Water Heaters to the Rescue

Demand Bidding in Electric Reserve Markets

With just a few changes in reliability rules, regulators could call on consumer loads to boost power reserves for outages and contingencies.

In proposing a standard market design (SMD), the Federal Energy Regulatory Commission (FERC) makes clear that it wants customers to participate in wholesale power markets, such as by bidding an offer to curtail consumption, increase supply, and reduce upward pressure on prices.

“We believe in the direct approach of letting demand bid in the market,” says FERC.

In fact, FERC much prefers this demand-response strategy to the more traditional special programs for load reduction, whereby regulators typically promise an incentive or subsidy to customers in return for cutbacks in usage.

According to the commission, letting customers bid their own demand directly in the market as a system resource “will be less costly than a program where an end-user receives pay-to customers in return for cutbacks in usage.

Whereby regulators typically promise an incentive or subsidy to customers in return for cutbacks in usage.

To ensure power system reliability, grid system operators require, and that the ISO can meet the disturbance control requirements on resources that provide contingency reserves.

To ensure power system reliability, grid system operators require, and that the ISO can meet the disturbance control requirements on resources that provide contingency reserves.

Consider the possibility of allowing resources with shorter ramping capability (in megawatts per minute, or MW/min). In addition, the resource must be able to sustain the interruption conveniently for only short periods—about an hour. Should regulators bar such resources from providing contingency reserves because of requirements developed with generators—and only generators—in mind?

Consider the possibility of allowing resources with shorter sustainable deployment time to provide reserves. That would accommodate loads with limited storage. And with more sophisticated deployment of resources—dispatching one set of electric water heaters when the outage occurs, for instance, and a second set 30 minutes later, when the first set is restored to normal operation—grid operators could expand the range of reliability resources.

To date, however, regulators have used retail loads to support power system reliability primarily through special demand-response programs—not through bidding, in markets for energy, congestion management, and ancillary services. Yet, if regulators would provide for retail loads to participate directly in wholesale power markets, those markets would expand in scope. Such participation would likely lead to lower prices (especially price spikes that are less severe), fewer opportunities for the exercise of market power, and improved reliability.

Yet encouraging such demand participation requires a careful review of existing reliability rules and market designs to ensure they do not unfairly exclude resources that can provide valuable services to the grid.

The fundamental issue here is how to get the regional reliability councils and the ISOs to think more broadly about the resources that can provide reliability services, how to value and pay for the reliability services these resources provide, and how to cost-effectively deploy such resources.

In this article we explore those options. We explain the nature and characteristics of ancillary services for contingency reserves, including the technical and reliability requirements imposed on resources that now provide these services. Also, we examine the design and results of markets for contingency reserves, plus the desirable characteristics of retail loads that might provide such reserves, and various ideas that might encourage participation in reserve markets.

Overall, we believe retail loads offer a substantial potential for aiding power system reliability through the supply of contingency reserves. Modifying the reliability requirements to accommodate demand resources and include them in revised markets should improve the efficiency of wholesale energy, ancillary-service, and congestion-management markets.

Reliability Rules

To ensure power system reliability, grid system operators impose various performance, metering, and communication requirements on resources that provide contingency reserves.

In terms of performance, the resource must demonstrate the claimed ramping capability (in megawatts per minute, or MW/min). In addition, the resource must be able to sustain the committed output for a minimum amount of time, typically an hour or more. Also, the resource must then be able to ramp down within a specified time to its pre-contingency level so that it is positioned to respond to another outage (restoration).

These capability requirements ensure that, during an emergency, the resource will be able to respond as rapidly as required, and that the ISO can meet the disturbance control standard (DCS), as defined by the North American Electric Reliability Council (NERC), in its Policy 1, “Generation Control and Performance.” This policy specifies two standards that control areas must meet to maintain reliability in real

By Eric Hirst and Brendan Kirby
time. The Control Performance Standard (CPS) covers normal operations and the DCS deals with recovery from major generator or transmission outages. Three contingency reserves are deployed throughout the Northeast: the 10-minute spinning reserve, 10-minute nonspinning (supplemental) reserve, and 30-minute (replacement) reserve. The three services are used to help control-area operators meet the DCS. For our purposes, note only that DCS requires that the system recovers from a major outage—one between 80 percent and 100 percent of the largest single contingency—within 15 minutes. For more details, see Table 1, Definitions of Real-Power Ancillary Services.

The three reserve services provide responses of different quality. Spinning reserve is the most valuable service, and therefore generally the most expensive because it requires the generator to be on line and synchronized to the grid. Because such generators are on line, 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 within seconds. Supplemental reserve, which could include generators that are already on line, is less valuable because it does not necessarily provide an immediate response to an outage. Both spinning and supplemental reserves must reach their committed output within 10 minutes of being called on by the system operator. Replacement reserve is less valuable still because it need not respond fully until 30 minutes after being deployed. Replacement reserves are used to permit the restoration of the 10-minute reserves so that these faster-acting resources are, once again, able to respond to a new emergency.

NERC’s DCS is a performance measure; it specifies what must be accomplished (recovery within 15 minutes) without specifying how that goal must be reached. Only in California do some retail loads (large water-pumping loads, to be specific) provide contingency reserves.

New England. Since ISO New England began operating real-time markets for energy and ancillary services in May 1999, it has experienced problems with its markets for the reserve services. Complications in the design of the ISO’s day-ahead unit-commitment and its five-minute security-constrained dispatch prevented it from notifying beforehand the winning bidders in its ancillary-services markets. As a consequence, generators did not know whether they were “selected” to provide operating reserves until after the fact. In addition, the ISO might, during a major outage, call upon units that were not selected to provide reserves, and therefore they did not get paid for providing the service. In August 1999, ISO New England filed emergency market revisions with FERC. In response to the ISO’s request, FERC permitted the ISO to cap the prices of operating reserves at the current hour’s energy price.

The prices paid by ISO New England for reserves may have little meaning because of flaws in the ISO’s reserve markets. During the three-year period from January 2000 through December 2002, the price of spinning reserve averaged $1.15, the price of supplemental reserve averaged $2.08, and the price of replacement reserve averaged $0.81/MWh. (During 2002, the prices averaged $1.68, $1.67, and $1.10/MWh, respectively.)

New England implemented a new, improved market design in March 2003, based on the PJM design. This new market system, however, does not include PJM’s two-part market for spinning reserve. ISO New England has not yet decided on the structure of its markets for contingency reserves and, therefore, may have no operating markets for any of the contingency reserves until late 2003.

New York. The New York ISO operates an integrated set of markets for energy, real-power ancillary services, and congestion management. Because of the severity of transmission constraints in New York, especially in New York City and Long Island, New York’s reserve markets have three zones. Prices in the New York ISO ancillary-service markets, which do not contain the flaws that the New England markets have, might be a more reasonable indicator of what prices should be in a well-functioning market. New York, like New England, acquires roughly 600 MW of each of the three reserve services each hour. For the two-year period from January 2001 through December 2002, the prices of spinning, supplemental, and replacement reserve in New York averaged $2.74, $1.69, and $1.16/MWh, respectively. This ordering of prices is consistent with the value of each service. Both spinning reserve is the most valuable and replacement reserve the least valuable. (The New England prices, on average, did not follow this order.)

Mid-Atlantic. Until December 2002, PJM had no markets for contingency reserves. Any generator committed for service by PJM is guaranteed recovery of the costs associated with startup and no-load costs. To the extent these costs are not recovered from energy markets each day, PJM pays these units the difference between their operating costs and revenues for the day. These uplift costs were collected from PJM customers through an operating-reserve payment, although the nexus between these costs and reserves is ambiguous.

Beginning Dec. 1, 2002, PJM began operating a two-tier market for spinning reserve. PJM does not yet operate markets for the other contingency reserves. Tier 1 consists of units online, following economic dispatch, and able to ramp up in response to a contingency. These units receive no upfront reservation payment but do receive an extra $50 to $100/MWh for energy produced during a DCS event. Tier 2 consists of additional capacity synchronized to the grid, including condensing units, that can provide spinning reserve. These units are paid a reservation charge, based on a real-time market-clearing price, but receive no extra energy payment during a reserve pickup. FERC approved the PJM market, noting, however, that it “does not contain all the attributes contemplated by the Commission in the SMD NOPR, and the PJM proposal is different from the spinning reserve markets in New York and New England.”

The PJM markets for spinning reserve appear to be aimed at particular kinds of generating units, perhaps in recognition of the fleet of generators within its control area. As a consequence, the market design is hostile to demand resources in that there is no way for retail loads to participate in these markets.

Confederated with SMD. The SMD as proposed by FERC would require day-ahead markets for spinning and

### Table 1: Definitions of Real-Power Ancillary Services

<table>
<thead>
<tr>
<th>Service</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spinning reserve</td>
<td>Generators on line, synchronized to the grid, that can increase output immediately in response to a major generator or transmission outage</td>
</tr>
<tr>
<td>Supplemental reserve</td>
<td>Same as spinning reserve, but need not respond immediately; therefore units can be offline but still must be capable of responding</td>
</tr>
<tr>
<td>Replacement reserve</td>
<td>Same as supplemental reserve, but with a 30-minute response time, used to restore spinning and supplemental reserves to their pre-contingency status</td>
</tr>
</tbody>
</table>

### Table 2: NERC Contingency-Reserve Requirements

<table>
<thead>
<tr>
<th>Requirement</th>
<th>10-minute reserve</th>
<th>30-minute reserve</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amount required</td>
<td>100% of first contingency</td>
<td>50% of second contingency</td>
</tr>
<tr>
<td>Maximum response time</td>
<td>10 minutes</td>
<td>30 minutes</td>
</tr>
<tr>
<td>% of reserve that must be spinning*</td>
<td>25 to 100%</td>
<td>0</td>
</tr>
<tr>
<td>Minimum sustainable time</td>
<td>1 hour</td>
<td>1 hour</td>
</tr>
<tr>
<td>Maximum restoration time</td>
<td>90 to 100 minutes*</td>
<td>4 hours</td>
</tr>
</tbody>
</table>

* The percentage of 10-minute reserve that must be spinning (synchronized) depends on the performance of the control area in recovering from DCS events, within the required 10 minutes. The percentage of 30-minute reserve that must be spinning (synchronized) depends on the performance of the control area in recovering from DCS-reportable events within the required 15 minutes.
Thus, a retail load selling reserves can view in this context, the participating retail load should be able to interrupt service (90 to 105 minutes after the contingency occurred), within the time required by the regional reliability council for restoration. These real-time markets would differ from the day-ahead markets in that potential suppliers would not be permitted to submit availability bids. In other words, the prices for each reserve service in real time would be a function of the contingency-related opportunity costs. FERC is clear that it wants these ancillary-service markets to be open to demand-side resources as well as generators.

Using Demand Resources: Needs and Opportunities

In the first instance, the characteristics required of contingency reserves, as determined by NERC and the regional reliability councils (see Table 2), should determine the desirable attributes of the demand resources that might provide these services. Ideally, the participating retail load should be able to interrupt service immediately, sustain the interruption for the amount of time required by the regional reliability council, restore to full load within the time required by the regional reliability council for restoration (90 to 105 minutes after the contingency occurred), and then be ready to be interrupted again. Then, the economic value of DCS events occurs rarely enough to be worth the effort. Thus, a retail load selling reserves can count on a modest reservation (capacity) payment hour after hour, and only an occasional interruption. Viewed in this light, the desirable demand characteristics might be driven as much by financial and convenience considerations as by physical characteristics.

Some industrial loads (such as a production line) might be able to shut down in response to an emergency on the electrical system. The high cost of shutting down and restarting an entire production process suggests that such a resource might be called upon only when the interruption is long. Such a large industrial load, therefore, is quite different from residential water heaters. Households with electric water heaters are unlikely to notice any performance degradation if the duration of the interruption is short. In addition, water heaters can be turned back on again very quickly, and be turned off, once again, to provide contingency reserves. Other resources take much longer to be restored and rearmed to provide reserves. Thus, different retail loads are well suited to provide different services to the bulk electric system.

An alternative way to view demand-side provision of contingency reserves is to ask what the system operator really needs to maintain reliability. After all, the current rules were designed to accommodate large generating units, not demand resources. A more flexible set of performance-based requirements would likely encourage demand participation and improve reliability. For example, rules that require a certain resource must maintain its emergency output or load reduction for the 60 minutes specified by NPPC. DCS performance could be just as good if some loads responded immediately and were then replaced by other load reductions after, say, 30 minutes. With this simple modification to the NPPC requirements, loads that can interrupt for 30 minutes, but not for 60 minutes, would be able to provide contingency reserves. However, the 60-minute requirement would reduce by 50 percent the amount of contingency reserves provided by loads relative to a

30-minute requirement for sustained output. Such a rule change would expand the amount of resources that could participate in ISO contingency-reserve markets, thereby improving reliability and reducing the costs of doing so. Table 3 summarizes the characteristics loads must meet to provide contingency reserves.

Achieving the Vision: Nine Recommendations

To help realize the potential benefits of demand-side participation, we suggest a list of nine recommended actions.

1. Set Up Revenue Markets. ISOs should, as soon as possible, design and open markets for all three contingency-reserve services—the 10-minute spinning reserve, 10-minute nonspinning (supplemental) reserve, and 30-minute (replacement) reserve. Without functioning markets for the reserves, it is difficult to see how retail loads could provide—and be compensated fairly for—these services.

ISOs should implement markets that follow closely FERC’s SMD proposal, as exemplified by the New York market. In particular, they should adopt a day-ahead market design that integrates availability bids for the reserve services with energy bids and integrates reserves and energy in real time.

2. Invite Loads to Bid. Loads would participate in the day-ahead reserve markets by submitting availability bids and the energy strike price (both in $/MWh) above which they would be willing to interrupt some load. Accepted load and generator bids would be treated the same way; in the event of a major outage, the ISO would dispatch generators and loads in economic merit order. Loads and generators that failed to respond to the ISO’s dispatch signal during a DCS event would face the same nonperformance penalties.

3. Review Regional Reliability Rules. The regional reliability councils should continue to review their requirements related to DCS and contingency reserves to ensure they are truly technology neutral. In addition, the councils should publish the results of the engineering and economic analyses used to justify these standards and rules.

4. Make Rules Technology-Neutral. The NPPC requirements (see Table 2) were designed to accommodate typical generating units and are likely unsuitable for demand resources that might fully satisfy appropriate real-time contingency requirements. For example, NPCC offers no justification for the 60-minute minimum duration of reserves. Longer duration may improve reliability, but it also raises costs and limits the number and type of resources that can provide reserves.

Where, one might ask, are the data and analysis showing the economic costs and benefits of different duration times? Or, for that matter, the other parameters shown in Table 2? The rules should recognize the technical differences between reserves provided by large resources (whose expected performance is generally deterministic) and small resources (whose expected performance is generally statistical). The rules also should accommodate resources whose availability and size vary, especially for those resources where the variability is positively correlated with system load (in particular, weather-sensitive loads). These rules should address the reliability requirements associated with speed of response, duration of response, and speed of restoration.

4. Examine Metering Rules. The ISOs should review the requirements they impose on resources that provide contingency reserves with respect to the frequency of metering output (or consumption) and the frequency with which these megawatt values are communicated to the ISO’s control centers.

Lengthen Intervals for Reporting. The four-second recording and reporting requirement imposed on generators is probably not needed for retail loads that provide contingency reserves, primarily because of the much smaller size of these demand resources. It may be sufficient for large loads to record load data at the one- or five-minute level for 10-minute reserves and the five- or 10-minute level for 30-minute reserves and then report results to the ISO at the end of each month for verification and billing purposes.

For small load resources, such as residential water heaters, it should be sufficient to carefully meter only a small fraction of the loads and then scale up to the population of participating loads. In both cases, there may be no reliability reason to report performance results to the ISO in near real time; it may be sufficient to provide such data at the end of each month for billing and settlement purposes.

7. Assess Load Characteristics. ISOs, distribution utilities, and state energy offices and regulatory commissions should work together to characterize the potential demand resource for reserves in each region. This assessment would examine opportunities in the residential, commercial, and industrial sectors to see which customers and which end uses are suitable for the provision of contingency reserves. This characterization also would examine the seasonal characteristics of different resources that can provide reserves.
loads, their storage capabilities, the speed with which the load can be interrupted and rearmed (restored), and the costs of the necessary metering and communications equipment. The resulting estimates of resource potential will be a function of reliability and market rules as well as the payments to retail loads for provision of reserve services.

8. Encourage Demand Participation. ISOs, distribution utilities, and state energy offices and regulators should encourage loads to provide contingency reserves and to participate in the ISO markets for these reserve services. To stimulate such participation, the ISO should work with load-serving entities and other load aggregators to combine many small loads. Such aggregation should improve greatly the economics of load participation in these markets.

The ISO could, based on the prior recommendation, work with the load aggregators to develop metering and communication requirements that meet the ISO’s legitimate reliability needs and accommodate the needs of the load aggregators and individual retail customers. In addition, ISOs and load-serving entities (LSEs) should educate customers on bulk-power reliability issues, the importance of contingency reserves, and the role that demand resources can play in cost-effectively providing these reserves. Finally, ISOs might establish pilot programs to demonstrate the market barriers, benefits, and costs of using large and small loads to provide contingency reserves. Such programs could involve a few large industrial loads and an aggregation of residential loads (perhaps through a utility’s existing direct-load-control program).

9. Design Protocols for Load Aggregation. The ISOs should, working with LSEs and others, design load-research protocols that could be used when reserves are provided by aggregations of many small loads and which could substitute for the traditional performance measurement used for generators. Such protocols would measure the load-reductions of various types of loads under different conditions (time of day, day of the week, and season) and develop methods to forecast expected load reductions from different types of loads participating in contingency-reserve markets.

These recommendations, while required to accommodate demand-side resources in participating in markets for contingency reserves, need not imply preferential treatment for any one class of resources. Rather, rules should be modified simply to incorporate a broad consideration of economic costs and benefits.

Eric Hirst is an electric-industry restructuring consultant in Bellingham, Wash. Brendan Kirby is a senior researcher at Oak Ridge National Laboratory in Oak Ridge, Tenn.

Endnotes


3. Although NERC requires recovery from a major disturbance within 15 minutes, the control-area operators require the resources providing contingency reserves to respond fully within 10 minutes. The extra five minutes often are needed by the operators to decide whether a major contingency has occurred and, if so, how best to respond.

4. Until November 2002, NERC’s Policy 1 was prescriptive. NERC required that, with some exceptions, at least 50 percent of the 10-minute reserves be spinning. Perhaps more important, NERC restricted spinning reserve to “unloaded generation that is synchronized and ready to serve additional demand.” Clearly, this statement excluded customer loads from providing that valuable ancillary service. NERC’s new Policy 1 permits contingency reserves to be “supplied from generation, controllable load resources, or coordinated adjustments to Interchange Schedules,” a very important change for demand resources.


9. It is baffling that a competitive market would be designed to pay resources providing the identical service different amounts, and in different ways, based solely on the cost of the resource to providing the service.


11. New England has averaged 14 DCS events a year during the past five years. This is about the same rate experienced in New York and PJM.

12. We assume that retail loads will be paid for reserves just as generators are. They will receive an hourly reservation payment based on the price set in the day-ahead market and, when called upon to reduce load, they will enjoy the benefit of a lower energy payment during this time of higher energy prices. That is, loads would not receive an additional energy payment for interrupting during a DCS event.