NEW ENGLAND DEMAND RESPONSE INITIATIVE

LONG-TERM RESOURCE ADEQUACY: DEMAND RESPONSE OPTIONS FOR NEW ENGLAND*

Eric Hirst
Consulting in Electric-Industry Restructuring
Oak Ridge, Tennessee

November 13, 2002

How can ISO New England best encourage demand resources to qualify as long-term resources (LTRs), and what criteria should such resources meet? A key principle to consider in assessing alternative approaches is to ensure comparability# in treatment of demand and supply resources.

In deciding how to encourage demand resources to participate in long-term resource-adequacy programs, it might help to view the LTR requirements as an umbrella under which existing (and future) demand-response programs fit. That is, a regionwide LTR requirement need not engender the creation of new demand-response programs.

Several of the characteristics and their choices include:

- <u>Forecast period:</u> How far ahead of the operating day should the regional transmission organization (RTO) (1) set the resource requirements for each load-serving entity (LSE) and (2) qualify resources as meeting all the requirements for eligibility? FERC, in its July 2002 notice on Standard Market Design (SMD), suggests a forecast period that is long enough to allow time for new facilities, including generation and associated transmission or demand management, to be built. This suggests a forecast horizon of two to four years. The existing ISO ICAP programs, on the other hand, use a 1-year horizon. A long forecast period will, as FERC notes, allow enough time for new "construction" to occur. On the other hand, the longer the time between the forecast and real time, the less accurate the load forecast will be and the less reliable (certain) the resource developers will be about the availability of their resources in real time. A

*This paper was prepared under contract with Lawrence Berkeley National Laboratory, Berkeley, CA (Charles Goldman, project manager) for the New England Demand Response Initiative.

#I emphasize comparability rather than equality because demand and supply resources, although they may provide the same services to the bulk-power system, are different.

multiyear forecast period suggests that qualification of resources might change over time. For example, a credible plan might be sufficient three years ahead of operations. One month ahead of operations, however, the RTO might require that the resource be physical and available.

Recommendation: The RTO require a plan three years ahead of operations from qualifying demand resources. The plan would specify the participating customers, the systems at each customer's facilities that would be interrupted, the method to be used to notify the customer of the impending interruption, the method used to measure and pay for the load reduction (e.g., baseline methods and metering requirements), and other details required to implement the specified load reductions.

Questions: What should the RTO require from resource developers during the interim period (e.g., between the plan several years in advance and real time)? And what would constitute a "credible plan" for demand resources?

- <u>Certification Requirements:</u> What proof must a resource provide to the RTO that it can, in real time, deliver the capacity and energy it claims it will provide? The answer to this question is closely related to the prior issue on forecast period. Far ahead of real time, certification might be relaxed, but close to real time, certification might involve a physical demonstration of the capacity, energy, and ramping capability of the resource as well as its metering and communications infrastructure. Once a resource is physical (i.e., operational), how often must it document its capability?

  Recommendation: Nonseasonal resources either deliver the capacity they committed during an RTO contingency or, if no contingencies occurred during the year, they conduct a test to demonstrate that capability. The test would be conducted by the RTO, it would be unannounced, and it would require the demand resource to perform as it is expected to during an emergency.

- <u>Qualifying Resources:</u> What kinds of demand resources might be eligible for qualification as long-term resources? Demand resources can be categorized as (1) interruptible load, including direct load control; (2) dynamic pricing, including participation of retail loads in the RTO's day-ahead and real-time markets for energy; (3) participation of retail loads in the RTO's day-ahead and real-time markets for contingency reserves; and (4) energy efficiency. Each of these categories might be treated differently.

  - Interruptible loads are provided by retail customers that agree to reduce their load by a minimum amount in response to dispatch instructions from the RTO. This is the type of demand response that participate in the existing ISO emergency demand-response programs.

    Recommendation: These programs would qualify as LTRs.

  - Dynamic pricing involves retail customers with the required metering and communications system (interval meters and the ability to receive information

on day-ahead and real-time hourly prices, and perhaps intrahour prices) and the ability and willingness to modify electricity use in response to these prices.
Recommendation: Because these responses to time-varying prices cannot be defined precisely beforehand, they may not qualify as resources an RTO can call on. On the other hand, these price responses should show up in an LSE's load forecast because prices and loads are highly correlated.[*]

- Loads can provide 10-minute and 30-minute reserves if they have the required metering and communications systems and can respond fully within the required time. That is, loads must meet essentially[#] the same requirements as generators do to be eligible to provide contingency reserves.
  Recommendation: Such load reductions should qualify as LTRs.[§]

- Energy efficiency involves the use of equipment that provides the same level of end-use amenity with less electricity use (e.g., compact fluorescent lamps to replace incandescent lamps). Because energy-efficiency measures are passive rather than dynamic, system operators cannot call on them as resources in real time, and they don't meet the security requirements of NERC's reliability definition. That is, they operate much as large nuclear, baseload power units. By lowering demand, these measures and technologies reduce the need for new generating units and transmission facilities.
  Recommendation: These efficiency improvements should be incorporated in the long-term load forecasts that determine how much resources are required, but should not be considered qualifying resources.[†]

■ Seasonal v Annual Requirements: The current ISO installed-capability programs (ICAP) divide each year into two or more seasons. The FERC SMD notice is silent on this point. This issue is important for demand resources that are weather dependent (e.g., air

---

[*]As noted in Exhibit 3 of the companion paper, "Long-Term Resource Adequacy: The Role of Demand Resources," A 1-MW reduction in the load forecast is more valuable to the LSE than a 1-MW qualifying resource.

[#]I write "essentially" because there is no reason to require loads to telemeter their consumption levels to the RTO every few seconds. It is surely sufficient for loads to record their consumption once every minute. They may not even have to report these data to the RTO in real time. It may be enough for the RTO to obtain this 1-minute data at the end of each billing cycle.

[§]Some retail customers might prefer to provide contingency reserves rather than experience emergency-driven interruptions. In the former case, the load reductions will be called on roughly once a month (when a major generator or transmission line fails), but the deployment period will be only an hour or two. Loads with some storage capability may be well suited to provide contingency reserves. On the other hand, some customers may be more willing to suffer a rare, but longer-duration load reduction, which is typical of the emergency programs.

[†]By purchasing or committing to acquire efficiency resources, LSEs avoid a portion of the LTR they would otherwise be required to deliver. This provides a source of value for efficiency resource in LTR markets that can be used to help pay for these resources.

conditioning and space heating) or seasonal (e.g., lighting).* Some demand resources, such as the cycling of air conditioners and swimming-pool pumps, can provide large load reductions during the summer, but none during the rest of the year. Some load reductions are coincident with system peaks, while others are not.
Recommendation: The LTR requirements should be seasonal, rather than annual. If an RTO chooses annual requirements, it should permit resources that can only respond during certain months or seasons to receive partial credit, based on the amount of time and load response it can provide.#

- <u>RTO Role in Markets:</u> How much responsibility should the RTO take in establishing long-term reserve requirements and in assisting market participants in meeting their obligations? Should it operate markets to permit suppliers and LSEs to buy and sell the rights to these LTRs? Or should it refrain from running markets and, instead, require LSEs and suppliers to make their own bilateral arrangements?
  Recommendation: Clearly, the RTO should develop the long-term load forecast, establish the annual or seasonal reserve requirement, and determine what fraction of this overall requirement each LSE must provide. The RTO should also run LTR markets because market participants are free to eschew the RTO markets and make their own bilateral deals. That is, participation in the RTO capability markets is entirely voluntary. Demand resources should be permitted to participate in RTO markets for LTRs.

- <u>End-Use Infrastructure Requirements:</u> What metering and communications systems must a qualifying retail load have?
  Recommendations: The answer depends on the kind of demand resource, as discussed above. For most demand resources, interval meters that record and store electricity consumption at the hourly level are sufficient. However, if the demand resources are participating in markets for contingency reserves, which require full response within 10 or 30 minutes, the RTO might reasonably require metering that records data every minute. (The comparable requirement for generating units is usually to telemeter data to the RTO once every several *seconds*.) With respect to communications, the retail customer that is providing the demand resource must be able to receive, confirm, and act upon instructions from the RTO, which might be provided with as little advance notice as 10 or 30 minutes (for reserves) or an hour or two (for other programs). Such two-way communications could involve the use of telephones, pagers, fax machines, or emails. Whether customer data on electricity use (in particular, the load reduction achieved in response to the RTO directive) must be provided to the RTO in real time is an open question. Although some believe that the RTO must know, in real time, whether

---

*The capability of many generating resources also varies across seasons, but to a much smaller extent.

#This load-reduction credit would be annualized on the basis of the size and timing (relative to the system peak) of monthly load reductions and monthly estimates of loss-of-load probability. To the extent that reliability problems are more likely to occur and be more severe during the summer months, the LOLP will be higher, giving more credit to load reductions during these months than other months.

and to what extent each resource is providing the reliability resources it was asked to deliver, it may be sufficient for the RTO to receive information on the actual demand reductions at the end of each billing cycle.[*]

Question: Who should pay for these metering and communications systems, the participating customer, the LSE, or electricity consumers in general (through an RTO uplift charge)?

- RTO Resource Rights: What rights does the RTO have to resources that qualify as LTRs? These rights (and limits) include (1) the maximum number of times a year (or season) the RTO can call on the demand resource, (2) the maximum amount of time the load can be curtailed during each event, and (3) the minimum amount of advance notice the RTO must provide. For generating units, the RTO can call on qualified resources as often as needed and for as long as needed each time, which suggests that generators provide a more valuable resource, per MW, than do demand resources; these differences between supply and demand resources complicate the reservation payments for the two sets of resources. Under what circumstances can the RTO call on these resources? Generally, the RTO has a prescribed set of procedures (OP-4 in New England) that specify what actions are taken in what order as the amount of contingency reserves falls (or is expected to fall) below required levels. This procedure will specify at what point in the sequence these LTRs are armed and subsequently dispatched.[#]

  Recommendation: For demand resources, which may participate in other shorter-term RTO programs (such as a day-ahead demand-response program or a real-time emergency program), it is important to specify in what order the demand resources will be called. Because the LTR resources are long-term in nature and are likely to receive a capacity (reservation) payment (in $/kW-month), they should probably be called before demand resources that do not receive an upfront reservation payment. An alternative is to permit each LTR to set an energy strike, above which, it would be called on. The RTO would then stack all its LTRs in order of increasing energy strike price.[§]

---

[*] Supply and demand resources can reasonably be held to different standards with respect to the immediacy with which production (or consumption) data are provided to the RTO. A difference is reasonable given the size difference between the two classes of resources. Generators are typically large, 100 to 1000 MW, whereas load reductions are typically small, 1 MW and up. While the RTO might reasonably need to know the output of each large generator in real time, it has much less need to know how individual demand resources are responding in real time, although it will want to know how the aggregate of such resources is performing.

[#] In New York, for example, the ISO notifies ICAP-qualified demand resources day ahead that they might be called on, and then provides an in-day 2-hour notice for dispatch. The ISO can call on supply resources with fewer restrictions, which makes them more valuable than demand resources.

[§] This proposed use of an energy strike price to trigger dispatch of LTRs does not imply that loads should be paid explicitly for their reductions. The load's strike price is merely the price at which it is willing to lower consumption and enjoy the benefits of a lower electric bill (in addition to the capacity payment).

- Obligations of Retail Customers and LSEs: The flip side of the RTO rights are the obligations of the entity that committed these resources to the RTO (the LSE, the retail customer itself, or perhaps a third-party provider).
  Recommendation: The providing entity must keep the RTO informed of the current status of the resource, in particular whether it is unavailable (the equivalent of a forced or maintenance outage for a generator) at all times. In addition, the contracting party must maintain the metering and communications systems in working order and otherwise ensure that the demand resource is available to respond to RTO directives as called for in the contract. Finally, to the extent that the contracting entity receives upfront capacity payments, it might be required to demonstrate its creditworthiness.

- Minimum Resource Size: The standard practice is to limit resources to no less than 1 MW (although demand resources as small as 0.1 MW can participate in New York's ICAP program as Special Case Resources). Three key criteria are (1) whether the resource must be visible to the RTO (i.e., must the resource be large enough that the system operators can see the changes in output or consumption on the RTO's SCADA system), (2) whether there is an upper limit on the number of resources with which the system operators can communicate before and during an emergency (in part, a function of how automated these communication systems are), and (3) the ability of the RTO to bill and pay for reliability services.
  Recommendation: Smaller loads can be aggregated by the LSE (or some other party) and presented to the RTO as a block of load reduction of 1 MW or more.[*] Such small loads need not be separately metered; the expected load reductions can be calculated statistically on the basis of periodic tests performed on a valid statistical sample of the small participating customers.

- Customer Baseline Level: This issue occurs with the short-term demand-response programs.
  Recommendation: Whatever works for those programs should be sufficient for LTRs. Calculating the baseline for demand resources that participate in the markets for reserve services is much simpler (and less amenable to gaming) than for interruptible-load programs. The advance notice for provision of reserves is only 10 or 30 minutes and the period of deployment is only one or two hours, which suggests that the average of customer-load levels for a few intervals before dispatch should be sufficient to establish the baseline. The amount of advance notice and the deployment period are longer for interruptible-load programs, which makes it more difficult to establish an accurate baseline.

[*] In such cases, the providing party is responsible for communications with the retail loads, for collection and aggregation of data from the loads, and for payments to the individual loads for their participation in RTO programs. That is, the RTO deals with the providing party in such cases and not with the individual loads.

- <u>Payments:</u> Should LTRs receive capacity payments (e.g., in $/kW-month), and should they also receive energy payment (e.g., in $/MWh)?
Recommendation: If the RTO does not operate markets for LTRs, the capacity payments are of no consequence to the RTO and are a private matter between the parties to bilateral arrangements. Supply resources that convert capacity into energy in real time in response to RTO dispatch signals would get paid the real-time spot price. Demand resources that convert capacity into energy reductions in real time in response to the same RTO dispatch signals would see a lower electricity bill, with the reduction equal to the load reduction times the real-time spot price. In other words, the RTO should not pay twice for demand response.

- <u>Penalties:</u> If demand resources are treated as explicit long-term resources, they should be subject to the same penalties as participating generators face. If, however, the demand resources appear only as a reduction in the load forecast, they would not be subject to penalties because they would have received no upfront capacity payment. In other words, the quid pro quo for receipt of capacity payments must be acceptance of nonperformance penalties. How these penalties are determined (e.g., as a multiple of the annual carrying cost of a new combustion turbine or a multiple of the real-time energy price) and how high they should be is a separate matter, but it is not a demand-resource specific issue.[*]

Table 1 summarizes some of the points made above, organized by type of demand resource.

C:\Data\Docs\NEDRI\LTRAOptions.wpd

[*] These substantial penalties involve *non*compliance, rather than slight *under*compliance. Presumably, a resource that provided, say 95% of its contracted amount, would have to reimburse the RTO for 5% of its capacity payment but would not face any penalties.

Table 1. Types of demand response and their characteristics as long-term resources

	Interruptible load	Ancillary services	Dynamic pricing	Energy efficiency
Treatment as LTR	Yes	Yes	No, part of LSE load forecast	
Performance requirements	Commit to reduce demand by contracted amount under specified conditions when called by RTO	Bid into day-ahead reserve markets	None	
Metering requirements	Hourly metering	1-minute metering	Hourly metering	None
Communication requirements	Ability to receive and confirm operator requests		Ability to receive day-ahead and real-time hourly energy prices	None
Advance notice	30 to 120 minutes	10 or 30 minutes	Customer discretion	None possible
Duration of response	Up to several hours	1 hour	Not applicable	
Payments				
Capacity	Yes, long-term	Yes, day ahead or real time	No	
Energy	No	No	No	
Customer baseline level	Same as for existing and planned demand-response programs	Because advance notice and dispatch times are short, could be average of a few intervals before dispatch		
Frequency of dispatch	Specified by program	As needed, based on energy-strike price of resource	Not applicable	
Penalties	Same as for generation providing LTRs	Same as for generation providing ancillary services	None	