Demand Response

This is the first of the Council's power plans to treat demand response as a resource.[1] The experience with this resource is limited, and we have much to learn about the size of the resource, its costs and benefits, and the mechanisms available for its acquisition. This section defines the resource and describes some of the potential advantages and problems of the development of demand response.

WHAT IS DEMAND RESPONSE?

Demand response is a change in customers' demand for electricity corresponding to a change in the incremental cost of providing electricity. To understand the implications of this definition fully, it's important to appreciate some additional points:

1. Currently, demand response is weak or nonexistent, because most users of electricity have no indication of changes in costs of providing electricity. These costs vary considerably across hours of the day, days of the week and seasons of the year.
2. To achieve increased demand response will require the introduction of changed pricing and/or incentive programs.
3. Demand response as defined here does not include involuntary curtailment imposed on electricity users, but is a voluntary response by those users to price signals or program incentives, financial or otherwise.
4. The "incremental cost of providing electricity" includes the potential cost of new generation, transmission and distribution facilities if they are nearing their capacity for a specific time and/or place."
5. Demand response needs to correspond to the cost of involuntary curtailment if the power system can't meet all loads reliably.

The problem is that while the region's electricity supply is generally responsive to conditions in wholesale power markets, its electricity demand is not. This situation has a number of adverse effects. It's widely recognized as one of the factors contributing to the high and volatile electricity prices experienced on the West Coast in 2000-2001.

How did this situation arise? As described earlier, the electricity market is currently a mix of competition and regulation. Producers of electricity, who sell into the competitive wholesale market, generally see prices that reflect the marginal cost of production. These wholesale prices vary substantially from one hour to the next; hourly prices can vary by multiples of three to one or more over a day or two. When supplies are short, prices rise and producers expand supply. In the short-term, supply expands through operation of more expensive units. In the long-term, supply expands through the building of new power plants. When supplies are ample, prices moderate, and producers cut back the operation of their most expensive units and review their plans to invest in new generating units.

But most consumers of electricity see retail market prices that are set by regulatory processes. These retail prices do not follow wholesale market prices except over the long run. It may take a

[1] According to the strict legal definitions of the Northwest Power Act, demand response is probably not a "resource" but a component of "reserves." For ease of exposition, the Plan refers to demand as a resource in the sense of the general definition of the word - "a source of supply or support."

year or more for high wholesale prices to be reflected in retail consumer prices. The good news is that retail customers are buffered from the hour-to-hour and day-to-day volatility of the wholesale market. The bad news is that retail customers have little immediate incentive to respond to shortages and high wholesale prices (e.g. caused by extraordinary weather, poor hydro conditions, by temporary generating or transmission outages or even market manipulations) by reducing demand for electricity.

In the absence of such response, overall system costs are increased. More expensive generators are dispatched and eventually, when there are no additional supplies available, prices can become extremely high as load serving entities bid against one another for power. As the experience of the last couple of years has shown, higher costs to load-serving entities eventually make their way into retail rates and customers' bills. Without demand response,[2] the electricity market lacks one of the mechanisms that moderate prices in most other markets.

In the traditional world of regulated monopoly utilities, inaccurate retail market signals led to a power system that was inefficient but tolerable. Without much demand response, we probably built more generation, transmission and distribution facilities than would have been necessary otherwise. However, utilities were able to build the extra facilities, recover their costs and make returns on their investments. The lights stayed on, but average costs were higher than they needed to be. Even in that world demand response would have offered cost savings, by reducing the need for generating and distribution capacity that was used only rarely.

But in the electricity industry we have now, and many believe we will continue to have in the future, the potential benefits of demand response are even greater. We now rely on a mix of regulated and unregulated power producers to build many new generating plants. The unregulated producers have no obligation to build, and no assurance of making a return on investment. Regulated producers, too, may regard construction of a new generating plant as a risky investment because of uncertainty regarding their ability to recover costs for regulatory and other reasons. There is no guarantee that either group will find it worthwhile to build to the same reserve margins as we have enjoyed in the past.

The region needs to maintain the reliability of the system and moderate the volatility of wholesale prices, without giving up the potential benefits of a competitive wholesale market. In our current situation, demand response can reduce the overall cost of the system, and play a critical role in ensuring reliability and price stability as well.

HOW IS DEMAND RESPONSE DIFFERENT FROM CONSERVATION?

The distinction between "demand response" and "conservation." needs to be clear. "Conservation," as the Council uses the term, is improvement in efficiency that reduces electricity use while providing an unchanged level of service (e.g. a warm house in winter, cold drinks, light on the desktop). "Demand response," as the term is used here, is a change in the service (level, quality or timing) <u>that is chosen voluntarily by the consumer</u>, which reduces electricity use or shifts it to a different time. If the change in service were imposed on the

[2] In fact we have had some limited demand response mechanisms in the past. For example, in the past Bonneville had the right in contracts with the Direct Service Industries to reduce power deliveries under certain conditions. However, under current contracts this right is much more limited. The significance of this right is further diminished if DSI load declines in the long term, which seems quite possible.

consumer <u>involuntarily</u> it would be "curtailment" and it would be evidence of an inadequate or unreliable power system.

Demand response could result from rescheduling an industrial customer's production, resetting a commercial customer's heating system thermostat, or a utility's direct control of a residential customer's water heater. Demand response could also be a customer's substitution of self-generated electricity for electricity provided by the power system (e.g. the use of a backup generator for a few hours at the system's peak load).

There is an important implication of the difference between demand response and conservation. Since conservation leaves service unchanged, the costs of alternative ways of providing the service can be compared (e.g. conservation and generation) and a cost-effective level of conservation in kilowatt-hours estimated. The estimate will be somewhat uncertain because of the quality of data, but the conceptual process is straightforward -- that is, start with the cheapest conservation measures and add more measures until saving another kilowatt-hour costs as much as generating and delivering another kilowatt-hour. The total conservation measures at that point represent the cost-effective level of conservation. The Council's plans have used this level as the basis for efficiency standards and implementation targets.

But this approach can't be used to set a kilowatt-hour target for demand response. To estimate a cost-effective level of demand response in kilowatt-hours would require putting a value on the changes in service levels for the whole range of services that might be affected, which is unfeasible.[3] But it is reasonable to assume that each consumer's choice of service level is best for him given the prices he faces, and would be best for the region as well if the consumer saw the region's cost of electricity. Instead of a policy goal specified in kilowatt-hours, we can adopt a goal of identifying incentive mechanisms (e.g. prices paid or payments received) that will lead each consumer's chosen level of service to be best for the region as well. To the extent consumers see these incentives, their demand response to changing conditions will be appropriate for them and for the region as a whole.

There are a number of approaches available to develop greater demand response, each with its own advantages and disadvantages. No one of these mechanisms will be the best for every situation – it seems more likely that some combination of mechanisms will be a sensible strategy, particularly while the region is still learning about their strengths and weaknesses. At the most general level, the approaches can be categorized as price mechanisms and payments for reduced demands. This chapter examines these approaches very briefly, with more detailed examination in Appendix H.

PRICE MECHANISMS

Real-time prices

The goal of price mechanisms is the reflection of actual marginal costs of electricity production and delivery, in retail customers' *marginal* consumption decisions. One variation of such mechanisms is "real-time prices" -- prices based on the marginal cost of providing electricity for each hour. <u>This does not mean that every kilowatt-hour customers use needs to be priced at marginal cost.</u> But it does mean that consumers need to face the same costs as the power system

[3] The <u>cost</u> of a changed level of service can be calculated, but to calculate the <u>value</u> it would be necessary to see into each consumer's head.

for their *marginal* use. The "two-part" real-time prices used by Georgia Power and Duke Power provides the needed marginal cost signal without charging real-time prices for all usage. The "two-part" tariff charges customers the traditional average-cost based rate for the customer's typical usage, and applies real-time prices to deviations from the typical usage level.

Compared to payments for reductions, real-time prices offer significant advantages, including low transaction costs[4], broad reach, and a very close match of market conditions and customer incentives. Real-time prices also face significant disadvantages, including a requirement of more sophisticated metering and communication equipment than most customers[5] have now, and concern about the volatility and fairness of real-time prices. Real-time prices have not been widely adopted as yet. Because of their problems (discussed in more detail in Appendix H), the pace of future adoption may be gradual at best.

Time-of-use prices

"Time-of-use prices" -- prices that vary with time of day, day of the week or seasonally -- could be viewed as an approximation of real-time prices. Time-of-use prices are set a year or more ahead and are generally based on the expected average costs of the pricing interval (e.g. 6 a.m. to 10 a.m. and 5 p.m. to 8 p.m. winter weekdays). Time-of-use prices have many of the same metering requirements as real-time prices. Compared to real-time prices, they have the advantage of more predictable bills and they do not require the same ability to communicate constantly changing prices. On the other hand, time-of-use prices cannot communicate the effects of real-time events on the cost to the system of providing electricity. Compared to real-time prices, time-of-use prices trade reduced efficiency in price signals for greater acceptability to customers and regulators, but has nonetheless achieved only limited adoption as yet.

"Critical peak pricing" is a variant of time-of-use pricing that could be characterized as a hybrid of time-of-use and real-time pricing. This variant leaves prices at preset levels, but sets the price of a small number of hours (e.g. 1 percent or 87 hours per year) at a relatively high price (e.g. 4-5 times average price). The hours these prices apply to are not set until conditions warrant, and customers are notified 24 to 48 hours in advance. Utilities are able to match the timing of highest-price periods to the timing of shortages as they develop, providing improved incentives for demand response at times when it is most valuable.

Any of the pricing mechanisms could be offered to customers as voluntary options, or they could be mandated for classes of customers (e.g. industrial or commercial). The voluntary option has the advantage of greater acceptability to customers, but would tend to attract customers who expect their bills to go down with little or no change in their patterns of use. The mandatory option would likely stimulate greater demand response, but customers who are faced with significant changes in their patterns of use could be expected to see such pricing as burdensome.

PAYMENTS FOR REDUCTIONS

Given the obstacles to widespread adoption of pricing mechanisms, utilities have set up alternative ways to encourage load reductions when supplies are tight. These alternatives offer

[4] Compared to the status quo of average cost pricing, all of the alternatives impose some transactions costs on consumers. In the case of real-time pricing, the consumer would experience "transactions costs" in the form of time spent monitoring frequent price variations and deciding what actions to take in response. For consumers with small electricity bills, these transactions costs could outweigh the benefits of demand response.

[5] Although many large customers already have the metering equipment.

customers payments for reducing their demand for electricity. In contrast to price mechanisms, which vary the cost of electricity to customers, these offers present the customers with varying prices they can receive as “sellers.” Arrangements can vary widely in the degree of control given to the utility in exercising the demand reduction, and in the demand reduction’s required duration.

Short-term buybacks

Short-term programs are primarily directed at reducing system peak demand (e.g. by reducing loads on a hot August afternoon or a cold January morning). The total amount of electricity used may not decrease, and may even increase in some cases, but the overall cost of service is reduced mostly because of reduced investment in generators and the moderating effect on market prices. Short-term programs can be expected to be exercised and have value in most years, even when overall supplies of energy are adequate.

Utility payment for load reductions

One variant of this approach is a utility offer of compensation for short-term demand reduction (e.g. for a 4-hour period the next day), giving the customer the choice whether or not to accept the offer and reduce load. Generally the customer is not penalized for not responding to the offer, but if the customer accepts the offer there is usually a penalty if the load reduction isn’t delivered. Other variations of this approach are described in Appendix H.

Such programs require that customers have meters that can measure the usage during buyback periods. The programs also require that the utility and customer agree on a base level of electricity use from which reductions will be credited. The base level is relatively easy to set for those industrial customers whose use is usually quite constant. It’s more difficult to agree on base levels for other customers, whose “normal” use is more variable because of weather or other unpredictable influences.

Demand side reserves

Another mechanism for achieving demand response is “demand side reserves,” which can be characterized as options for buybacks. The power system needs reserve resources to respond to unexpected problems (e.g. a generator outage or surge in demand) on short notice. Traditionally these resources were generating resources owned by the utility, but increasingly other parties provide reserves through contracts or an “ancillary services” market. In such cases, the reserves are paid for standing ready to run and usually receive additional payment for the energy produced if they are actually run.

The capacity to reduce load can provide much the same reserve service as the capacity to generate. The price at which the customer is willing to reduce load, and other conditions of participation (e.g. how much notice the customer requires, maximum and/or minimum periods of reduction) will vary from customer to customer. In principle, customers could offer a differing amount of reserve each day depending on their business situation.

The metering and communication equipment requirements, and the need for an agreed-upon base level of use, are essentially the same for demand side reserve participants as for short-term buyback participants. Compared to stand-alone buyback programs, demand side reserve programs may have an advantage to the extent that they can be added to an existing ancillary services market.

Payments for reductions -- interruptible contracts

Interruptible contracts give the utility the right to interrupt a customer's service under certain conditions, usually in exchange for a reduced price of electricity. Utilities have negotiated interruptible contracts with some customers for many years. An important example of these contracts was Bonneville Power Administration's arrangement with the Direct Service Industries, which allowed BPA to interrupt portions of the DSI load under various conditions.

In the past, these contracts have usually been used to improve reliability by allowing the utility to cut some loads rather than suffer the collapse of the whole system. In practice, service was rarely interrupted. Now these contracts can be seen as an available response to price conditions as well as to reliability threats. We can expect that participants and utilities will pay close attention to the frequency and conditions of interruption in future contracts, and we can imagine a utility having a range of contract terms to meet the needs of different customers.

Payments for reductions -- direct control

A particularly useful form of interruptible contract gives direct control of load to the utility. Part of BPA's historical interruption rights for DSI loads was under BPA direct control. Not all customers can afford to grant such control to the utility. Of those who can, some may only be willing to grant control over part of their loads (e.g. a specific production line, or a domestic water heater or furnace thermostat). Direct control is more valuable to the utility, however, since it can have more confidence that loads will be reduced when needed, and on shorter notice. The adoption of advanced metering and other technologies can be expected to facilitate the use of direct control.

There is an interesting and potentially very attractive form of direct control technology that could be available in the near future. This technology would need no intervention by a utility to reduce load, but would respond directly to stress on the power grid, indicated by grid frequency below standards. Recent work by Pacific Northwest National Laboratory and others has raised the possibility of low-cost controllers in millions of appliances, controllers that reduce loads temporarily in response to grid frequency. These controllers currently cost about $25 per appliance, but they are produced in large numbers the costs are likely to be reduced by 90 per cent or more. Appliances with such controllers represent a potentially very significant short-term "peaking" resource that could address spinning reserve requirements at very attractive cost.

Longer-term buybacks

Longer-term reductions in load, from buybacks or other incentives, are uncommon in most parts of the world but have been a useful option in the Pacific Northwest, given the year to year variability of hydroelectric production. Such programs, in contrast to short-term buybacks, generally result in an overall reduction of electricity use. They are appropriate when there is an overall shortage of electricity, rather than a shortage in peak generating capacity.

Most utility systems, comprising mostly thermal generating plants, hardly ever face this situation. The Pacific Northwest, however, relies on hydroelectric generating plants for about two-thirds of its electricity use. In a bad water year we can find ourselves with generating capacity adequate for our peak hours, but without enough water (fuel) to provide the total electricity needed over the whole year.

This was the situation in 2000-2001, an unusually bad supply situation for our region. The longer-term buybacks that utilities negotiated with their customers were reasonable and useful responses to the situation. Even though these longer-term buybacks might not be used often, there will be other bad water years in the future, and it's prudent to preserve long-term buybacks as an option for those years. Most of the long-term buybacks in 2000 and 2001 were with aluminum smelters. If, as seems likely, much of that capacity does not resume operation, aluminum smelters would no longer be as significant a source for long-term buybacks. However, there are some other activities that could also be sources for long-term buy-backs.

ADVANTAGES OF PRICE MECHANISMS VS. PAYMENT FOR REDUCTIONS

Generally, buybacks avoid some of the problems of price mechanisms, and they have been successful in achieving significant demand response. Utilities have been able to identify and reach contract agreements with many candidates who have the necessary metering and communication capability. The notification, bidding and confirmation processes have worked. Utilities have achieved short-term load reductions of over 200 megawatts. Longer-term reductions of up to 1,500 megawatts were achieved in 2001 when the focus changed from short-term capacity shortages to longer-term energy shortages because of poor water conditions.

But buybacks have limitations relative to price mechanisms, even though the marginal incentives for customers to reduce load should be equivalent in principle. Buybacks generally impose transaction costs by requiring agreement on base levels of use, contracts, notification, and explicit compensation. The transaction costs mean that they tend to be offered to larger customers or easily organized groups; significant numbers of customers are left out.

Transaction costs also mean that some marginally economic opportunities will be missed. There may be times when market prices are high enough to justify some reduction in load, but not high enough to justify incurring the additional transaction cost of a buyback.

POTENTIAL BENEFITS OF DEMAND RESPONSE

The benefits of demand response depend on: 1) the cost avoided by an incremental megawatt-hour of demand response, 2) the total amount of demand response that can be achieved, and 3) the cost of achieving that amount of demand response. This section will describe approaches to estimating the first two factors. While experience with the cost of achieving demand response is beginning to accumulate, it is not yet practical to translate that experience into a "supply curve" of demand response.

Avoided cost

The cost avoided by an increment of demand reduction is the cost of generating and delivering the extra electricity that would have been needed otherwise. The avoided cost is the value of demand reduction to the power system. The system could afford to pay up to the avoided cost for demand reduction and still reduce the system's total cost.

It's important to understand that the short-run avoided cost can be substantially different than the long-run avoided cost. In the short run the power system may have adequate peak capacity, so that the cost of meeting peak load is simply operating the existing generators and using the existing transmission and distribution system to deliver the energy. In the long run, with growing demand for electricity, the cost of meeting peak also includes the construction and operation of new generating plants and perhaps the expansion of the transmission and distribution system. These extra construction costs can increase avoided cost by multiples of five to 20. This means that 80 percent to 95 percent of the value of demand response is in avoiding construction of unnecessary generators in the long run. Accordingly, this plan is concerned with long-term avoided cost[6].

The avoided cost varies widely across the hours of the year as supply and demand for electricity is affected by season, weather and other conditions. The avoided costs will be highest when demand is highest and/or supply is tightest. Estimates of these costs depend on assumptions regarding availability of imports, the degree of flexibility available in the hydroelectric system, the cost of peaking generators, and others.

Council staff has made preliminary estimates of avoided costs that are described in more detail in Appendix H. These estimates range from several hundred dollars to more than 1,000 dollars per megawatt-hour, substantially higher than the rates paid by most retail customers, which are based on average costs. Retail rates vary by utility but average about $60/megawatt-hours over the Pacific Northwest. To the extent that avoided costs and retail rates diverge, retail customers lack incentive to adjust their electricity usage appropriately, and demand response programs are worth pursuing.

Potential size of resource

Since short-term demand response affects customers differently than does long-term demand response, it is to be expected that different amounts of each will be available. Some of the limited historical experience with short-term demand response has been translated into a range of short-term price elasticities.[7] By using elasticities from the lower end of that range, modest avoided costs, and modest peak loads,[8] it was estimated that short-term demand response of at least 1,603 megawatts could be developed in the Pacific Northwest.

Any estimate of longer-term demand response must be based on the region's recent experience using buy backs to respond to the tight supply and high prices that persisted for weeks and months in 2000-2001. In that case, load reductions varied from month to month but totaled over 2,000 megawatts for significant periods. Many of these reductions came from the aluminum industry, which has unique characteristics that made it particularly attractive to reduce loads in the economic environment of 2000-2001. Similar reductions could be difficult or impossible to repeat if, as seems possible, the aluminum industry's presence in the region does not recover in the future. However, other economic activities, particularly those for which electricity is a significant part of the cost of production, may be candidates for long-term or at least seasonal demand response.

These very rough estimates of potential could be refined, although the basic conclusion to be drawn seems clear – even if they are wrong by a factor of two or three, the potential is significant.

---

6 In some cases costs of construction of distribution and/or transmission could also be avoided by demand response. These costs are location specific and are not included in these avoided cost estimates. If it were possible to include distribution and transmission in the calculations, avoided costs would be higher.

7 Price elasticity is a measure of the response of demand to price changes -- the ratio of percentage change in demand to the percentage change in price. A price elasticity of –0.1 means that a 10 percent increase in price will cause a 1 percent decrease in demand.

8 Our estimation process is described in more detail in Appendix H.

Experience

Programs to stimulate demand response are gaining experience, in our region and nationally. In our region, a number of utilities have run short-term buyback programs; Bonneville, PGE and Pacific Power have the most experience in this area. Longer-term buyback programs were run in 2000-2001 by these utilities and others, including Avista, Chelan County PUD, Grant County PUD, Idaho Power and Springfield PUD. While this region has no significant experience with real-time prices, several utilities, including Tacoma Power, Puget Sound Energy and Montana Power (now NorthWestern Energy) have offered service to customers at prices that followed the wholesale market on a daily or monthly basis. Puget Sound Energy, Portland General Electric and Pacific Power and Light have experience with pilot programs in time-of-day pricing. Milton-Freewater Light and Power has a program that allows the utility to control residential water heaters directly, and Puget Sound Energy ran a pilot program in which it directly controlled thermostats of residential heating systems. More detailed information about this experience is presented in the Appendix H.

Nationally, the best-known real-time price programs are at Duke Power, Georgia Power and Niagara Mohawk. Gulf Power has a voluntary residential time-of-day price program that incorporates a critical peak price for no more than 1 percent of all hours. Finally, there are a number of short-term buyback programs, run by utilities or independent system operators; some of the best-known are those run by PJM Interconnection, ISO New England, New York ISO and by several utilities and agencies in California.