LOAD AS A RESOURCE IN PROVIDING ANCILLARY SERVICES

BRENDAN KIRBY and ERIC HIRST
Oak Ridge National Laboratory
and
Consultants in Electric-Industry Restructuring
Oak Ridge, Tennessee

Most of the commercially important ancillary services involve maintaining or restoring the generation and load real-power balance over varying time frames. Traditionally utilities have addressed this problem almost exclusively by controlling generation. It does not have to be this way, however. The important concept is to balance load and generation, which can be done using either side of the equation. Controlling load may be the single largest untapped resource currently available to the electricity industry. Restructuring is beginning to provide the framework within which this resource could be exploited. Several obstacles exist (primarily related to aggregation, communications, and economic incentives) but technical and commercial solutions to these problems exist.

WHICH SERVICES MIGHT LOADS SELL?

Table 1 presents the 12 ancillary services commonly discussed (Hirst and Kirby 1998). Clearly loads will not sell System Control, System Black Start, or Dynamic Scheduling. Energy Imbalance and Real-Power-Losses are primarily accounting services with the required physical energy and capacity coming from other services. Reactive Supply and Voltage Control from Generation and Network Stability are also services that loads are not likely to have the resources to supply (though if they do they should be encouraged to participate).

The five remaining services (Regulation, Load Following, Frequency Responsive Spinning Reserve, Supplemental Reserve, and Backup Supply Plan) deal with maintaining or restoring the real-energy balance between generators and loads. These services are characterized by the required response time, the response duration, and the communications and control required to facilitate the service. Figure 1 shows the required response for the five energy balancing functions. Because regulation requires continuous adjustment of real power transfers between the resource and the system it is unlikely that many loads will be capable of delivering that service at this time. (Water-pumping load, such as municipal-water and irrigation systems, may be a possible exception if they use variable-speed drives for their pumps.) The contingency reserves are especially amenable to being provided by loads. Load following could be provided by loads directly or through the use of a spot market price response on a shorter time frame than 1 hour.

Table 1. Key Ancillary Services and their definitions
System Control: Control-area operator reliability and commercial functions
Reactive Supply and Voltage Control from Generation: Injection and absorption of reactive power from generators to control transmission voltages
Regulation: Maintenance of the minute-to-minute generation/load balance to meet CPS 1 and 2
Load Following: Maintenance of the hour-to-hour generation/load balance
Frequency Responsive Spinning Reserve: Immediate (10-second) response to contingencies and frequency deviations
Supplemental Reserve: response to restore generation/load balance within 10 minutes of a generation or transmission contingency
Backup Supply Plan: Customer plan to restore system contingency reserves within 30 minutes if the customer's primary supply is disabled
Real-Power-Loss Replacement: Compensation for transmission-system losses
Energy Imbalance: Accounting for the hourly discrepance between scheduled and actual transactions
Dynamic Scheduling: Real-time metering, telemetering, and computer software and hardware to electronically transfer some or all of a generator's output or a customer's load from one control area to another
Network Stability: Use of fast-response equipment to maintain a secure transmission system
System Black Start: The capability to start generation and restore all or a major portion of the power system to service without support from the outside after a total collapse

The same types of restrictions apply to loads supplying ancillary services that apply to generators supplying the same services. For a generator to supply contingency reserves it must have capacity available to respond to the contingency. The generator cannot be operating at full load. Similarly, a customer selling contingency reserves must have capacity it

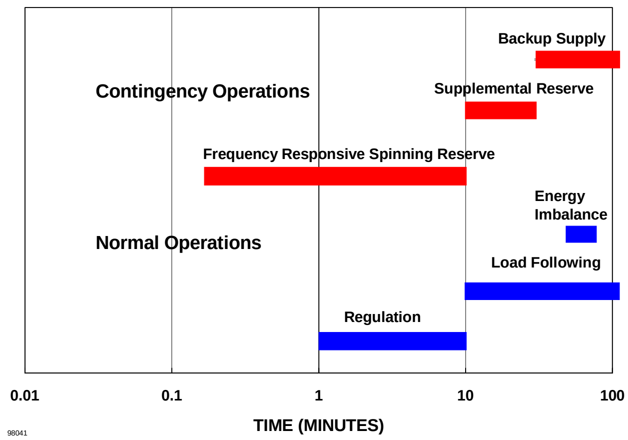

Fig. 1. Ancillary service dynamics.

can unload when the contingency occurs. The customer cannot already be at minimum load.

WHY UTILIZE LOAD AS A RESOURCE?

There are several reasons that loads should be encouraged to sell ancillary services. FERC is encouraging open competitive markets for generation. FERC ordered the unbundling of ancillary services to promote competitive markets, which should improve economic efficiency and lower electricity prices. These markets should be open to any technology capable of providing the service, not just to generators, which will expand supplies and reduce horizontal-market-power problems.

Beyond the argument of fairness, having loads participate as suppliers, as well as consumers, of electricity services improves resource utilization. Ancillary services consume generating capacity. When loads provide these reserves, generating capacity is freed up to do what it was designed for, i.e., generate electricity.

Loads will probably respond more quickly to control-center requests than large generators because the load response is composed of many small resources. This will likely more than overcome the communications and control delays associated with their greater numbers.

Loads should also be a more reliable supplier of ancillary services than conventional generators. Because each load will generally be supplying a smaller fraction of the total system requirement for each service, the failure of a single resource is less important. Just as a system with ten 100-MW power plants requires less contingency reserves than one with a single 1000-MW plant so too a system that utilizes a large aggregation of loads as a resource to supply reserves will require less redundancy in the basic resource than one that carries all of its reserves on a few large generators. There can still be common-mode failures in the facilities of the aggregator but it is easier and cheaper to install redundancy in this portion of the system than with an entire 1000-MW plant.

PROVISION OF INDIVIDUAL SERVICES

In all cases the owner of a load, in cooperation with an aggregator and the system operator, would determine the portion of the load which could provide the service. Metering, communications, and control requirements would then be established.

Looking first at the services required to restore the generation/load balance after a contingency *Supplemental Reserve* is a likely candidate for many loads. The resource must fully respond within 10 minutes of the contingency.* Response must be maintained for an additional 20 minutes, i.e., until 30 minutes after the contingency. This is a short interruption that many customers may find acceptable. Candidates include water pumping, building temperature control, water heaters, and air compressors. Anything that inherently has some storage in the process, or any process for which storage can be readily added is a good candidate.

The system operator takes some of the 10 minutes to recognize the contingency and to call for response. The aggregator's communications process will also consume some time. This leaves a few minutes for the load itself to respond.

Obviously, the load itself must be consuming power for it to curtail consumption during the contingency. Many candidate loads cycle as they provide service (e.g., hot water heaters). Since individual loads do not cycle together, the aggregation will always have some load available for curtailment. The aggregation has to be accurately characterized to know how much operating reserve is available at any time. Individual loads also have to be controlled after the reserve is released to prevent them from all returning to service simultaneously.

Frequency Responsive Spinning Reserve is both easier and more difficult for loads to provide. Because the service responds to system frequency, each load has the triggering signal available at all times. The service only has to be provided until it is replaced by Supplemental Reserve, 10 minutes, creating a shorter interruption. Full response is required within 10 seconds, however, which may make it harder for some loads to provide. A typical generator droop characteristic could be created by having each load in an aggregation respond at slightly different frequencies.

---

*Specific timing requirements for each service vary from region to region. The requirements referenced here are from NERC (1998) Draft Policy 10.

Frequency Responsive and Supplemental Reserves restore the system's generation/load balance and maintain it for 30 minutes. Thirty minutes after a contingency occurs the customer that was receiving the lost generation is responsible for making other arrangements or curtailing its load. The *Backup Supply Plan* is a pre-arrangement that tells the system operator how to proceed for each load's loss of primary supply. Some loads may find it attractive to provide Backup Supply for other loads. The 30-minute warning provides time for communications and for the curtailing load to take actions to reduce its own costs.

Loads may also wish to participate in maintaining the generation and load balance during normal operations, though this seems less likely. A load or aggregation of loads could provide *Load Following* by cycling their daily operations in response to direct MW commands from the system operator or by responding to short-term price signals (Hirst and Kirby 1997).

Regulation is the least likely of the generation/load balancing services to be provided by load. It is possible, however, that loads with variable speed drives (e.g., water pumping) could accept automatic-generation control signals from the system operator. Municipal water pumping accounts for approximately 1% of electricity consumption nationally, providing a potentially significant source of load-based regulation or other ancillary services.

PAST USE OF LOAD CONTROL

Load control has been and is currently used in a number of locations [NERC (1997b) *Operating Manual*]. Some implementations have been successful but the idea has not been universally adopted. This is at least partly because of traditional rate structures, which provide little flexibility to customers. The customer must agree up front to be subject to utility control, usually for a year or more. There is no ability to enter and leave the market as the customer's economic conditions change. The customer often gets paid a flat fee independent of how or if the resource is actually used. This provides little flexibility for the load and little incentive to actually perform.

Similarly, the cost of peaking generation or peak reserves are typically spread over an entire season or year. Charges (both operating and capital) are not assigned exclusively to the hours when the generation or reserves are required. Assigning the costs to the hours when they are needed would result in much higher prices for those services during specific hours (and lower prices at other times). Under either good economic regulation or a truly competitive market, the result would be the same total revenue collection (that required to pay for the resource). Providing a price signal that accurately reflects the real-time cost to provide the service will encourage all suppliers, loads and generators, to offer supply when it is needed most.

SUPPLIER CONTROL OF ITS FACILITY

While automatic deployment is necessary when selling some reserves, it is often important to allow the load to decide when it will participate and when it will not. Just as the price of hourly energy and each of the ancillary services vary, so do customer economics. For many customers there are times when less flexibility exists and the load cannot be interrupted without high costs being incurred. These times are often independent of anything happening on the power system and are therefore unrelated to the price of the service. For the right price, a residential customer might be willing to automatically curtail air-conditioning use for 30 minutes to supply contingency reserves, for example. This same customer would probably be unwilling to curtail use at any price on the evening when he was holding a dinner party, however. Similar restrictions might apply for an industrial customer such as a continuous chemical processing plant while it is taking a monthly inventory and needs a stable process. In both cases the customer choice not to participate is unrelated to the utility economics; neither load is trying to avoid providing the service when it is highest in value. In fact, the chemical plant may intentionally select times for its inventory when the power system is not stressed, such as at night or on weekends. It would do this not because of a concern for the power system but because that may be a time when the chemical process is stable as well due to reduced activity at the chemical plant.

The utility needs information about which loads will be supplying services ahead of time. The load must declare that it is available before it enters or leaves the market. Perhaps this declaration would be one day in advance for the following 24 hours. Both the utility and the load will need the ability to change the availability on shorter notice, perhaps with economic consequences. A load that experiences technical difficulties and is suddenly incapable of supplying the service must be able to leave the market. Conversely, if the power system finds itself unexpectedly short of reserves it will need to be able to call for additional reserves quickly, perhaps by raising the current price. Indeed, this is how the day-ahead, hour-ahead and real-time markets are intended to operate in California's competitive bulk-power system.

It is critical to avoid providing an incentive for a resource (either load or generation) to declare itself available when it is not (as is done in the United Kingdom). Equipment failures are inevitable but service providers should have an incentive to maintain the reliability of their resources. They should never find it profitable to sell a service that they know they cannot deliver.

TRADING CERTIFICATION FOR METERING

Most of the generators on a typical power system are relatively large and expensive. It is reasonable for the system operator to monitor unit output and bus voltage every 2 to 8 seconds. The amount of data and the expense per MW are both reasonable. When the operator calls for response the response can be monitored in real time.

Providing the same information from hundreds or thousands of individual resources would be prohibitively expensive and would provide an overwhelming amount of data that could not be managed in real time. An alternative to real-time monitoring of each individual resource exists. Loads could be certified, either individually or in aggregation, for the provision of each ancillary service. Certification would consist of exercising the resource under controlled conditions to determine the reliable response (NERC 1998). Testing of the contingency reserves, for example, would not be announced to the resource. The response would be measured on control-area metering. Periodic testing would monitor continued capability. Recording meters at each resource could also be audited to verify performance for both actual events and tests.

AGGREGATION AND COMMUNICATION

The major objection often voiced to customer supply of ancillary services is that the system operator cannot deal with the large number of individual resources and that the communications requirements would be overwhelming. These are valid concerns but ones that can be addressed. Aggregators can provide a genuinely valuable function here. By handling the communications with a large number of loads they can present the system operator with a single point of contact for a reasonable amount of capacity, similar to the system operator's interface with generating resources. They can also be an interpreter between the electrical system and customers. The system operator is not interested in learning the details and concerns of each customer. Similarly, customers are in businesses of their own and have neither the time nor the interest in learning all about the power system. The aggregator can bridge this gap, creating a valuable resource in the process.

Communications are inherently different with an aggregation of resources than with a single entity. As mentioned above, it is not currently practical to collect data from thousands of individual loads every 2 to 8 seconds. It is practical, however, to send instructions to those loads as fast as necessary. That is because it is the same signal going to large groups of the loads. That signal could be "deploy now" or it could be "the current price for response is $X".

CUSTOMER ECONOMICS

In competitive bulk-power markets, customers will have many choices with respect to their use of electricity and their payment for electricity services. In the context of this discussion, they can choose to participate in hourly markets and face spot prices that can vary widely in response to supply/demand relationships. Alternatively, they can sell reserve services (options) as discussed above. Decisions on whether to participate in spot markets or sell reserves will be based on the customer's flexibility in modifying its electricity use (in particular, its fixed and variable costs to modify its electricity use in real time), the prices of energy and reserve services, and the frequency with which outages occur.

For example, higher reserve prices and less frequent outages will lead customers to sell reserve services, forgoing opportunities to reduce consumption at times of high spot electricity prices. On the other hand, increasing flexibility (i.e., declining cost) in modifying electricity use will lead to more decisions to participate in spot energy markets.

CONCLUSIONS

Loads can provide several of the energy-balancing ancillary services. They should be encouraged to do so because of the reliability and commercial benefits provided by expanded supplies of these resources. Artificial barriers to the entry of customers into these markets should be removed. The customers themselves will have to determine their economic costs and benefits to see if and when they will participate. Loads must be given the flexibility to respond to their own economic constraints.

Technical problems associated with the need to aggregate many individual loads to achieve a resource of sufficient size to be useful can be overcome. This may be one of the areas where load aggregators can play a genuinely useful role. An aggregator can relieve the system operator of the burden of dealing with an excessive number of individual resources while still providing the required response. Similarly, the high communications burden associated with a large number of resources can be alleviated by broadcasting control signals to the resources. Rigorous certification, coupled with post-event meter audits, can substitute for real-time monitoring of each resource.

REFERENCES

E. Hirst and B. Kirby 1997, "Cutting Electricity Costs for Industrial Plants in a Real-Time World," *Public Utilities Fortnightly* **135**(22), 42–45, December.

E. Hirst and B. Kirby 1998, *The Functions, Metrics, Costs, and Prices for Three Ancillary Services*, Edison Electric Institute, Washington, DC, October.

North American Electric Reliability Council 1993, *Electric System Restoration*, Princeton, NJ, April 1.

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

North American Electric Reliability Council 1997b, *NERC Operating Manual*, Princeton, NJ, December.

North American Electric Reliability Council 1998, *Policy 10—Interconnected Operations Services*, draft, Princeton, NJ, November.

D:\Wpd\ANC\LoadAPC.wpd
January 11, 1999