit the restoration of the 15-minute contingency reserves so that these faster-acting resources are, once again, able to respond to a new emergency.

(C) Installed Capability

The markets discussed so far in this section deal with the short term, from day ahead to actual operations. From a reliability perspective, these markets help system operators maintain security. The long-term equivalent, needed to create adequacy, includes markets for installed generating capability¹⁰ and transmission rights. Traditionally, vertically integrated utilities built or bought the rights to enough generation to meet a loss-of-load probability of not more than one day in ten years. This criterion ensured that the utility had enough generating capacity to meet peak demand with a probability of 99.97%. (One divided by ten times 365 is 0.000274.)

The installed-capability requirements and markets implemented by the three Northeastern ISOs have all experienced problems. A fundamental problem with the requirement is the lack of a tangible product. The simple existence of “iron in the ground,” which could include a generator that is not able to produce energy (i.e., is unavailable), is of no value during a reliability emergency. A more useful product would be the *right to convert installed capacity into energy* at a predetermined strike price under certain conditions. Instead of recognizing MW of installed (or even unforced) capacity as a reliability product, system operators could require load-serving entities to obtain the rights to energy on demand, a true option. Another possibility is to replace installed-capability requirements with long-term contracts for contingency reserves. Developing demand-side analogs to installed capability will be difficult until the need for and definition of installed capability are clarified.

Many utilities offer their large industrial and commercial customers interruptible rates. These programs typically offer a discount in the demand charge (expressed in \$/kW-month) in exchange for the right to interrupt service to a portion of the customer’s load. These programs are characterized by a rigid structure that specifies months or years in advance the maximum number of times a year the utility can call for interruptions, the minimum amount of advance notice it must provide, the maximum time permitted for each interruption, and the penalty imposed on customers who do not meet their contractual obligation to interrupt demand when called upon to do so. In many cases, these programs were discounts in disguise and were never intended to be used for reliability purposes.

¹⁰ Considerable disagreement exists over the value of an installed-capability requirement in competitive markets; see Hirst and Hadley (1999) and Hobbs, Inon, and Stoft (2001) for perspectives on these issues. See ISO New England (2001) for a discussion of the issues and problems in New England.

The California electricity crisis of 2000 and 2001 demonstrates well the problems that can occur with these traditional utility interruptible-load programs (California Public Utilities Commission 2001). Although the three California utilities had been paying industrial and large commercial customers more than \$220 million a year for interruption rights, when the California electricity crisis occurred, actual operation of these programs “identified serious problems.” In particular, many of the customers participating in Southern California Edison’s program did not interrupt their loads as required, resulting in a compliance rate of only 60 to 70% (achieving about 1,200 MW of load reduction instead of the 1,800 MW under contract). In addition, many customers dropped out of the programs as soon as they could, once they realized that the utilities’ contractual rights would, under emergency conditions, be exercised.

The three northeastern ISOs have installed-capability requirements and companion markets. Long-term contracts for load interruptions generally qualify as installed capability. PJM’s Active Load Management program, operated primarily by the distribution utilities, includes direct control of residential equipment, customer load reduction to a firm level (interruptible contracts), and guaranteed load drops implemented through the use of onsite generation. In this program, PJM provides no monetary payment. Instead, participating load-serving entities receive installed-capability credits for the load reductions, which reduce their costs of installed generating capacity. Participating loads must be available for up to ten PJM-initiated interruptions during the planning period (October through May and June through September), for interruptions lasting up to six hours between noon and 8 pm on weekdays, and within two hours of notification to the load-serving entity by PJM. The baseline is either the customer’s load one hour before the event or the customer’s hourly load on a comparable day, as determined by the load-serving entity. Failure to perform can lead to penalty charges related to PJM’s capacity deficiency charge; that is, the penalty is comparable to that which would apply for providing insufficient generating capacity to meet the required installed-capability requirement.

Almost 2,000 MW of load (roughly half of which is residential and small-commercial direct-load control and half of which is industrial loads and onsite generation) qualify for installed capability in PJM. The program was called upon six times during the summer of 1999, not at all during the summer of 2000, and provided 1800 MW of load relief in 2001 (PJM 2001).

(D) Participation of Retail Loads in Reliability Service Markets

The explanations of these markets suggest that retail loads should be readily able to participate in the day-ahead market for energy, the day-ahead markets for the three reserve services, and the long-term markets for installed capability. Participation in the markets for regulation and real-time energy seem much more problematical for loads because these functions require the ability to modify loads frequently (several times an hour) with only a few minutes advance notice.

Many types of load could profitably and relatively easily provide reliability services to regional power grids. Indeed, the industry and its customers have a long history of interruptible contracts, time-sensitive rates, and other working relationships to draw upon. Utilities, system managers, and ESCOs have historically worked with flexible industrial processes, large building heating and cooling systems, ripple- and radio-controlled residential air conditioners, and water heaters, and certain public and municipal loads.

Large municipal water-pumping systems provide another useful model. Such systems typically have tanks, reservoirs, or lakes to store water for later distribution to consumers (Kueck 2002). These water-storage systems are a natural energy-storage system because they permit the water-treatment system to interrupt pumping operations for up to a few hours at a time. During such nonpumping periods, gravity will ensure sufficient water flow and the appropriate pressure to consumers. The California Department of Water Resources, as an especially important example, has pumping loads of more than 1,500 MW. In aggregate, municipal water pumping accounts for about 3 to 4% of total U.S. electricity consumption, which is about as large as the nation's entire spinning-reserve requirement¹¹. Thus, municipal water systems are serious candidates for provision of spinning reserves, both because of their large storage capacity and because of their large size

Water-pumping loads are not the only ones that could provide the reserve services. Pumped-storage hydroelectric projects, when in the pumping mode, can provide large amounts of reserves very quickly¹². More generally, any customer facilities or processes with large storage capacity or that can tolerate short-term interruptions on short notice (ranging from residential water heaters, to commercial HVAC systems, to industrial processes that store energy-intensive intermediate product) are good candidates for provision of spinning and supplemental reserves.

Table 2 summarizes the requirements for retail loads to participate in the relevant markets discussed above:

Aggregation refers to a minimum size requirement the RTO might impose on all entities connected to the grid. Smaller loads wanting to participate in these markets need to be aggregated so the total exceeds the RTO's minimum size requirement.

Meters -Participation in the day-ahead energy market requires meters that can record and store hourly consumption data. Participation in the ancillary-services markets might, however, require meters that can record and store subhourly consumption data, perhaps at the 5-minute interval. The ISO or load aggregator needs this detailed information to determine whether individual customers responded to the ISO's call for the reserve service and did so within the required time. Aggregation of some load reductions, such as the use of switches to turn off water heaters, will likely not require interval meters. If enough such loads are aggregated that the load

¹¹ Permitting these municipal loads to provide (sell) spinning reserve could also improve the efficiency of local water operations. The best way to provide spinning reserve from these loads would be to install adjustable-speed drives on the large pumping motors. The use of such drives, rather than throttling valves, would maintain high levels of pumping efficiency, provide much greater control over pumping operations, and reduce maintenance costs.

In addition to providing spinning reserves, these pumping loads, depending on their locations, might also be able to provide congestion relief and voltage support if they are located in areas with highly-loaded transmission lines and low voltages Kueck (2002).

¹² For example, a pumped-storage facility in Connecticut has four 250-MW pumping motors. Turning off the pumps could provide up to 1,000 MW of reserves. In some cases, the system can switching rapidly from pumping to generation and, therefore, provide almost double the nameplate rating.

reduction is “visible” to the system operator in real time and the system operator is permitted to test this resource periodically, meters that are read monthly should suffice.^{13 14}

Communications include the transfer of information from the RTO to the customer and the transfer of information from the customer to the RTO. The former include acceptance of bids in the various markets, the associated market-clearing prices, and the real-time calls to provide the ancillary services purchased in the day-ahead markets. The latter include bids into the energy and ancillary-services markets and information on actual electricity consumption.

Advance notice refers to the amount of time the customer has to respond to a particular request. Participation in the day-ahead energy market imposes no requirement on the customer. Customers can consume as much or as little electricity in real time as they want; however, any differences between these actual amounts and those contracted for in the day-ahead market are settled at the real-time energy price rather than the day-ahead price. Customers selling load reductions as ancillary services, on the other hand, *must* comply with the RTO's request to reduce load. Depending on the service, the leadtime available is 15 minutes for the two contingency reserves or 30 or 60 minutes for replacement reserves. Failure to comply, either in magnitude or in time, will result in a penalty. At a minimum, the RTO will likely withhold the day-ahead capacity payment for the service; in addition, it might impose a penalty related to the reliability risk imposed on the system by the load's failure to comply.

There is no explicit payment for participation in the day-ahead energy market. Rather, the customer pays for the contracted amount at the day-ahead hourly price. Typically, customers selling load reductions into the reserve markets are paid for the capacity they have agreed to provide, a reservation payment in \$/MW per hour. In addition, if the load is called upon in real time to provide the reserves, it is also paid for its energy, typically at the higher of its day-ahead energy bid or the current market-clearing price for energy.

Baseline - Some demand-response programs require definition of a baseline against which the actual load is measured to determine the amount of load reduction. This concept does not apply to the day-ahead energy market. And it is straightforward to apply to the reserve markets because the amount of time between the call for the reserves and their delivery is so short. Typically, consumption during a few 10-minute periods before the call is used to define the baseline. Alternatively, the baseline is defined as the average consumption during the same hour for several previous days of the same type.

¹³ Periodic testing will determine how reliable a resource this load reduction is (in particular, what fraction of the connected load reduction actually responds to activation) and how the magnitude of this resource varies with certain factors such as weather (e.g., water heater interruptions will provide more load relief in the winter than the summer) and time (e.g., water heaters are more likely to be on early in the morning when people first get up).

¹⁴ The load aggregator must still develop an equitable method for sharing the payments from the RTO to the participating customers, perhaps seeking to recognize that not all customers will respond every time.

Penalties - Participation in markets for installed capability provide a long-term equivalent to participation in markets for operating reserves. A customer could choose to participate in an installed-capability program (such as PJM's) and receive monthly payments for the capacity it provides. In return for this stream of payments, the customer makes a long-term commitment (i.e., sells an option on interruption rights) to the supplier, permitting the supplier to take the contractual amount of load relief under prespecified conditions. These conditions typically include the number of times a year or month interruptions can be called, the minimum advance notice to be given to the retail customer, the maximum duration of the interruption, and the penalties for failure to comply.

Participating in the day-ahead reserve markets, on the other hand, involves much more modest and short-term obligations. The customer is obligated to interrupt load when called upon to do so only during those hours on the following day for which its bid was accepted. The penalty for noncompliance is likely to be much more modest also.

Table 2. Characteristics of load participation in wholesale power markets

	Day-ahead energy	Spinning reserve	Supplemental reserve	Replacement reserve
Aggregation	RTO might require minimum size, say 1 MW, which would require aggregation for all but the larger industrial loads			
Meters	Hourly	Interval meters capable of recording consumption at the 5- or 10-minute level		
Communication	Daily submission of hourly bids to RTO, daily receipt of hourly prices	Daily submission of hourly capacity and energy bids to RTO, RTO must be able to call on winning bidders to reduce loads within required times		
Advance notice	Day ahead	10 or 15 minutes for full compliance	30 or 60 minutes for full compliance	
Frequency	Customers are free to participate in these markets as they choose; once having chosen on a day-ahead basis to sell reserves during certain hours, they are then committed to providing that service if called upon			
Duration	Not applicable	The load reductions might need to be sustained for as long as an hour (spinning and supplemental reserves) or two hours (replacement reserves)		
Penalties	None	Penalties applied because load committed to make reductions upon RTO call for reliability service (quid pro quo for reservation payment)		
Payments	Day-ahead market clearing price for energy	Day-ahead market clearing prices for capacity plus energy payments for actual load reductions when called upon		
Baseline	None	Because advance notice is so short, baseline is usually consumption during one or a few intervals before the ISO call		

^a Where generators are permitted to bid in startup and no-load costs, retail customers should be permitted to bid in curtailment-initiation costs to reflect any costs they might incur in getting ready to modify their hourly electricity use.

IV. DEMAND-SIDE RELIABILITY RESOURCES -- RECENT ISO EXPERIENCE

Since the summer of 2000, the three Northeastern ISOs have operated targeted demand-side programs to improve system reliability. Based upon initial experience in 2000, these programs were refined and expanded (in particular, to include economic demand-side programs) for summer 2001.

These special reliability markets procure contingency-reserve services from loads. The ISOs likely established these separate markets, rather than encouraging retail loads to participate in ancillary-service markets, for a variety of reasons. PJM has no markets for reserves, while New

England recognizes serious limitations in its ancillary service markets¹⁵. More generally, market participants likely felt that the metering and telecommunications requirements for participation in what had historically been generation-only functions, were too onerous. Whether these demand-management pilot programs disappear after the ISOs create or modify their markets to encourage retail-load participation or whether they become permanent is unclear.

Because these programs are so new, their effects have generally been modest. For example, in the initial year only 7 MW of demand reduction signed up for the New England program, and only 63 MW signed up in 2001. The PJM reliability program achieved a maximum reduction of 62 MW one day in August 2001, with an average reduction of 21 MW (PJM 2001). The New York program had a greater impact, achieving an average reduction of 355 MW over 23 event hours, and a peak of 425 MW when most needed. (Neenan Associates 2002). Experiences of the three regions with reliability demand-response programs¹⁶ are summarized below.

(A) Program Designs

ISO-NE offered the Demand Response Program (Class 1 Demand Response Program), which was designed to enroll large commercial and industrial customers. All participants in the program were required to have web based technology for notification and load monitoring supplied by RETX. Compensation to participating customers was based on day-ahead price forecasts. Participants received 100% of the difference between the bid price and their retail rate.

The NYISO operated the Emergency Demand Reduction Program (EDRP) in 2001. Most of the participants were large (500-1,000 kW) and very large (greater than 1 MW) industrial customers. However, the aggregators participating in the program in many cases bundle together a number of smaller customers. The EDRP is a “Call”-type program but is voluntary in that there are no penalties and payment is based on the performance of a participant in each given hour of a curtailment event. The EDRP only operates when there are Power Emergency Advisories issued by the NYISO.

Advisory alerts are sent by burst e-mail and by phone to the CSP or LSE. The LSE or CSP is asked to provide an estimate of the load reduction likely from their customers. Notifications are sent two hours ahead of the beginning of the curtailment. Performance is measured using a customer baseline calculation derived from a 10-day rolling average of each hour of the curtailment period. Interval load data is usually not available to participants except as part of their normal billing process.

¹⁵ The markets for spinning and supplemental reserves in New England are, as the ISO itself put it, “fundamentally flawed” (ISO New England 1999). In addition, even though ISO New England acquires more than 1,000 MW for the equivalent of replacement reserves on most days, it has no formal market to acquire those resources.

¹⁶ Data in these summaries are taken from work of Chuck Goldman (ongoing study of PRL discussed in Framing Paper #1) and from Neenan, 2002.

For the EDRP, payments by the ISO are the lesser of \$500/MWH or the local marginal spot price. Participants in the LSE programs will be paid 90% of \$500/MWH less the retail rate they pay. Participants in the CSP programs will be paid a somewhat smaller amount depending on the terms of their contracts.

PJM - PJM operated the Emergency Load Response Pilot in 2001, focusing on demand response from very large industrial customers – more than 90% are greater than 1 MW in size and more than 80% are industrial customers, including pumping stations, air reduction facilities, and cement plants. Performance is calculated using a Customer Baseline Load derived from a 10-day rolling weekday average of each hour of operation. Participants must have interval meters but interval load data is not provided to participants except as part of their normal billing cycle.

The program is voluntary, and is based on economic incentives to customers. It passes through to participants 100% of the avoided purchase cost savings (less retail rates) based on actual market prices during the hours of operation. The program uses bulletin board-style web pages. Notification of events is by pager and e-mail.

It is important to note that PJM, in conjunction with its members, also operates legacy Active Load Management (ALM) programs. The ALM programs are a resource embedded in the operations of PJM's LSE members. The legacy ALM programs involve sizeable investment and maintenance on the part of the LSEs. PEPCO, for example, has an ALM program consisting a total ALM credit of 272 MW. PECO has a relatively small ALM program, about 100 MW, consisting of an interruptible rider on their large steel customers. PP&L has a very large interruptible/curtailable program

(B) Summer 2001 Results:

New England - The NE-ISO Demand Response Program was larger than the program offered in 2000, but was still relatively small. It enrolled 101 customers with 63 MW of potential demand reduction. The program operated for three days during the week of August 7, 2001 when the system operator declared an electric supply emergency. During this period the energy clearing price reached \$1,000/MWH for several hours. The program achieved load reductions of 25 MW on average over this period, of which about 14 MW was due to operation of back-up generators.

New York - The EDRP included both the interruptible loads and standby generation of 300 participants served by 12 load-serving entities and 9 curtailment service providers in New York State. The total load reduction potential was 700 MW, of which 150 MW is stand-by generators. In Summer 2001 the EDRP was operated four times (Aug. 7,8,9, and 10) due to an operating reserve shortage. The program delivered 450 MW on average, which is a significant share of the 1800 MW operating reserve that NYISO tries to maintain, and more than 2/3 of 700 MW demand reduction potential. Observers credit the economic benefits of the program for customers with the solid enrollment and positive customer response.

PJM - As of September 2001 the Emergency pilot included 40 participants totaling 155 MW. The PJM pilots (economic and emergency) were operated on three occasions in 2001, July 25, August 8, and August 9, when Maximum Emergency Generation events were declared. On July 25 the two pilot programs contributed 22 MW over and above the 150 MW of ALM achieved by the legacy programs. On these days the bid-in price and the LMP (Locational Marginal Price) were well over the \$500/MWh minimum guaranteed to Emergency Pilot participants. PJM believes that this pilot accomplished its “Proof of Concept” objectives. The key motivator to participation was access to economic incentives; the key barrier was need for additional meters and the environmental restrictions many participants faced regarding use of their backup generators.

(C) Program Experience Summary

Table 3 provides a brief comparison of some of the elements of these programs.

Table 3. Comparison of ISO summer 2001 reliability load-reduction programs

	PJM	New York	New England
Payment basis	Actual load reduction		Reservation payment plus payment for load reduction
Payment amounts	Higher of zonal price or \$500/MWh		Energy clearing price
Availability of load reductions	9 am to 10 pm		8 am to 10 pm
Minimum capacity per customer	0.1 MW, interval meters required		
Penalties	None		Refund of reservation payment to later of first of month or last successful performance
Baseline	Load during hour before PJM calls for load reduction	Average of hourly loads during highest 5 of the last 10 weekdays	Average of hourly loads during past 10 business days
Dispatch	Maximum emergency generation	Operating Reserve Shortage or Major Emergency	Operating Procedure 4
Advance notice	one hour	two hours	30 minutes, up to 2-hour duration

Source: Biewald and Johnston (2001).

V. POLICY AND PROGRAM ISSUES

This paper describes current North American reliability rules and practices. Because of the traditional belief that retail customers could not and would not want to manage their loads in response to economic incentives, these rules and practices have historically excluded loads from the provision of, and payment for, reliability services. This need not and should not be the case. Expanding reliability functions and markets to include retail loads can improve power system reliability or lower the costs to maintain reliability at current levels.

Tapping demand-side resources to improve reliability cost-effectively will require removal of existing barriers in a number of areas, as well as positive decisions to develop customer and market capabilities to deliver load-side resources when needed by system operators. To begin with, NERC and the ten regional reliability councils will need to review their reliability rules to remove the technology biases that favor traditional supply-side resources over potential demand-side solutions. Reliability standards should specify what is to be accomplished through a standard, but not how that reliability goal is to be achieved. And these performance standards should be written so that customer loads are permitted to provide those reliability services they can technically and economically provide.

There are also important actions that need to be taken by FERC,¹⁷ at the regional/ISO/RTO level; and by state utility and environmental regulators. We will not focus here on the FERC agenda, although that may become relevant to NEDRI Participants as our process evolves. Regional and state options and policy issues are briefly addressed in the section below.

Policy Options and Topics for Discussion

1) How should ancillary service markets be organized in New England to realize the potential value of demand-side resources?

Most of the focus of reliability programs has been on the emergency curtailment programs operated by the ISO. However, in the long-run, routine ancillary service markets will provide many reliability services. What needs to be done in New England to design and operate these markets so that demand resources can participate on a technology-neutral basis?

2) Should customers be permitted to contract with the ISO for reliability load response programs (directly or through a Curtailment Service Provider), or must those arrangements be made through LSEs?

¹⁷ A recommended action agenda for FERC is set out in R. Cowart, "Demand-Side Resources and Regional Power Markets: A Roadmap for FERC," (October 2001). Many of these topics were addressed in FERC's March 2002 Standard Market Design Working Paper, discussed in Part I of this paper.

One of the major challenges to both price- and reliability-focused programs at the regional level is the tension between openness and effectiveness. On the one hand, LSEs, particularly traditional utilities, have a great deal of customer knowledge and have longstanding relationships with them. ISOs do not have those relationships. On the other hand, if LSEs are the essential gate-keepers to customer participation, they may block participation or dampen customer interest by requiring a substantial share of the economic value of the customer's load response. New England, which initially required approval of incumbent LSEs, has seen slow penetration in its reliability programs. In New York, a uniform tariff requires utilities to pass 90% of the DR value to the end-user, if they are the customer's representative; however, customers may sell DR into the market without going through the local utility. In 2001, about 72% of enrolled customers in New York worked through their LSE, and about 25% enrolled through a Curtailment Service Provider.

3) How can regional DR reliability programs be coordinated with legacy interruptible contracts operated by LSEs?

Concerns about "double counting" and concerns over potential conflicts among program claims on customer resources have sometimes led decision-makers to forbid loads from participating in regional reliability programs if they are also participating in a local utility interruptible program. However, demand response may well have both a local and a regional benefit at the time it occurs. How can the appropriate values be communicated and paid to participating customers without undermining reliability and cost-effectiveness goals?

4) How should demand resources be integrated into "installed capability" markets?

As noted in the text, there is a robust debate over the value of ICAP markets, at least when energy markets are more fully developed. However, if ICAP continues to be a means of ensuring adequate long-term investment in reserves, how should demand resources be treated? New York permits LSEs to meet their ICAP obligations through curtailable load-side "special case load resources" (ICAP/SCR). There is substantial evidence that a reliable stream of ICAP-type payments may be critical to continued availability of demand-side capacity in regional markets, and that programs that merely "pay when called" will not elicit sufficient enrollment. This should be discussed.

Moreover, the value of enrolled resources varies with location, so ICAP-type payments to DR should also reflect those values. Payments in 2001 in the New York City region, for example were four times as high as for Western New York, providing strong incentives for LSEs and CSPs in that region to sign up customers. How is locational value reflected in the ISO-NE programs recently announced for Southwest Connecticut, or for other high-value locations?

5) Can reliability-oriented programs be designed to recognize the unique characteristics and needs of loads, just as they do now for generators?

System operators have traditionally understood and accommodated the needs of different types of generation in dispatch queues and, more recently, in competitive markets (e.g., energy-limited

hydro, ramp rate constraints, minimum run times). Has the ISO conducted sufficient research on customer needs to understand how reliability rules could be adjusted to accommodate the particular needs of loads, and thus enhance the reliability contribution of demand response resources? What rule changes might be needed?

6) What types of demand-side resources should be eligible to participate in “emergency” load response programs?

(A) What is the role of and/or limits on use of diesel-fired back-up generators?

As with PRL programs, those designing and administering emergency demand-response programs must confront the environmental consequences of the program designs. In the 2001 New York reliability program (EDRP), 85% of the curtailment performance came from load reduction, while 15% was accompanied by some degree of on-site generation. While the NYISO limited participation in their *economic* program (DADRP) to non-diesel fired BUGs, many PRL programs in other regions allow diesel BUGs to participate. This may be a relatively straightforward issue for economic programs in New England, but the trade-offs are somewhat different for reliability programs. Some analysts argue that limited-run *emergency* programs should permit diesel back-up generation, at least to some degree. NEDRI participants should consider the types of demand-side resources that should be eligible to participate in these programs.

(B) Will ISO rules permit participation and compensation to aggregations of small, widespread load responses resources?

Utilities have long understood the value of widespread load management programs, such as water heaters or air conditioners managed through ripple- or radio-controls. However, reliability managers and system operators have been reluctant to accept such resources for reliability purposes, and have generally sought individual metering and/or operations center control of the resource. Reliability managers have legitimate concerns, but widespread resources can be quite valuable, and statistical verification should be possible. Will rules in New England promote such programs, which may be the best way to secure reliability resources from residential and small commercial customers?

7) What actions should state regulators take to enhance customer participation in reliability-focused DR programs?

- Should state regulators mandate interval meters, at least for larger customers?
- Should PUCs redefine Standard Offer service, at least for larger customers, as a service with interval meters and real-time pricing?
- Should interval meters be required and/or subsidized?

One of the persistent barriers to participation in load-response programs, including reliability programs, is the insulation of customers from exposure to real time prices. Some advocates of

market-based solutions urge regulators to remove the price caps and price stability of Standard Offer service, and to move those customers to real-time (or at least time-block-based) pricing. Recognizing that most customers do not want to become electricity “day traders,” others promote more modest steps, such as:

- compensation for demand interruptions at their market value, but permitting consumption at Standard Offer rates;
- limiting mandatory real-time prices to large customers only;
- accelerating deployment of interval meters, which would facilitate enrolment in reliability and price-response programs, without a *requirement* to pay real-time prices;
- and socializing the costs of meters and associated telemetry as a regional reliability enhancement that should be supported through a widespread uplift charge. New England has engaged and supported RETX as one way to advance customer interaction and load response. NEDRI Participants should evaluate these options to enhance customer reliability responses.

Note: many of these issues will be discussed at NEDRI in connection with Framing Paper #3 on metering and retail pricing.

8) Should the ISO and state PUCs support development of Curtailment Service Providers to accelerate and deepen DR for reliability?

Traditional utility relationships and lack of knowledge of the options by customers also stand as barriers to the rapid deployment of DR. Competition for the customer-response resource might accelerate innovative service packages that could combine efficiency, on-site generation, and load response options providing sufficient customer benefit to deepen enrollment in reliability-focused programs. Should regional and state decision makers seek to open up this avenue? If so, what barriers are most troublesome to potential CSPs and what incentives would be most important to provide?

9) What target levels, if any, should New England adopt as a benchmark level of success for reliability-oriented DR programs?

It is very difficult to know in the abstract whether New England is being “successful” or “failing badly” with the DR reliability programs it offers. Some would argue that the realistic potential is small, so the modest enrollments seen in past years simply reflect reality. Others would argue that low enrollments reflect numerous market barriers to success, both in the underlying electricity markets, and in the rules of the programs themselves.

Meanwhile, New England’s historic utility load management programs have been eroded by divestiture and competition. The region incurs large financial and environmental costs to maintain a large reservoir of spinning reserve, and is facing congestion problems in Massachusetts, Connecticut, and elsewhere. What benchmarks of success, if any, should New England adopt as a reality check on the progress of its reliability-focused DR programs? Can we evaluate the relative contributions of “economic” vs. “reliability” enrollments and responses?

10. What lessons should ISO-NE and Stakeholders in New England take from the experiences of New York and PJM with demand-side response for system reliability?

In 2001, New England obtained just 25 MW in its Demand Response Program, PJM called on just 22 MW of demand response in its summer pilot program, while the New York ISO cost-effectively called upon about 450 MW in its EDRP (reliability) program. Was this due to a better-functioning underlying market in New England, a better-functioning reliability program in New York, or better-functioning markets and legacy interruptible programs in PJM? PJM has relatively small pilot emergency programs, but maintains large Active Load Management (ALM) programs operated by the legacy utilities. New York appears to have had greater success in enrolling and calling upon DRR as part of its ISO-level programs. PJM appears to have had greater success with demand response through utility-run ALM programs. To what degree are these experiences relevant to New England, and what lessons can we take from them?

REFERENCES

- B. Biewald and L. Johnston 2001, *The Other Side of Competitive Markets: Developing Effective Load Response in New England's Electricity Market*, Synapse Energy Economics, Cambridge, MA, June 13.
- S. Braithwait and A. Faruqui 2001, "The Choice Not to Buy: Energy Savings and Policy Alternatives for Demand Response," *Public Utilities Fortnightly* **139**(6), 48! 60, March 15.
- California Public Utilities Commission 2001, *Energy Division's Report on Interruptible Programs and Rotating Outages*, Energy Division, San Francisco, CA, February 8.
- J. D. Chandley 2001, "A Standard Market Design for Regional Transmission Organizations," *The Electricity Journal* **14**(10), 27–53, December.
- R. Cowart 2001, *Efficient Reliability: The Critical Role of Demand-Side Resources in Power Systems and Markets*, National Association of Regulatory Utility Commissioners, Washington, DC, June.
- R. Cowart 2001, *Demand-Side Resources and Regional Power Markets: A Roadmap for FERC, RTO Futures*, Montpelier, VT, October.
- E. Hirst 2001, *Real-Time Balancing Operations and Markets: Key to Competitive Wholesale Electricity Markets*, Edison Electric Institute, Washington, DC, and Project for Sustainable FERC Energy Policy, Alexandria, VA, April.
- E. Hirst and S. Hadley 1999, *Maintaining Generation Adequacy in a Restructuring U.S. Electricity Industry*, ORNL/CON-472, Oak Ridge National Laboratory, Oak Ridge, TN, October.
- E. Hirst and B. Kirby 1998, *Unbundling Generation and Transmission Services for Competitive Electricity Markets: Ancillary Services*, NRRI 98-05, National Regulatory Research Institute, Columbus, OH, January.
- E. Hirst and B. Kirby 2001, *Retail-Load Participation in Competitive Wholesale Electricity Markets*, Edison Electric Institute, Washington, DC, and Project for Sustainable FERC Energy Policy, Alexandria, VA, January.
- B. F. Hobbs, J. Inon, and S. E. Stoft 2001, "Installed Capacity Requirements and Price Caps: Oil on the Water, or Fuel on the Fire?" *The Electricity Journal* **14**(6), 23! 34, July.
- ISO New England, Inc. 1999, *Request for Expedited Approval of Revisions to NEPOOL Market Rules 1 and 10*, Docket No. ER99-4002-000, submitted to the Federal Energy Regulatory Commission, Holyoke, MA, August 5.

ISO New England, Inc. 2001, *Compliance Filing*, Docket No. EL00-62-015, submitted to the Federal Energy Regulatory Commission, Holyoke, MA, June 4.

J. Kueck 2002, personal communication, Oak Ridge National Laboratory, Oak Ridge, TN, January 3.

R. Lafferty, D. Hunger, J. Ballard, G. Mahrenholz, D. Mead, and D. Bandera 2002, *Demand Responsiveness in Electricity Markets*, Office of Markets, Tariffs, and Rates, U.S. Federal Energy Regulatory Commission, Washington, DC, January 15.

Neenan Associates 2002, *New York ISO Price-Responsive Load Program Evaluation: Executive Summary*, prepared for the New York Independent System Operator, Altamont, NY, January 15.

North American Electric Reliability Council 1993, *Demand-Side Management, The System Operator's Perspective*, Princeton, NJ, December.

North American Electric Reliability Council 1997, *NERC Planning Standards*, Princeton, NJ, September.

North American Electric Reliability Council 2001a, *Reference Document: Interconnected Operations Services*, Interconnected Operations Services Subcommittee, Princeton, NJ, March 28.

North American Electric Reliability Council 2001b, "Policy 1 — Generation Control and Performance," Version 2 of proposed policy change, Princeton, NJ, June 17.

North American Electric Reliability Council 2001c, "Policy 1 — Generation Control and Performance," *NERC Operating Manual*, Princeton, NJ, November 15.

PJM Interconnection LLC 2001, *Report on the 2001–2002 PJM Customer Load Reduction Pilot Program*, Norristown, PA, submitted to the Federal Energy Regulatory Commission, Docket No. ER01-1671-000, December 28.

U.S. Federal Energy Regulatory Commission 2002, *Working Paper on Standardized Transmission Service and Wholesale Electric Market Design*, March 15.

U.S. Federal Energy Regulatory Commission 1996, *Promoting Wholesale Competition Through Open Access Non-Discriminatory Transmission Services by Public Utilities; Recovery of Stranded Costs by Public Utilities and Transmitting Utilities, Final Rule*, Docket Nos. RM95-8-000 and RM94-7-001, Order No. 888, Washington, DC, April 24.

U.S. Federal Energy Regulatory Commission 1999, *Regional Transmission Organizations*, Order No. 2000, Docket No. RM99-2-000, Washington, DC, December 20.