

DEMAND RESPONSE: A MARKET OVERVIEW



Issue Brief

Katrina Managan

Program Manager, Institute for Building Efficiency

INTRODUCTION

Utilities carefully manage the electricity supply to be sure there is always enough available on the grid to reliably meet customer demand. Sometimes the most cost-effective way to balance supply and demand is to pay willing customers to curtail demand. These are called demand response programs, and customers who participate in them are called participating customers. Demand response programs have evolved and expanded over time. This paper looks at that evolution and describes each type of demand response program.

Historically, demand response was designed to help utilities and grid operators lower the cost of meeting large peaks in demand. The advent of demand response coincided with the proliferation of residential air conditioning units in the 1970s, which caused particularly high demand peaks on the hottest summer days. Demand response can also help with utilities deal with winter peaks. These are generally less extreme than the highest summer peaks, but often occur when some power plants are taken down for maintenance and less supply is available. In addition, demand response programs have evolved to provide grid regulation services.

Demand response helps operators facing four types of challenges:

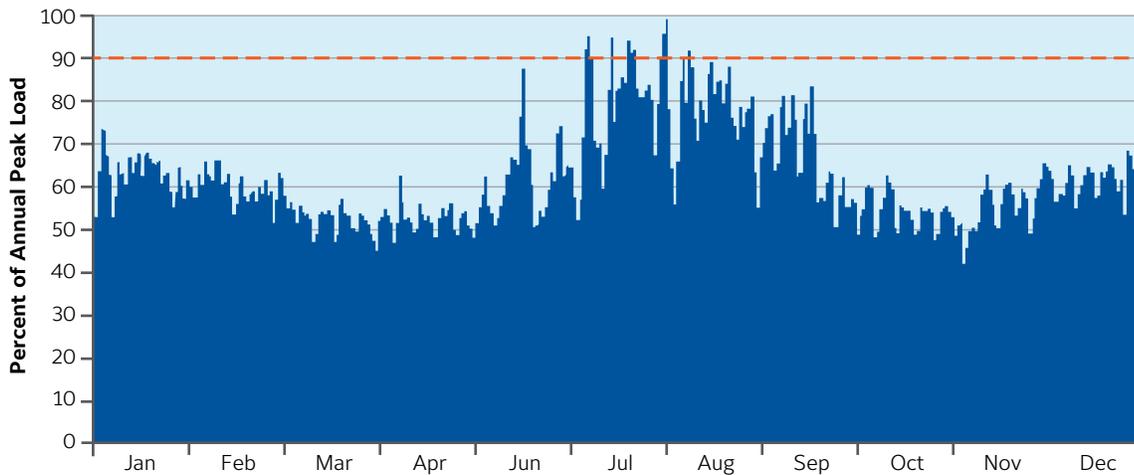
- The need to forestall building additional power plants as demand rises
- More intermittent renewable energy resources connecting to the grid
- Old power plants being retired
- Increasing transmission congestion

With these grid changes, there are now two major categories of demand response: emergency response programs and price response programs. In addition, new demand response programs are emerging to address and support the transmission of electric power from seller to purchaser, a concept called ancillary services.

CAPACITY AND ENERGY PROGRAMS – DEALING WITH MAJOR FLUCTUATIONS IN SUPPLY AND DEMAND

Demand response programs were originally created to help utilities lower the biggest peaks in demand on hot summer days. The following chart shows that the total system load for utilities or grid operators is rarely above 90 percent of total capacity. Without demand response, these peaks in load (generally two to 10 per year) were expensive to deal with, as utilities had to build and run extra capacity just to cover a few peak hours on the hottest days of summer. As an alternative, they established demand response programs in which customers are paid to reduce demand during those peak times.

Figure 1. Peak Load Planning¹



¹ PJM system load data from 2012 (data on PJM web site).

Demand response programs pay participating customers in two ways: either for a reduction in capacity (kW) or energy (kWh). Capacity (kW) is the ability of a power plant to make power – the maximum output capacity at a given moment in time. For a demand response participating customer, capacity is the ability to curtail load, or the maximum capacity for curtailment. Energy (kWh) is the total amount of energy used over time (an hour, month or year). For a demand response participating customer, energy is the total actual curtailment.

Programs also differ in how they “call” in the required changes. Programs either sign up customers to curtail load when the utility calls them (emergency/standby), or allow customers to bid a power reduction into the market a day ahead of time when electricity prices are projected to be high (price response).

EMERGENCY/STANDBY

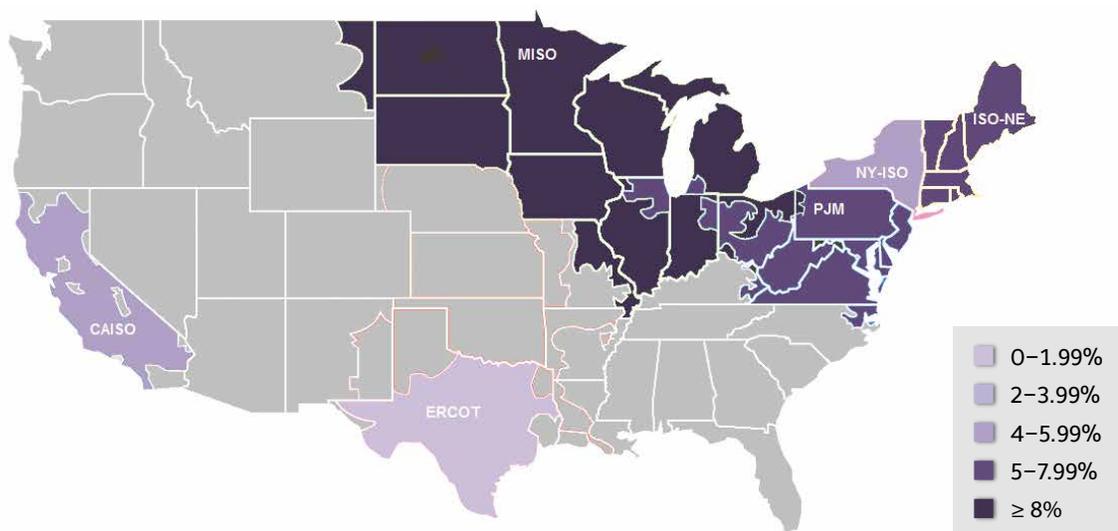
The basic idea behind emergency or interruptible load demand response programs is that the utility pays participating customers to shed a portion of their electric load during peak load events on the grid, or during power system emergencies or contingencies, including system component failures. In order to maintain grid reliability, utilities always have extra capacity available to cover spikes in demand created by any of these events – this is called capacity reserve margin. Demand response can be part of how utilities quickly replace that capacity reserve margin in case another event should occur before the first one is over.

Customers participating in demand response programs agree to a pre-specified amount of demand reduction. This allows the utility to call on those customers and maintain capacity reserve margin, protecting grid reliability. The payment ideally benefits both the customer and the utility. For the utility, the demand response payments are generally more economical than the investment in extra generation capacity to cover the peak hours. For the participating customers, any loss in production or inconvenience in operations due to curtailing energy must cost less than the payments they receive for that curtailment.

Participating customers receive payments for being on stand-by (capacity payments for the kW promised). Under some programs, they receive additional payments when they are actually called (energy payments for kWh avoided). Payments are proportional to how much a customer commits to reduce. Participating facilities receive advance notice of grid stress so they know ahead of time when they will be called. Failure to curtail when called can result in penalties. Typically, participating customers receive only a handful of calls per year, although the number of calls has increased recently in some regions as utilities take greater advantage of the demand response resource.

Emergency or standby DR makes up the majority of demand response programs today by megawatt, representing 87 percent of demand reduction capabilities across all reliability regions.² An estimation of these programs' saturation for summer 2013 is shown below.

Figure 2. Demand Response Emergency/Standby Saturation (as a percentage of total internal demand)³



² NERC's 2012 State of Reliability

³ Map produced by John Ellis, Global Energy & Sustainability MBA Intern, Institute for Building Efficiency, Johnson Controls, based on data from the NERC 2013 Summer Reliability Assessment.

Total internal demand is the sum of the metered (net) output of all generators within the system and the metered line flows into the system, less the metered line flows out of the system. DR participation can vary wildly for states within the same ISO. This is due to the overarching price for DR resources within the ISO vs the actual price of electricity. How the states set their electricity rates will have a huge effect as to whether or not DR payments are worth it for consumers.

Generally, the types of demand reduction actions taken in emergency or standby programs include:

- Direct load control, in which the utility remotely cycles residential cooling and water heating
- Customer-controlled shutdown of some process equipment (for industrial participating customers)
- Customer-controlled modification of temperature set-points or shutoff of lighting (for commercial customers).

All emergency/standby programs are either energy or capacity programs, or both.

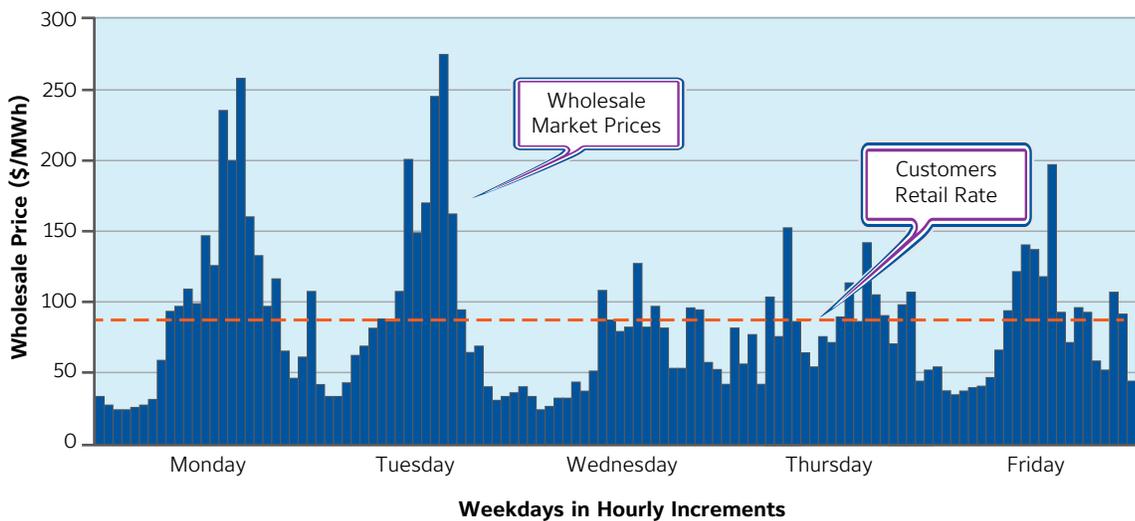
The success of emergency DR programs and DR participation in wholesale markets is driven by a number of factors including adequate incentive payments, infrequent calls, and short-duration curtailment periods (the average response period in summer 2011 was 2 hours and 51 minutes). Success has also been driven by regulatory support, participation by aggregators (curtailment service providers), reductions in the cost of hardware and software required for participation, and uniform rules for participation in large regions like the PJM regional transmission organization (versus different requirements for separate utilities).

PRICE RESPONSE

As utilities gained confidence in the ability of emergency and standby demand response programs to provide reliable extra capacity during grid events, they began to look at the potential for demand response to meet other needs of the grid. One key challenge utilities face is spikes in the wholesale price of electricity. Wholesale prices can fluctuate dramatically, rising especially high on a hot summer day when the demand for electricity is high.

The following chart shows the fluctuations in the wholesale price of electricity in Philadelphia during a hot summer week. The graph shows how the wholesale market prices often rose above the rate the utility could charge retail customers. The flip side to this challenge, as shown in the graph, is that historically, electricity rates are often flat or two-tiered, meaning there is usually a complete disconnect between what customers pay for electricity and what is actually costs. That is changing and, in theory, could eventually lead to real-time pricing. But without tiered/real-time pricing, customers have no price signal to restrict demand on hot days and no incentive to curtail use, unless enrolled in a demand response program.

Figure 3. Wholesale Prices for a Hot Summer Week in Philadelphia⁴



⁴ PJM pricing data in the PECO zone from a week in 2007.

Opportunities for load to respond to wholesale market price signals have been developed to achieve more competitive markets by enabling buyers to respond to price signals and obtain financial benefits for doing so. These demand response programs help protect utilities from large fluctuations in wholesale electricity market prices. Offers to produce electricity historically only involved generators. The paradigm has expanded so that in markets with price response programs, energy consumers can offer to curtail their load for the going price of power as a "demand capacity resource." The system operator can then select the lowest-cost approach to ensuring that the grid is balanced – by accepting offers for more power, selecting lower-cost offers to reduce demand, or a combination. These programs are also referred to as demand bidding and buy-back.

On the customer side of the meter, there may be a variety of ways to manage power demand when a bid is accepted. For example, the building may have technology on site to make ice overnight, and then use the ice to cool the building during the heat of the day, thereby reducing the peak-hour electricity demand. Similarly, some companies use batteries to store electricity during low-cost times and deploy it during peak times. Price response programs are flexible, allowing participating customers to decide when to participate. In some programs, capacity bidding is done the day ahead; in others, customers can react to hourly price signals. Participation is accomplished through either automation or manual control. Interval metering and communication systems are required so that the utility can measure and verify that the customer did in fact curtail load when requested. Customers may not participate in both emergency and price response opportunities in the same hour because the resulting reductions could then be double-counted.

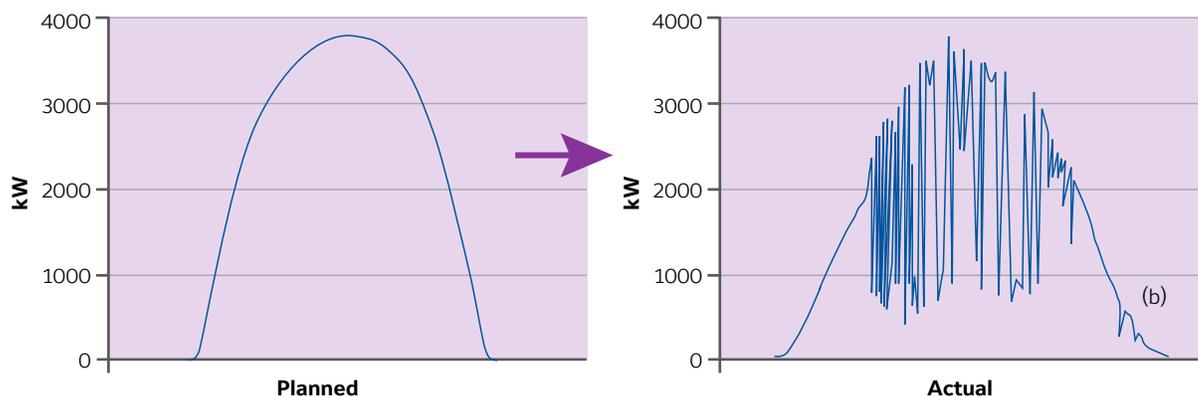
ANCILLARY SERVICES – DEALING WITH SMALLER FLUCTUATIONS IN SUPPLY AND DEMAND

While both emergency and price response programs are designed to provide a volume of capacity or energy reduction, another form of DR is emerging to help improve grid reliability by aiding the transmission of electricity. Ancillary services on the grid enable the reliable transmission of electric power. The need for ancillary services is growing as grid operators face new challenges, including transmission congestion and the increasing role of intermittent renewable power generation.

Demand response has the potential to be a critical part of the response to the intermittency of renewable energy and other challenges facing the grid: It could provide ancillary services to curtail demand at times when it would be costly or difficult for the grid operators to supply that excess demand. Ancillary services make up only a small portion of demand response program enrollment today, but these programs are growing quickly.

Figure 4. Intermittency of Renewables Generation⁵

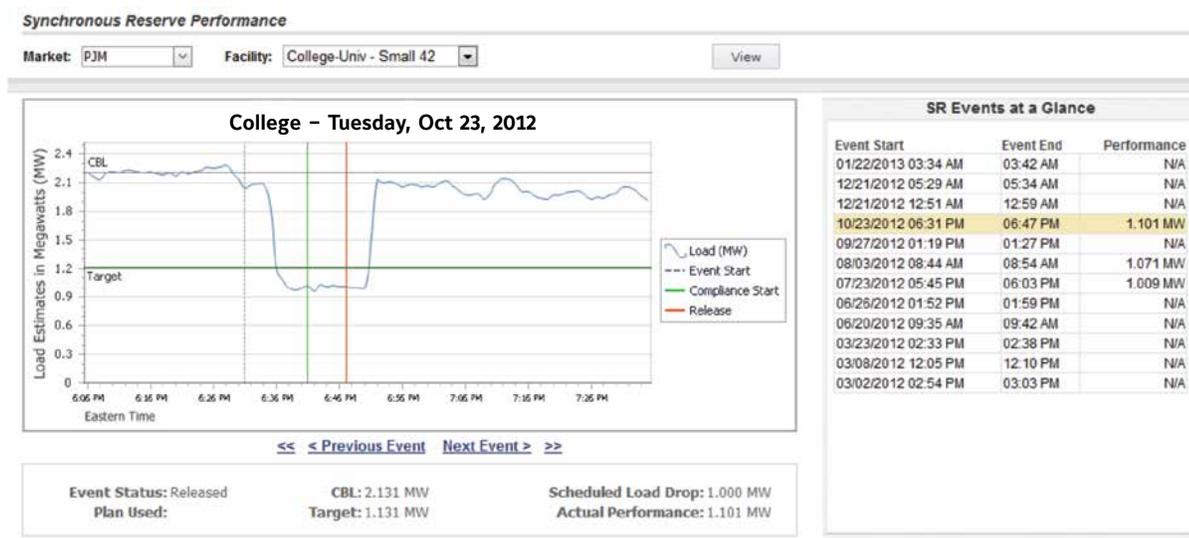
⁵ Jay Apt CMU, 4.6 MW TEP Solar Array (Arizona)



Types of ancillary services include synchronized reserve and frequency regulation. Synchronized reserve supplies electricity if the grid has an unexpected need for more power on short notice – it is the rapid balancing of supply and demand. Frequency regulation manages the quality of energy on the grid, balancing generation and load to keep the system frequency reliably at 60 Hz. Both synchronized reserve and frequency regulation are extremely important because problems with either can lead to brownouts/blackouts and damaged equipment for both the utility and customers. The response time required for ancillary services tends to be faster than for other demand response programs, and therefore the payment per megawatt curtailed tends to be larger.

Synchronized reserve (also called spinning reserve or 10-minute reserve) demand response programs compensate customers for the ability to provide a predetermined amount of energy in order to meet a reliability standard. Utilities can lower excess standby capacity if they know they can call on reserve resources if demand rises higher than expected for a brief period, or if a generator suddenly fails. Both generators and demand resources can provide this response, by either adding or removing energy quickly. These resources usually respond within 10 minutes with typical response duration of only 5 to 30 minutes. Resources participating in synchronized reserve programs are generally dispatched a few times a month. These programs are a good fit for utility customers with large chillers that have automatic controls. The following chart shows how a building on a college campus was able to drop its load by just over 1000 kW in about 10 minutes when called upon as synchronous reserve.

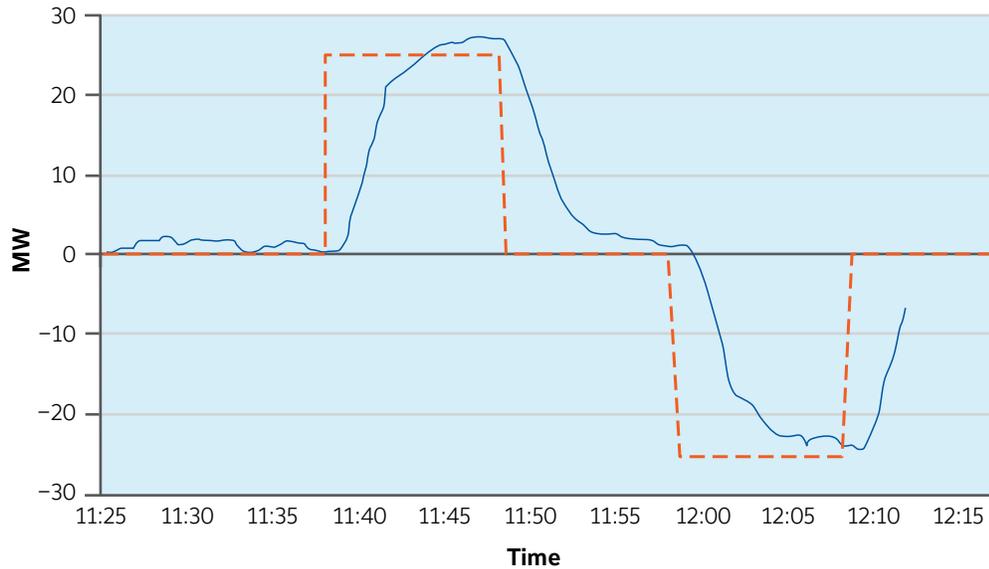
Figure 5. Synchronous Reserve Performance⁶



⁶ Screenshot from JCI demand response application.

Frequency regulation demand response programs compensate customers for the ability to modulate demand up or down (up-reg, down-reg) in order to maintain grid frequency and support reliability. Frequency regulation services ideally need to respond within seconds to balance out the frequency on the grid. A utility customer gets real-time signals from the grid and must have the capacity to react quickly by either increasing or decreasing their load to balance frequency (usually automatically). Customers can opt to automate their equipment control systems to rapidly adjust their load in response to utility calls for a demand response event; a full response is required within 5 minutes.

Figure 6. Frequency Regulation



Demand response allows the utility to have less excess capacity ready to balance frequency on the grid. As renewables become more prevalent, it will be increasingly important to have demand response frequency regulation resources. This is because with distributed generation, utilities cannot regulate power production the way they do with large centralized generators. As demand response programs find a way to help larger portions of the demand pool to participate, the grid will operate with greater overall efficiency and will require fewer large power plants.

FERC Order 755 established pay-for-performance requirements for frequency regulation. The order requires transmission providers to compensate participating customers based on the speed and accuracy of their response, in addition to the quantity of demand curtailed. Energy storage systems inherently excel in speed, accuracy, and ramping ability, so the order may create more opportunities for energy storage projects to participate in the ancillary services market.

OVERVIEW OF FOUR KEY MARKETS: CALIFORNIA, ERCOT, NY, AND PJM

The table below shows an overview of the types of demand response programs available in each of four key markets for demand response.

	Emergency	Price Response	Reserve	Regulation
California	✓	✓		
ERCOT	✓		✓	✓
NYISO	✓	✓	✓	✓
PJM	✓	✓	✓	✓

California Independent System Operator (CAISO)⁷

California's Renewable Portfolio Standard (RPS) establishes a 33 percent target for renewable energy sources in the statewide generation portfolio by 2020.⁸ Due to the real-time variability of renewable power generation and typical seasonal demand peaks, demand response will be needed to help balance supply and demand. Currently, California utilities offer only energy and capacity programs. The CAISO market is structured so that most demand response programs are operated by local utilities, the largest of which are Pacific Gas & Electric (PG&E), Southern California Edison (SCE), and San Diego Gas & Electric (SDG&E). All three offer price response programs as well as interruptible load and capacity bidding programs, which deliver energy and capacity payments. However, reserve and regulation programs are either in a pilot stage or nonexistent. This creates a gap between the needs of a more intermittent statewide renewable energy portfolio and the support that could be provided through demand response. The lack of a program structure to capture reserve and regulation demand resources to match the evolving grid generation resources limits customer participation in ancillary services, reducing the system's capability to respond to fast-changing grid conditions.

⁷ <http://www.aiso.com/market/Pages/ProductsServices/Default.aspx>

⁸ Office of the Governor of California, 2008 <http://gov38.ca.gov/index.php?/executive-order/11072/>

Electric Reliability Council of Texas (ERCOT)⁹

Texas has an unusual electricity market in that it is for energy only, with no capacity market. This means that within the marketplace, customers participating in demand response programs are compensated only for demand reductions achieved and not for being available for demand response. Of the programs available, the Emergency Response Service (ERS) program operates the most similarly to those in other markets, with both an energy and capacity incentive payment for demand reduction. Auctions happen every four months in ERCOT, so there is no price certainty outside each four-month block.

⁹ <http://www.ercot.com/mktinfo/>

The ERS program offers regular incentives, and enrollment is easy, but there is only a short advance notice when a business's resources must be "cycled" (turned-off). Other demand programs in ERCOT are operated as reserves with short times for participating customers to comply with calls for demand reduction. In the Load Acting as a Resource (LaaR) program, customers offer bids into the marketplace and, if their bid is accepted, they are compensated only for the energy demand that is reduced. This program incorporates regulation services through a frequency responsive relay (on/off switch) that automatically removes the electrical load from the grid. Reserve margin projections show that Texas may be very short on capacity in a few years, and demand response may play a critical role in filling that gap.¹⁰

¹⁰ <http://www.greentechmedia.com/articles/read/5-trends-to-watch-in-demand-response>

New York Independent System Operator (NYISO)¹¹

New York's market offers a wide range of both economic and reliability-based demand response programs. Economic programs include the Day-Ahead Demand Response Program (DADRP), which allows demand response customers to bid into the market with a unit of demand reduction, and the Demand Side Ancillary Services Program (DSASP) for reserve and regulation resources. NYISO also offers an Emergency Demand Response Program (EDRP), used in energy shortage situations to maintain the reliability of the bulk power grid. Participating companies are paid by the NYISO for reducing energy consumption when asked to do so by the NYISO. Reductions are voluntary for EDRP participants. The NYISO's Special Case Resource (SCR) program allows demand response resources to participate in the NYISO's capacity market. Three types of auctions (six-month strip, monthly auction, spot market auction) are held, providing an advantage to aggregators who can opt in new customers as needed. The use of demand curves (developed on a three-year cycle) in the spot market auction as a proxy for load bids provides some degree of price certainty.

¹¹ http://www.nyiso.com/public/markets_operations/market_data/demand_response/index.jsp

PJM Interconnection LLC (PJM)¹²

PJM arguably has the most mature market, offering a wide variety of opportunities to participate in the energy, capacity, price response (day-ahead scheduling reserve), synchronized reserve and frequency regulation markets. Registration of customers by aggregators as Emergency Load Response resources enables participation in energy, day-ahead scheduling reserve, synchronized reserve and frequency regulation markets.

PJM has a three-year forward auction for its emergency demand response program; so prices are guaranteed for two to three years out, but not beyond that. To participate in the auction, aggregators need to know years in advance how much capacity they will have in each zone. This can be challenging.

¹² <http://www.pjm.com/markets-and-operations/demand-response.aspx>

The Role of Aggregators

Today, the market is structured so that aggregators (also called Curtailment Service Providers) play a role in helping customers find and participate in demand response programs. On the flip side, aggregators help utilities meet their demand reduction and pricing goals. Often, a single customer cannot contribute the minimum capacity required by utility to participate in demand response programs. Aggregators group multiple customers into one response unit, giving grid operators the volume of resources they need. Aggregators market programs to customers, not only helping enable their participation through aggregation, but also helping them find the appropriate program to fit their business needs.

In addition, by grouping customers into a large portfolio, aggregators also reduce the risk of any single customer not curtailing demand when called. This helps ensure the reliability of the demand response resource.

Depending on the particulars of the regional marketplace, the aggregated customer demand resources are offered during regular market cycles. In ERCOT, aggregators offer blocks of demand resources in the market every four months, while in PJM the bidding occurs each year in May. When a request (also called a dispatch) for demand response occurs, the need for demand response is relayed from the grid operator to the aggregator, which then relays the dispatch to the demand response customers. In the aggregated model, customers are compensated a portion of value of their demand reduction in exchange for participating. Some programs compensate a flat rate, while others offer a variable rate based on market energy prices.

CONCLUSION

Historically demand response was designed to help utilities and grid operators lower cost and ensure reliability during large peaks in demand on a few of the hottest summer days. These programs evolved into the two major categories of emergency and price response programs, which now represent 87 percent of demand response enrollment and can meet a significant portion of grid capacity when called upon. That said, enrollment in many regions could be significantly higher.

As more intermittent renewable energy comes onto the grid, old generation resources are retired, costly new plant expenditures are avoided, and transmission congestion becomes a larger problem, new demand response programs are emerging. These are designed to enable demand resources to provide ancillary services and mitigate small imbalances between supply and demand. Ancillary services demand response programs are still developing and have relatively low enrollment to date.

As all types of demand response programs continue to grow, they face the challenge of attracting participants. Specific structural features can make programs more or less attractive. The form of payments and incentives, the level of complexity, the degree of customer control, the frequency of calls, the length of curtailments, and the amount of notice vary by program design. It is essential to get these attributes right for each customer type to entice them to supply their demand response resources to the grid.

The second paper in this series from the Institute for Building Efficiency, "Increasing Demand for Demand Response," looks at these program attributes and at the process of matching customers to an appropriate program. It also presents a framework for analyzing where gaps may lie in programs' alignment with customer interests and needs.

ACKNOWLEDGEMENTS

Thank you to the following individuals who participated in background interviews, development of the paper, and/or provided a review of the paper. The views contained in this analysis are those of the authors alone and are not necessarily shared by individuals or organizations listed here.

Bruce Campbell, Johnson Controls

Chris Head

Colleen Snee, Johnson Controls

Jennifer Layke, Johnson Controls

John Ellis, Johnson Controls

Nick McLellan, Johnson Controls

Rich Quattrini, Johnson Controls

Susan Covino, PJM Interconnection LLC

The Institute for Building Efficiency is an initiative of Johnson Controls providing information and analysis of technologies, policies, and practices for efficient, high performance buildings and smart energy systems around the world. The Institute leverages the company's 125 years of global experience providing energy efficient solutions for buildings to support and complement the efforts of nonprofit organizations and industry associations. The Institute focuses on practical solutions that are innovative, cost-effective and scalable.

If you are interested in contacting the authors, or engaging with the Institute for Building Efficiency, please email us at: InstituteforBE@jci.com.

institute
for **building**
efficiency
an initiative of Johnson Controls