ANCILLARY SERVICES

ERIC HIRST and BRENDAN KIRBY
Oak Ridge National Laboratory
Oak Ridge, Tennessee 37831

INTRODUCTION

Ancillary services are those functions performed by electrical generating, transmission, system-control, and distribution-system equipment and people to support the basic services of generating capacity, energy supply, and power delivery. The Federal Energy Regulatory Commission (FERC 1995) defined ancillary services as "those services necessary to support the transmission of electric power from seller to purchaser given the obligations of control areas and transmitting utilities within those control areas to maintain reliable operations of the interconnected transmission system."

FERC identified six ancillary services: reactive power and voltage control, loss compensation, scheduling and dispatch, load following, system protection, and energy imbalance. Our earlier work identified 19 ancillary services (Kirby, Hirst, and VanCoevering 1995). Here we offer a revised set of seven ancillary services and mention several other services that merit consideration (Hirst and Kirby 1996). The services presented here are based on the work of several others, including FERC (1995), Houston Lighting & Power (1995), the Michigan Public Service Commission (1995), the New York Power Pool (1995), and the North American Electric Reliability Council (NERC 1995).

In developing this set, we identify those services that are essential to maintain electric-system reliability, are required to effect a transaction, or are a consequence of a transaction. We exclude services that are optional, long-term in nature, too cheap to warrant the costs of metering and billing, naturally bundled with other services, or very location specific.

The existence, definition, and pricing of ancillary services is a function of industry structure. We assumed the continued existence of control areas and the NERC control-area concepts and requirements. Based on several recent proposals, we anticipate creation of independent system operators (ISOs) that will replace today's utility operation of control areas. These entities will not be controlled by generation-owning organizations. Rather, they will be *independent* organizations with the primary goal of operating the electrical system in real-time to maintain reliability. The ISO may also dispatch some or all of the generating units within the control area to minimize the cost of electricity production. Similarly, the ISO may create a wholesale spot market for power, although in some proposals such a power exchange is a separate entity.

Our list recognizes that the ISO might unbundle some services that it purchases (i.e., when facing suppliers) and unbundle services differently that it sells (i.e., when facing customers). In addition, services should be unbundled only if the ISO can identify and measure the amount of service provided by suppliers and/or consumed by customers. As computing and communication technologies improve, it may be possible to unbundle additional services. Finally, the incremental metering, accounting, billing, and auditing costs of unbundling must be less than the benefits of unbundling.

Our set of services includes scheduling and dispatch, which is a control-area operator function requiring few resources (computers, meters, communications equipment, and people). The set also includes several generating services, such as load-following, reliability, and supplemental reserves, as well as loss replacement and energy imbalance. For various reasons, discussed below, there is considerable confusion about the definitions and purposes of these services. Finally, we include system voltage control, which requires both generating units and transmission-system equipment (Table 1). All of these services are required. Altogether, these services cost U.S. electricity users almost $14 billion annually.

SCHEDULING AND DISPATCH

Although scheduling and dispatch are two separate services, we lump them together because they are inexpensive and both are performed, or at least coordinated, by the ISO. Scheduling is the before-the-fact assignment of generation and transmission resources to meet anticipated loads. Because the ISO has the ultimate responsibility to maintain reliability within a control area, the ISO must coordinate the schedules. Scheduling can encompass different time periods: a week ahead (e.g., a utility will schedule its units on Thursday for each hour of the following week), a day ahead, and a few minutes before each hour.

Dispatch is the real-time control of all generation and transmission resources that are currently online and available to meet load and to maintain reliability within the control area. Dispatch can include decisions on which generating units to operate at what levels to minimize fuel and variable operating costs, but such least-cost dispatch is not necessary. That is, buyers and sellers, acting through bilateral contracts, can decide which units to operate at what levels. However, the system operator must have control of enough generation and transmission resources to minimize equipment damage and service interruptions and to redispatch generating units because of transmission constraints. The ISO needs information from generators and customers concerning the value of transactions to economically redispatch.

Scheduling and dispatch are very inexpensive, requiring only computers, metering, and communications equipment plus the control-room operators. Overall, this service costs less than 0.2

Table 1. Proposed set of electric generation and transmission ancillary services

Service	Unbundle to Suppliers	Unbundle to Customers	Controlled by ISO?	Can resource be provided competitively?	Must be inside control area?
Scheduling and dispatch	Y	Y	Y	N	Y
Generating reserves					
- Load following	Y	Y	Y[a]	Y[a]	N[a]
- Reliability	Y	Y	Y[a]	Y[a]	N[a]
- Supplemental operating	Y	Y	Y	Y	N
Energy imbalance	Y	Y	Y[a]	Y	N
Real-power loss replacement	Y	Y	Y	Y	N
Voltage control					
- Generation	Y	N	Y	?[b]	Y
- Transmission	N	N	Y	N	Y

[a] If dynamic scheduling is feasible, these services can be provided and controlled by another supplier.
[b] Whether the market for generator VARs is competitive depends on the specifics of each situation.

mills/kWh. Because only the system operator can perform these services, it cannot be provided competitively.

LOAD-FOLLOWING RESERVE

The various definitions of generation-reserve services that exist today are generally based on control-area concepts. These concepts start with the basic principle that loads and resources will maintain an instantaneous balance. In addition, frequency will be maintained close to 60 Hz. Each control area in an interconnected system will maintain enough generating capacity online to provide for the area's loads, including the provision for contingencies, and to help maintain constant frequency (i.e., NERC's A1 and A2 criteria).

The NERC control-area performance criteria require control areas to maintain their Area Control Error (ACE) within tight limits. The first of the two criteria requires that, on an instantaneous power basis, the control area be in balance with the rest of the interconnection at least once every 10 minutes. The second criterion requires that the control area's energy imbalance be within a certain limit (roughly 0.2 to 0.4% of peak demand) every 10 minutes. Accumulated A2 discrepancies are called inadvertent interchange.

These criteria and their underlying concepts have given rise to various generation-related ancillary services, including frequency control, regulation, load following, energy imbalance, spinning reserve, supplemental reserve, nonoperating reserve, and standby service. Unfortunately, the definitions of, and boundaries among, these services are often unclear. The primary reasons for this lack of clarity, we believe, are (1) ambiguity about what can be purchased from another supplier and (2) differences between a control area and an individual customer's load. In addition, the evolution of the relationships among utilities with each other and with the regional reliability council also complicates clear definitions for these services.

Figure 1 shows the load for a hypothetical customer from 7 am to 8 am on a weekday morning. The total load consists of three primary components. The first element is the minimum constant (base) load during the hour, 80 MW in this example. The second element is the trend during the hour (the morning pick-up in this case); here that element increases monotonically from 0 MW at 7 am to 14 MW at 8 am. The third element is the random fluctuations in load around the underlying trend; here the fluctuations range over ±2 MW. Combined, the three elements yield a range of loads during this hour of 78 MW to 96MW, with a mean of 85 MW.

Fig. 1. Components of a hypothetical customer's time-varying load, from 7 am to 8 am on a weekday morning.

In addition to the three elements discussed above, uncertainty about loads further complicates the decision on how to meet load. All three elements identified in the preceding paragraph are likely to vary from day to day with changes in weather, employment, and other factors unique to the particular customer. For example, a sudden change in wind direction might drive clouds over the service area, in response to which loads will increase as customers turn on lights. Under the traditional industry structure, the local utility must have available additional generation capacity to respond to these unanticipated changes in load. In the emerging structure, the customer may need to purchase additional services, either from its control area or from another supplier.

Given these different components of a customer's load, what can that customer reasonably purchase from a supplier? The customer can easily purchase a block of power consistent with its base demand. Even here, however, ramping requirements at the start and end of the hour complicate the situation. Can a remote supplier provide the second element, ramping? Not under today's protocols for scheduling transactions. Can a remote supplier provide the third element, random fluctuations? Again, not under today's protocols. A remote supplier could fully meet this customer's time-varying load only if that load was telemetered to the supplier's control area and to the supplier's generating units, as well as to the customer's control area. Under these conditions, called dynamically scheduling, the load would effectively be removed from the customer's control area and placed in the control area of the supplier. Dynamic scheduling is widely accepted in principle, but its implementation is still rare.

Given this parsing of a customer's load, it is clear that the various ancillary services discussed today do not match the elements of the load. For example, is load following the same as element 3 or is it equal to the sum of elements 2 and 3? How does regulation differ from load-following? Where does energy imbalance fit in?

The generating units that provide ramping (element 2) are chosen because they can respond to controls and will fit into optimal (i.e., least-cost) dispatch when they are loaded. The units that follow fluctuations (element 3) need to respond more rapidly to control signals (in terms of MW/minute) and, because they oscillate throughout the day, may not need to fit into the least-cost dispatch. Both types of generating unit respond to unscheduled changes in load.

We believe that elements 2 and 3 (ramping and fluctuations) could be combined into a single service. One could define the base component of demand as the lowest likely level of demand during a particular time period (80 MW in our example) and have the service that meets elements 2 and 3 provide both capacity and energy.

The volatility of arc-furnace loads provides a vivid example of the complications in defining load following, as well as energy imbalance (Fig. 2). Although this steel-mill load averages 38 MW, its load varies from 9 MW to 76 MW during this hour.

This discussion suggests that some of the generation-related ancillary services can be combined into one customer load-following service (the sum of elements 2 and 3 in Fig. 1). If this service accurately follows the time-varying loads of customers, the control-area requirements traditionally met by frequency control and tie-line regulation will be automatically satisfied, except for outages and losses. Also, by definition, energy imbalance will be zero under these conditions. Operating reserves will still be needed to protect against generator and transmission contingencies.

Fig. 2. Minute-by-minute variation in the load of an electric-arc steel mill.

Thus, load-following reserve, in our view, includes four separate components. The two control-area functions included in load following are maintenance of interconnection frequency at 60 Hz and maintenance of generation/load balance within the control area. The two customer functions include following the moment-to-moment fluctuations in load and following the longer-term (e.g., hourly) changes in load.

The output of the generating units used to provide load-following reserve is adjusted continuously and automatically to compensate for changes in aggregated customer load. These generating units have governors, which automatically adjust unit output in response to frequency changes. The units also have automatic generation control (AGC) equipment, which responds to signals from the system operator's computer to change output in response to changes in ACE. Typically, utilities assign about 1% of their generating capacity to load following.

Load following includes the fixed costs of the generating units used to follow load plus the variable costs associated with increased O&M, higher heat rates, and the requirement to operate these units out of merit order. Because these units are constantly increasing or decreasing output, their operating lifetimes may be shorter than if they were operated at a more nearly constant output; this lifetime loss adds to the fixed cost for this service. Finally, some of the capital costs of governors and automatic generation control equipment should be assigned to load following. Overall, load following costs about 0.5 mills/kWh.

In principle, customers should be charged for load following on the basis of the volatility of their loads (e.g., the standard deviation of the load). An interesting question concerns the appropriate time period over which to determine such temporal variations. The usual 30- or 60-minute interval may be too long to capture the effects of load volatility on utility costs. As shown in Fig. 2, some customers impose substantial load-following costs on the utility, while other customers (such as paper mills) may impose near constant loads, which require little of this load-following service.

Absent dynamic scheduling, only the system operator can provide the real-time signal to increase or decrease generator output. Any generating unit located within or close to the control area can provide the service. Of course, those units must be equipped with governors and automatic-generation control equipment.

OPERATING RESERVES

Operating reserves are, in some respects the supply side analogue of load-following reserve. While load following reserve is used to match generation to load based on the time-varying nature of demand, operating reserves balance generation to load in response to unexpected generation or transmission outages. Generating reserves used to meet generating and transmission outages are split into two pieces:

Reliability reserves, which include spinning reserves and other generating units that can be started quickly, all of which must be fully available within 10 minutes, typically about 3% of peak demand; and

Supplemental-operating reserves, which include generating units that can begin to provide power within 10 minutes and are fully available within 30 minutes, typically about 3% of peak demand.

These reserves are controlled in the same way as load-following reserve. Both detect and respond to discrepancies between generation and load. An important difference is that load-following spinning reserve is responding all the time to small changes in system load while operating reserves respond to infrequent, but usually larger, failures of generation or transmission. We split these reserves into their two component parts because they differ in the types of equipment used to provide the service, the number of potential providers (including interruptible loads for supplemental-operating reserve), the extent to which they must be controlled by the system operator, and the cost.

The costs of reliability and supplemental operating reserves include both fixed and variable components. The fixed-cost component is the annualized cost of the generating units plus control equipment used to provide these reserves. When the reserves are called upon (i.e., to respond to generation or transmission outages), there will be additional fuel costs incurred. Some utilities impose both a fixed cost (in $/kW-month) and a variable cost (in ¢/kWh, when these reserves are used) for operating reserves. Overall, these reserves cost about 1.8 mill/kWh, with more than half the cost from reliability reserve.

Spinning reserve is spread over as many units as is practical because it is easier to get the required rapid response by adjusting several units a small amount rather than by adjusting a single unit a large amount. Any generating unit equipped with a governor and AGC can help provide this service.

Utilities maintain additional generation reserves to cover times when the spinning reserves are insufficient. These reserves not only back up reliability reserves but are also used to restore the generating mix to a least-cost configuration. Any generating unit or interruptible load could help supply this service if it can be fully available within 10 to 30 minutes. Supplemental reserves are less expensive than reliability reserves because the former do not require governors or automatic generation control. Also, supplemental reserves are not necessarily maintained in as ready a state as are reliability reserves.

As with load following, only the system operator knows when and how much operating reserves are needed. However, any generating unit or interruptible load within or near the control area can provide the service.

ENERGY IMBALANCE

Energy imbalance (EI) is unfortunately unavoidable because it is impossible to exactly match generation to load. As defined by FERC, EI is a confusing service. In some sense, EI is the customer equivalent of a control area's inadvertent interchange. At both the customer and control-area levels, the service is intended to serve primarily as an accounting mechanism to ensure appropriate compensation (to the local control area for EI and to other control areas for inadvertent interchange) for the unavoidable small discrepancies between actual and scheduled flows.

FERC's definition of EI specifies a deadband of ±1.5%. If the deviation between actual and scheduled flows, measured over each one-hour period, is within this deadband, the customer can return the imbalance "in kind" during a like time period (onpeak or offpeak) within a 30-day period. Within this deadband, over- and under-generation can offset each other.

If the deviation falls outside the deadband, then FERC proposes to charge the customer 10¢/kWh for imbalances outside the deadband. FERC is not clear on whether these charges apply to both undergeneration (where the customer is taking unscheduled energy and power from the local control area) and overgeneration (where the customer is supplying unscheduled energy and power to the local control area). We assume that the customer would pay for undergeneration and would receive neither compensation nor penalty for overgeneration. Alternatively, the customer could be charged for deviations outside the deadband in both directions.

In our view, this definition of EI is too broad. It encompasses both an accounting service (intended to compensate the control area for minor discrepancies) as well as a penalty for substantial deviations from schedule, which are more akin to backup services. The steel mill load is a real-world example of the complications in defining energy imbalance (Fig. 2).

We suggest a three-part split of EI.

Energy imbalance would include only the discrepancies within the defined deadband as measured over the defined time interval. The deadband could be set at ±1.5%, ±3%, or some other number (perhaps based on the NERC control performance criteria L_d, which is typically ~0.5% of system peak). The appropriate time period for measuring energy imbalance could be set at the 10 minutes used in the NERC A1 and A2 criteria or the 60 minutes proposed by FERC. To prevent chronic abuse, it may help to set tight limits on the deadband and the reconciliation period and allow occasional short deviations outside the deadband (e.g., in the event of a forced outage). For example, compliance could be required for at least 95% of the time periods. On the other hand, the benefits of tight limits on energy imbalance must offset the higher costs of metering, accounting, and billing. Any imbalance outside the deadband is handled with either standby service or unauthorized use, discussed below.

Standby service would be contractually arranged beforehand between the customer and a supplier (not necessarily inside the local control area). Presumably, the provider of this service would impose both demand and energy charges for this service. See discussion below.

Unauthorized use would be the penalty charges imposed by the local control area in the event that (a) the customer's load fell outside the deadband and (b) the customer had not arranged for standby service. Because unauthorized use is not a service, its charge would not be based on costs. Rather its price would be designed to encourage customers to obtain standby service and to discourage customers from leaning on the local control area.

Assuming that customers, on average, incur an energy imbalance outside the deadband equal to 1% of their loads, this service costs about 0.7 mills/kWh. These costs cover both the capital and operating costs of the generating units that provide the service.

REAL-POWER LOSS REPLACEMENT

Real power losses are the differences between generated real power and the real power delivered to customers. Moving power always results in losses because of the resistance of each element in the T&D system. The losses depend on the network's configuration, the location and output of the generators, and the location and demand of the loads.

Losses are composed of the excitation and load losses of each element. Excitation losses are voltage dependent and essentially load independent. Load losses for most devices are a function of the square of the load. For a typical transmission system, losses average 2 to 3% of the system load. However, losses vary greatly as conditions on the network change. In particular, at times of system peak demands, losses are often much higher than under average loading conditions. The nonlinear nature and temporal variations in losses make it difficult to compute their costs and to assign them unambiguously to particular customers.

Real power losses must be made up by generators. The ISO could run its own generators to compensate for the losses, it could contract with another supplier to provide for the losses, or customers could contract with other suppliers to provide for the losses. Retail customers usually pay for losses on a system-wide basis. Point-to-point transaction customers (where a customer contracts with the system operator to move a block of power from one point to another) can either pay the system operator for the losses or they can supply extra power to make up for system losses. The ISO must have control over online generation to compensate for real-time losses even if, on average, other suppliers make up these losses.

Typically, energy losses are paid for on a ¢/kWh basis and vary with time based on the variable operating costs of generating units. Demand losses are paid for on a $/kW-month basis and reflect the costs of additional generating and transmission capacity. Only the system operator has sufficient information to know what the losses are at any time. On average, losses amount to about 1.3 mills/kWh.

VOLTAGE CONTROL

System voltage control is used to maintain voltages within prescribed limits at various points in the transmission grid and to compensate for the reactive requirements of the grid. In that sense, it is analogous to reliability spinning reserve. Local voltage regulation is a customer service, intended to: (1) meet customer reactive-power needs and (2) control each customer's impact on system voltage and system losses and ensure that power-factor problems at one customer site do not affect power quality elsewhere on the system.

We split the services into a local component and a system component because the customer has sufficient information at its location to control local reactive-power demand and the local voltage, while only the system operator has sufficient information to know what the voltage regulation and reactive-power requirements are throughout the grid. Because local voltage control is a customer problem, not a grid problem, we do not consider it an ancillary service.

Reactive losses are much higher than real losses. Voltage drops are predominantly caused by the inductance of the lines and transformers, and can be compensated for by supplying reactive power. (Too much reactive compensation can produce excessively high voltages.) Because of the high inductance of lines and transformers, reactive power does not travel well through the grid, so reactive support must be provided much closer to reactive loads than real power needs to be provided to real loads. Voltage regulation is aimed primarily at maintaining voltages within certain ranges, but is also concerned with minimizing temporal variations in voltage.

Voltage is controlled throughout the transmission system through the use of ratio-changing devices (e.g., transformer taps and voltage regulators) and reactive-power-control devices (e.g., capacitors, reactors, static-var compensators, generators, and occasionally synchronous condensers). The system operator must monitor and control these voltages and supply the reactive-power requirements of the grid. At certain locations, it may be more economical for the utility to purchase reactive support from a customer or generator than it is for the utility to directly supply the reactive support it is responsible for. The equipment used to provide or absorb VARs can be categorized as dynamic (primarily generating units) or static (primarily transmission-system equipment).

The cost of supplying reactive power is primarily the capital cost of the equipment (e.g., generators and capacitors). In addition, the operating cost of over- or under-excitation of generating units should be assigned to reactive support. The primary cost of voltage support provided by generators is the opportunity cost associated with the reduction in real-power production capability caused by production or absorption of VARs. Transmission-related voltage-control devices have both capital and operating costs.

Because the cost of system voltage support cannot easily be assigned to individual customers, its cost should probably be included in the basic transmission tariff. However, the system operator could purchase VAR support from generators as a separate service. Thus, voltage control is a service that, in our view, should be unbundled to suppliers but not to customers. Overall, voltage control costs about 0.4 mills/kWh.

CONCLUSION

In preparing its final rule on open-access transmission service, we suggest that FERC consider splitting its system-protection service into its two primary pieces, reliability reserve and supplemental-operating reserve. We also suggest that FERC define more sharply all of the ancillary services, especially load-following reserve and energy imbalance. Finally, we suggest that FERC consider other services and their provision in a restructured electricity industry; these services include black-start capability, time correction, standby service, planning reserve, redispatch, transmission services, power quality, and planning and engineering services.

References

E. Hirst and B. Kirby 1996, *Electric-Power Ancillary Services*, draft, ORNL/CON-426, Oak Ridge National Laboratory, Oak Ridge, TN, January.

Houston Lighting & Power 1995, Comments to the Federal Energy Regulatory Commission in Docket Nos. RM 95-8-000 and RM 94-7-001, Houston, TX, August.

B. Kirby, E. Hirst, and J. VanCoevering 1995, *Identification and Definition of Unbundled Electric Generation and Transmission Services*, ORNL/CON-415, Oak Ridge National Laboratory, Oak Ridge, TN, March.

Michigan Public Service Commission 1995, *Opinion and Order After Remand*, Case Nos. U-10143 and 10176, Lansing, MI, June 19.

New York Power Pool 1995, *New York Power Pool Presentation on Ancillary Services*, Prepared for FERC Dockets RM95-8-000 and RM94-7-001, Schenectady, NY, November.

North American Electric Reliability Council 1995, *Interconnected Operations Services Reference Document*, Version 1.0, Princeton, NJ, November.

U.S. Federal Energy Regulatory Commission 1995, *Promoting Wholesale Competition Through Open Access Non-discriminatory Transmission Services by Public Utilities*, Docket RM95-8-000, Washington, DC, March 29.