Transmission System Operation and Interconnection

Transmission System Operation and Interconnection

Fernando Alvarado
The University of Wisconsin
Madison, Wisconsin

Shmuel Oren
University of California at Berkeley
Berkeley, California

Introduction

Stated simply, the ultimate objective of the transmission system is to deliver electric power reliably and economically from generators to loads. Power systems are large, highly complex, ever-changing structures that must respond continuously in real time. Electricity must be produced and delivered instantaneously when it is demanded by a load. Power outages are not acceptable, so the system must also tolerate sudden disruptions caused by equipment failure or weather. And the system must perform as economically as possible, with transactions and sales monitored accurately.

Another distinctive feature of the electricity system is its inherent dynamic effects, which must be considered at all times even though they are difficult to explain and fully anticipate. Dynamic effects can be illustrated if we liken the power system to a large ballroom with many chandeliers. Each chandelier (system load) is connected to one or two other chandeliers by (big) rubber bands (the transmission lines). At strategic points these rubber bands are also attached to the ceiling (these rubber bands, which support the whole structure by being attached to the ceiling, represent the generators supporting the system). The whole structure is quite precarious. Not only must it be strong enough to support the chandeliers, but the loss of any rubber band must also be tolerated. Because the loss of a rubber band will set the whole pattern of chandeliers in oscillatory motion, the system of interconnecting rubber bands (the transmission system) must be designed so that these oscillations do not become destructive and cause some or all of the whole ensemble to crash to the floor.

This paper addresses the operation of the U.S. electric power system in its evolution from a historic structure of regulated, vertically integrated, regionally franchised utilities to the present-day market in which competition and entry by new participants is encouraged. Our specific focus is the impact of this industry restructuring on system operations. Our analysis presumes that the current structure of interconnected generators and loads will not be fundamentally altered. Examples of fundamentally different structures for the delivery of electricity that are not considered here include such possibilities as the provision of electricity to customers by means of isolated, distributed generators at every customer site. Another drastic alternative that we do not consider would be the use of direct current (DC) transmission as the backbone of the entire system. The cost of this alternative for the entire grid would be prohibitive. We presume that if these more radical options are incorporated to some degree, they will be integrated into the existing, conventional alternating current (AC) generation supply and delivery scheme. New generation is presumed to be either connected or connectable to the existing grid. Likewise, we presume that control of the system will continue to require real-time coordination between production and consumption, so we do not address electricity storage (although increased use of energy storage whenever the economics justify it can be readily incorporated into the current or any future system structure). The main impediments to energy storage are the cost and the efficiency of the technology (a great deal of stored power is lost once it is reconverted back to electricity). In short, as we address power system operations and interconnection, we assume that most of the fundamental requirements for system operation cannot change although the rules for operating the system might (and most likely will) be altered.

Traditional power system operation relies on the concept of independent but coordinated functioning of multiple "control areas." A control area is a (usually contiguous) portion of the system (lines, transformers, generators, loads, and other equipment) under the supervision and control of a single operator (or group of operators at a single locations or under a single administrative structure). Control center operators maintain the system's integrity—prevent outages and insure reliable operation—by following reliability rules that every control area enforces. The rules are intended to balance supply and demand without creating overloads, congestion, or other similar problems. Operations are based not only on maintaining a balance between supply and generation but also on controlling the frequency of the system in a distributed manner. Sufficient reserves are provided throughout the system so that it can tolerate the loss of any one component at any time (the "N-1 criterion"). We do not anticipate that either requirement will change as result of restructuring.

The remaining sections of the paper address system operations as follows:

- Traditional operating policies and protocols associated with the role of operators.
- The evolution of system operations into a competitive environment by considering two models:
  - The "reliability-driven model," in which markets are permitted to operate but reliability concerns limit which transactions can take place, and, when necessary, previously approved transactions are curtailed in the name of reliability,

    and

  - The "market-driven model," in which the objective is to create a market that values reliability sufficiently and is nimble and precise enough that reliability problems are solved by market responses to price signals, which reflect system limits and thereby embody reliability rules in the prices paid to generators or paid by consumers at various times and locations.

- Additional possible directions in which system operations and the transmission grid might evolve.
- Alternative scenarios and specific recommendations.

The objective of this paper is to delineate the conditions that will permit the creation of a power system that supports and encourages competition without compromising reliability or operability. An underlying premise is that a properly designed market structure that reaches all the way to system operations (i.e., an increased use of markets for meeting operational needs) will yield higher throughput of electricity and more appropriate utilization of the transmission grid than is currently the case. Such a market should include sufficient incentives to grid operators to maximize their throughput not only in real time but also through greater use of existing assets, e.g., by optimizing maintenance schedules, increasing live maintenance, maintaining appropriate inventories of spare parts and components, and using dynamic line ratings to maximize grid utilization. In the long term, this market structure should also encourage appropriate transmission grid expansion.

Traditional System Operation Policies and Protocols

The primary, traditional objective of power system operation is to maintain system integrity. This means that uncontrolled cascading outages must be prevented. Maintenance of system integrity is referred to as "security." Closely related to the notion of security is the notion of "reliability." Although we use the terms interchangeably in this paper, reliability is generally understood to include the concept of adequacy of supply, which means that methods for procuring reliability can be devised. At bottom, however, both terms refer to the avoidance of unintended blackouts.

The system is expected to maintain its integrity and continue to operate properly without a major disruption even when a component fails. For example, if an overhead line fails because of a lightning strike, the resulting fault requires that the line be taken out of service immediately to prevent a further expansion of the problem or damage to system components. The protective relaying system is designed to accomplish this automatically and more or less instantaneously. The overall power system should, however, continue to operate even with this line out of service and in spite of the transient disruption caused by the fault. Likewise, even if the largest generating unit were to suddenly go out of service for any reason, the system should be able to recover and continue normal operation. Normal operation means that (1) the frequency of the system stays within acceptable bounds, (2) all voltages at all locations are within required ranges, (3) no component is overloaded beyond its appropriate rating, and (4) no load is involuntarily disconnected. The North American Electric Reliability Council (NERC) has, over the years, developed a number of criteria intended to assure this degree of system reliability and security. Regions such as the Western Systems Coordinating Council (WSCC) and the Electric Reliability Council of Texas (ERCOT) follow similar objectives and protocols.

A second traditional objective of power system operation is to minimize operating costs. Thus, a system operator traditionally had two roles: to assure reliability and, in effect, run a real-time market. When conditions were tight, security would take priority. Otherwise, economy of operation dictated the operator's objectives.

Tools for Managing Operations

In order to accomplish the two (sometimes conflicting) objectives of security and economy, system operators have at their disposal a number of tools to manage the system in real time. These tools range from Supervisory Control And Data Acquisition (SCADA) systems that monitor and display the status of the system in real time to more sophisticated tools such as State Estimator programs. A State Estimator program gathers all available telemetry data (real-time measurements) on the system and gives a complete, real-time picture of system status. An accurate, error-free picture of what is going on in the system is an important precondition for running the system reliably.[1]

Operators also have means at their disposal for direct control of the transmission grid. These include control of switching operations (inserting or removing lines and/or transformers), shunt injections (usually reactors and capacitors inserted at buses, mainly as a means to regulate the voltage profile), and control of regulating transformers and other series- and shunt-adjustable devices. Operators can also adjust system area set points to help regulate system frequency, control flows on exports/imports to/from other systems, and maintain the Area Control Error (ACE) within specified bounds. ACE is the difference between the total power exported by a control area and the intended exports from the area, plus a component that represents the required contribution by that area to the control of system frequency. Control of interruptible load is also often within the purview of an operator. In emergencies, many operators also have control of feeders and ordinary system load.

Control Areas

Control areas are central to system operation and interconnection and have well-defined boundaries. Flows of power across control area boundaries are always metered and monitored. Although it may be possible to operate a large interconnected system functioning as one control area, the practical difficulties of doing so have been insurmountable to date. Even if the entire grid were to be integrated and operated as a single whole, it is likely that the notion of control areas would survive in some form as a practical means to attain distributed, decentralized, and redundant control. NERC is currently reviewing the notion of control areas in order to better adapt it to a competitive environment.

For both historical and practical operational reasons, every location in the system is assigned to a control area. Every control area in the system is “responsible” for balancing its generation with its load because the amount of electricity generated must equal the amount of electricity consumed, plus losses. Whenever there is insufficient generation, the entire system “slows down” (i.e., the frequency drops). The opposite occurs when there is excess generation. Because the entire interconnected system is so large, it is most practical to balance generation and load on an area-by-area basis. However, it is necessary to precisely measure how much power is being exported or imported by an area to know whether the area is balancing its generation and load. Gathering this information requires that every line or transformer connecting an area to any other area be accurately metered and monitored in real time, and all measurements be aggregated at a central location so that an accurate ACE can be monitored.

[1] Improvements in the area of metering, monitoring, and state estimation are significant steps in improving the transmission grid. Although many improvements are currently technically feasible, investment in them has not been forthcoming for many of the same reasons that investment in new transmission has been lagging.

Related to the notions of control area and ACE is the concept of uninstructed deviations. As attempts are made to adjust the ACE for each area, errors inevitably accumulate because the control actions of generators (or loads, if permitted) are in reality different from what was intended or instructed; these are uninstructed deviations. From these deviations arises the notion of “energy imbalance” as an ancillary service. Consistent errors in one direction or the other by a number of participants (along with random changes in demand) also give rise to frequency drift, which must be corrected with frequency regulation. In traditional systems, uninstructed deviations have been handled by having systems “pay back” the energy at a later corresponding period (peak or off peak). This approach has worked well over the years, but as we shall see below, it needs to be revisited.

The question of whether the system would be better off with more or with fewer control areas (or whether control areas might be replaced by a superior concept) has not yet been answered. It may be desirable to have more numerous, smaller control areas to avoid communication problems, handling of large amounts of data, and complexity that might make the system difficult for operators to understand. More control areas, however, mean greater need for coordination among them and a considerable increase in the number of monitored tie lines that must be precisely accounted for at all times. Fewer, larger control areas, however, might make the system more vulnerable to the effects of failure of one control center. One key point is clear: all control areas need to follow uniform (or at least compatible) practices for both reliability and business activities.

Generation Redispatch

In traditional power systems, an additional tool for system operators to maintain secure operation of the system is generation redispatch, in which operators send orders to generators to increase/decrease their output based on system security needs. In many systems, operators have access to tools that permit them to estimate the cost (a proxy for price) of their actions. Thus, operators generally have at least some awareness of costs. Increasingly, however, a larger portion of system generation (and also of load) is being bought/sold under merchant contracts that specify specific levels of production at any given time. This effectively eliminates adjustment of generator output as a primary tool for maintaining system security unless contracts are written to grant the system operator this type of control. One approach to returning this control to the operator is to have generators offer incremental/decremental (inc/dec) bids for their output. That is, generators indicate the price at which they are willing to increase their output by one MW (inc bids) or decrease their output by one MW (dec bids), with limits for both. This arrangement permits the operator to reschedule generation as before, at an explicit price. If an inc bid is exercised (a generator is asked to increase its output), the price of increasing the generator output by one MW becomes the marginal price of electricity at the generator location (this cost is also known as the locational marginal price, or LMP).

Depending on the design of the market, the costs of redispatch can be either absorbed as part of the cost of system operations and paid by all participants using a cost structure in which these expenses are shared, or these costs can be charged to those “responsible” for the need to incur redispatch costs. It is more efficient to avoid the sharing of expenses by all because this approach tends to create incorrect incentives, although for practical reasons and in cases of “common good” facilities, some sharing of costs by all is sometimes necessary. An example of this type of situation would occur when a badly located generator for the conditions may elect to produce power because it knows it will be paid the system price (which is high), thus helping create congestion and preventing other more valuable trades from taking place. The operator is then forced to incur a redispatch cost to eliminate the congestion, but since all share on the resulting added cost, the party most responsible for the congestion benefits.

Security

System security is achieved by making system operation tolerant of the outage of any component (some multiple outages are also considered). That is, the outage of any single system component (or predefined set of components) should not cause a cascading outage of equipment that leads to a total or partial blackout. The system should be secure even when an outage is the result of a "shock" such as a short circuit or fault on a component prior to the component's outage. A system that is resistant to the outage of any one component is said to be N-1 secure. In a planning time frame, N-1 security means that the intact system must be able to tolerate the outage of a component. In a planning timeframe, some allowance is often made for limitations that the system will encounter in real time. One way in which this is sometimes done is by considering the simultaneous failure of any one line and any one generator when doing planning time frame studies. In an operations time frame, however, N-1 security means that the current system must be able to tolerate the "next worst" contingency. Because an actual operating system may have already sustained the outage of one or two components, this is tantamount to operating the system in an N-2 or N-3 condition from the planning point of view. Previous contingencies are "sunk events" from the perspective of system operations. This means that, once a contingency occurs, meeting the N-1 criterion means considering the altered system, not the original system, as the new base case to which the criterion must be applied.

It is almost universally accepted that N-1 security is fundamental to system operation and that achieving this level of security is in roughly the same category as making sure that generation meets load: it must be done, regardless of cost. However, once the goal is to make the system N-2 or N-3 secure, cost and other similar considerations enter the picture. Operators have traditionally handled the threat of multiple contingencies adaptively. For example, operators have been known to "move" generation closer to loads when storms approach and the likelihood of an outage (or multiple outages) increases. "Moving" generation means increasing generation at a location near the load and reducing the output of generators far from the load (these actions must be taken together because balance between generation and load must be maintained). Because of losses in the system depend on the pattern of flows in the transmission system, and changes in losses also depend on transmission system status, an increase in load by 1 MW may require more or less than 1 MW to attain a new system equilibrium. By moving generation around under stormy weather conditions, operators are, in effect, treating the weather as a contingency. Formalizing criteria for taking such measures is not always easy, but efforts are under way to do so. In a traditional environment, the costs of such redispatch are borne by all, but in a competitive environment these costs will be differentiated by time and location and borne in accordance with the marginal price of electricity at any point in space and time. That is, every node in the system has a possibly unique marginal locational price for electricity (an LMP) which, in theory, reflects the cheapest way to deliver one additional MW of electricity to the location in question without exacerbating problems on any line or other limits.

To maintain N-1 (or better) security and achieve a secure operating point that is resistant to cascading failures requires several preconditions:

- The system must have sufficient spinning reserves. Spinning reserves are generators that can instantaneously increase their output when a decrease in frequency signals that load is exceeding generation. If there are sufficient spinning reserves, system frequency will, after the loss of the largest generator, automatically settle to a new, acceptable value as a result because a sufficient number of other on-line generators will immediately pick up the deficiency. Generators already at their limit plus other generators that do not have Automatic Generation Control (AGC) cannot be counted on to provide spinning reserves. (The outage of a component may be caused by a fault, which may pose additional problems of a dynamic nature.) There is no fundamental reason why demand (load) could not provide spinning reserve by reducing consumption in response to a frequency drop. Traditionally, demand (particularly induction motor loads) provides spinning reserves by reducing consumption as the frequency drops. However, this automatic reduction does not take place in most new adjustable-speed drives (ASDs), which are electronic motor controls that adjust motor speed independent of system frequency (unless programmed to do otherwise). The increased penetration of ASD loads is reducing the "free" spinning reserves that loads have traditionally provided to the system.

- The system must have sufficient supplemental reserves to maintain system integrity after the initial shock of a contingency. As the result of an initial outage, some components may end up operating beyond their sustainable capability or may offer the system only limited-term assistance. In either case, it is necessary to restore operating conditions that are sustainable more or less indefinitely so that the system is ready to sustain a further outage without a major cascading failure. In effect, the objective of supplemental reserves is to re-establish spinning reserve margins.

- Both types of reserves must be located so that they can deliver power as needed for every possible outage condition. While spinning reserve is being relied upon, this delivery takes place more or less automatically. When supplemental reserves are needed, it must be possible to rapidly maneuver the power system to a condition in which it is capable of delivering the needed power. Such power readjustment must be possible after every event.

There is likely to be a tradeoff between the location of reserves and the strength of the transmission system. If the system's transmission capability is very limited (or fully utilized), reserves for possible contingencies must be provided "locally" so that transmission is not necessary to access the reserves. If the transmission system has ample capacity available, the use of remote reserves is practical. For radial situations, this assessment is quite simple. In contrast to a network, a radial system has no loops and has only one way of sending power via the transmission grid; if that one link fails, the system is disconnected. The only options after the loss of the link are to generate power locally or to reduce load. In network situations with many possible contingencies and changing flow patterns, the assessment of adequacy of reserves in relationship to the availability of transmission capacity can be a difficult problem. If a control area relies on remote reserves, the transmission system must be able to deliver the reserve power when and if required. Although the distinction between radial and network situations is complex, proper handling of the complexities and subtleties of networks is the key to proper operation and design of power transmission grids.

A comment pertaining to reserves (as well as other “ancillary services”) is that reserves and energy markets are in a sense complementary in grid operations but are substitutes in energy markets. As power markets mature, the market structure may evolve to encompass more open-market instruments, such as forward trading (this refers to the purchase of power ahead of the time where it is wanted rather than reliance on the spot or real time market for electricity). Another structure that can help with assuring adequacy of reserves that might evolve is the self-provision of reserves, where anyone in need of reserve services is responsible for providing it themselves.

Traditionally, operators have relied on experience, training, and prior off-line studies that specify parameters (often in the form of diagrams or nomograms) that indicate whether the current condition of a system is acceptable from a security point of view. Today, more and more operators also rely on sophisticated power-flow and contingency-analysis software that can assess system conditions in real or near-real time. Nomograms continue to serve a useful purpose to account for real-time incorporation of stability limits. If system conditions are determined not to be acceptable, the operator generally has at his or her disposal the tools mentioned above to help address the problem.

Balancing generation and load

A system operator is in charge of frequency regulation in addition to system security. Electric power customers expect that power will be a sinusoidal AC voltage waveform of 60 Hertz.[2] Many system and end user components rely on this frequency. However, without some control of frequency (frequency regulation), the frequency would quickly drift outside acceptable bounds as a result of even slight imbalances between generation and load. Responsibility for frequency regulation is almost universally organized based on control areas. As mentioned above, every formally defined control area must match its generation to its load. NERC rules (CPS1 and CPS2) specify exactly what is meant by proper balancing of load and generation. These two rules recognize the random nature of system variations and require the balancing of production and demand over designated intervals rather than at precisely all times (which would be virtually impossible). If a control area has “undergenerated” over a short period, it is expected to “overgenerate” to compensate for the shortage. This is accomplished by adjusting the ACE set point to increase the output of generators within the control area. In other words, areas adjust their generator outputs if necessary in order to balance power.

In addition to balancing load and generation, control areas handle transactions. Transactions are scheduled when the importing area schedules a net import and the exporting area schedules a net export. Even when private parties from different areas engage in transactions, the ACE must be adjusted. In a traditional operating environment, errors in energy exports or imports (“inadvertent errors”) can accumulate. According to traditional rules, energy errors must be “paid back” in a later corresponding period (peak or off-peak). This rule implicitly assumes that (1) the accumulation of error is small and unintended, and (2) all peak-period energy has approximately equal worth as do all off-peak-period energy.

No matter how good energy-balancing rules are, some frequency “drift” can develop because of sluggish response of frequency regulation equipment, slight metering errors, and random factors; as a result, another

[2] This voltage is expected to oscillate between a maximum positive value and a maximum negative value 60 times per second, in a “smooth” manner following the shape of a mathematical sine wave. Departures from this waveform are called “harmonics.” Departures in the frequency of oscillation are called “frequency deviations.”

role of the ACE has been to bias generation to regulate frequency system wide. The definition of ACE incorporates a term that becomes negative when the frequency is above its set point and positive when it is below its set point. An additional correction can be made to the ACE in order to maintain the exact number of cycles over every time period (time correction or isochronous control). This time correction control is done by temporarily adjusting the set-point frequency.

Operating the System Economically

In addition to maintaining security and regulating frequency, a traditional system operator was also in charge of system economy. This meant that the operator sought a generator dispatch pattern that was not only secure but was as economic to operate as possible under the security criterion in effect (for example, the N-1 criterion). With a given mix of on-line resources, no constraints, and no losses, the optimum operating point is known to be the point at which the marginal cost of production is the same for every generator. When transmission losses are taken into consideration, penalty factors or other schemes can be used to determine the optimum operating point by adjusting the marginal cost of production according to the location of each generator in the system. When constraints (or contingencies) must also be considered, the problem becomes one of constrained economic dispatch. Finally, when the operator also takes into consideration other controls available (such as tap adjustments or voltage settings), the problem becomes in effect a nonlinear constrained optimization problem, better known as an Optimum Power Flow (OPF). In other words, the system operator can use generator outputs alone, or can use additional means of control in order to make the system work better. All traditional systems use some form of economic dispatch.

Constraints increase operating costs. Thus, the existence of a transmission constraint *in a traditional power system operated to minimize cost* causes a higher operating cost than if the constraint were removed. To the extent that eliminating a transmission constraint permits a more efficient operating point, greater transmission capacity lowers operating costs. If the constraint is in the flow on a line or transformer, it does not follow that merely adding a line or otherwise increasing the capacity around the constraint will always result in lower operating costs. It is possible that such actions will simply “move” the constraint to a different (and perhaps less desirable) location, with negative consequences for the system. Examples of relocating constraints by adding capacity exist in both theory and practice. An additional issue is the possibility that there may be multiple ways in which a constraint can be addressed, in whole or in part.

Tightly coupled to the problem of dispatch are the problems of operations planning and unit commitment. If an insufficient number of resources are “on line,” (that is, already running and connected to the system) it may not be possible to respond to a particular contingency without shedding load. Because security is defined as the avoidance of (cascading) involuntary load curtailment, involuntary curtailment of load is almost never acceptable. To avoid load curtailments, traditional control systems have relied on unit commitment and operations planning to decide ahead of time which units should be in service at any given moment. Determination of the optimal schedule for units to be committed during any given period generally requires the solution of a rather involved mathematical problem known as a multi-period dynamic optimization. This means that one must figure out not only the best combination of units needed to run the system at any given future time, but also how that particular set of units affects the ability to run the system at a later time, since once a unit is on line it is often desirable to keep it on line (many units are designed to have minimum “up times”). True optimization is difficult to attain, so simpler heuristics and approximations are often used to make these decisions. For example, the problem is often solved in “whole hour” intervals, it is often solved for a one-week time horizon, and many units that are “known” to be on (e.g., nuclear plants) are not included in the mix. Another approximation often used is the assumption of a “perfect load forecast” over the interval under consideration. Market alternatives that shift some of the complexity of commitment decisions away from an operator to market participants have proven successful. Such strategies include reliance on “self-commitment,” where units decide on their own when they want to be committed. The requirement that offered prices by any unit be monotonically increasing in quantity (even if its marginal cost of production decreases) is another strategy aimed at simplifying the operator’s task.

Additional problems of system operation

Some other key problems related to system operation are complex and difficult to explain. These include dynamic problems (particularly the possibility of interregional oscillations between interconnected systems as were experienced for several years in the Western Interconnection), “loop flows,” “unexplained flows,” problems of reactive power and reactive power reserves, and problems associated with flow-control devices (phase-shifting transformers, some series flexible AC transmission system (FACTS) devices, and high-voltage DC (HVDC) transmission lines[3]). Interregional oscillations arise almost naturally in relatively weakly coupled systems with long transmission lines because of interactions between the controls of the governors and the exciters of the individual generators. The prevention of such oscillations generally requires system-wide studies that represent the dynamics of generators and their feedback controls. The most effective solution to these oscillations is often to install Power System Stabilizers at all or at least the most important generators. Stabilizers would not normally be installed except in cases of extreme need because of their cost and complexity. (Although the initial cost of purchasing a generator with appropriate power system stabilizer capability is higher than the cost of a generator without this capability, the added cost at initial purchase is significantly less than the cost of retrofitting a generator to add this capability.) All of these considerations add complexity to the problem of managing system operation.

Because power system stabilizers cost money, individual generators in competition with one another would not tend to install stabilizers; there is no direct benefit to the individual generator from installing a stabilizer unless the real-time value of a stabilizer is established (under many conditions, its value will be zero, but on a few occasions the value may prove to be extremely high) or a requirement is established that all generators must install this equipment and thus share in the burden of system stabilization. Specifically, the value of a stabilizer may prove crucial only under certain highly unusual conditions that may result as a consequence of several component outages and/or unusual load or generation patterns.

[3] A FACTS device is a high-power electronic device intended to rapidly control system flows and/or voltages. FACTS technology tends to be expensive to install and maintain but can alleviate many specific AC-system problems. HVDC refers to high-voltage DC transmission, in which a rectifier is used to attain high-voltage DC power which is subsequently reinverted to AC power (possibly at a different frequency). The advantage of HVDC is that it permits greater isolation between two regions of a system than AC transmission. However, the converter stations required at each end of an HVDC line stations are expensive and subject to many technical problems. Moreover, it is difficult (some argue that it is impossible) to build large HVDC networks.

Application of the N-1 Criterion

The N-1 criterion for system operation is deterministic. It requires that the system be able to tolerate the outage of any one component without disruption and does not concern itself with the probability of an outage. Even if an outage or contingency is highly unlikely, the criterion is still generally applied because system failure when a component is lost is unacceptable. The cost of meeting this criterion is not questioned; the criterion is generally considered as fundamental as the need to balance generation and load. (In practice, some probabilistic considerations do enter into the criterion in the definition of what constitutes a credible event worth guarding against. This issue is discussed in more detail in Kirby and Hirst, *Reliability Management and Oversight,* in this volume.) The consequences of a failure to balance generation and load are immediate and measurable: system frequency drifts. However, consequences of failure to meet the N-1 criterion may not be directly observable unless a critical component goes out of service. The absence of actual contingencies to reveal the failure to meet this criterion can create the false impression that a system is operating adequately when in reality it is operating at great risk.

The application of the N-1 criterion to generation outages is illustrated in Figures 1 and 2, which also show system reserves. Figure 1 illustrates the generator outputs. Figure 2 aggregates the outputs and the unused portion of the outputs (the reserves). Figure 3 illustrates a possible situation after one generator goes out of service. Likewise, Figures 4 through 6 illustrate the N-1 criterion for line outages. Initially all lines are below their flow limits. Figure 4 shows conditions before the outage, and Figure 5 shows conditions after the outage. Figure 6 illustrates a case where insufficient N-1 capacity has been reserved.

For radial connections, the N-1 criterion may be impossible to satisfy; if there is a single radial link feeding a particular load, there is no way to prevent at least some load outage if the link fails. The only way to avoid

Figure 1. Four generating units within one area. The height of the box represents the size of the unit. The shaded area represents the portion of the unit capability that is being utilized. Availability of reserves is represented by the unshaded area. For simplicity, only one type of reserve is illustrated. In reality, reserves in several time frames are of interest.

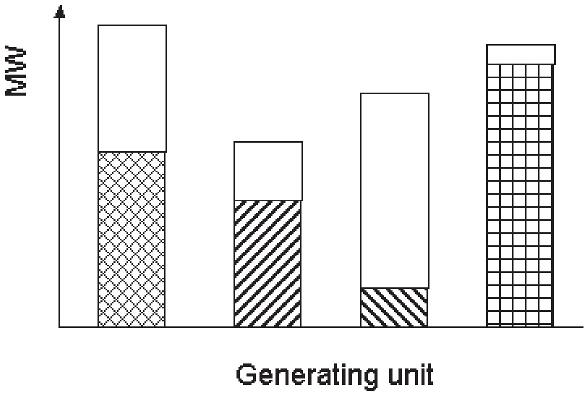


Figure 2. The vertical double-ended arrow represents the system demand (including losses, for simplicity). The stacked bars next to it illustrate how the generators are meeting the demand and the point at which total supply equals demand. Available reserves are illustrated by the rightmost bar.

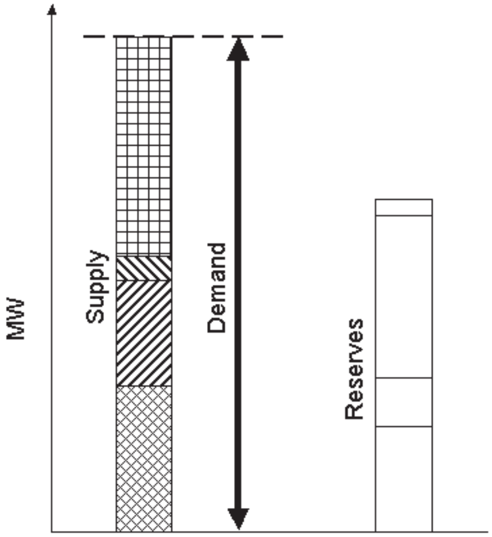


Figure 3. Outage of the largest generating unit requires reliance on reserves. In this example, we have assumed that one of the four generators goes out of service. We can see that the available reserves are enough to supply the load. The crossed-out bar suggests that the outage of the largest unit involved removal from the system of the unit itself as well as the reserves associated with the unit.

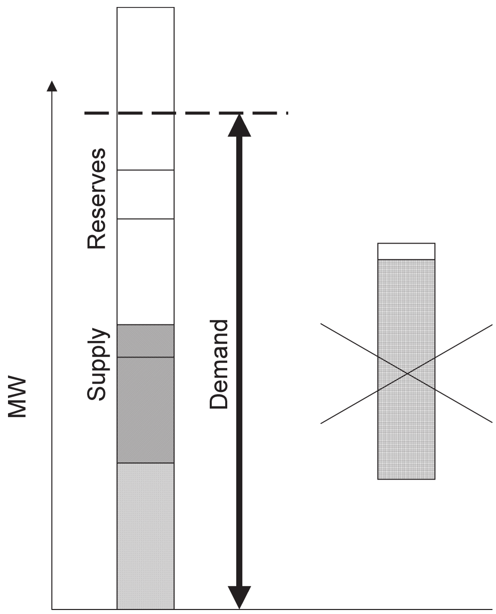


Figure 5. Flows after the outage of a line. The fourth flow in this example has been reduced by the outage. The N-1 criterion is satisfied because all new flows are still within acceptable bounds.

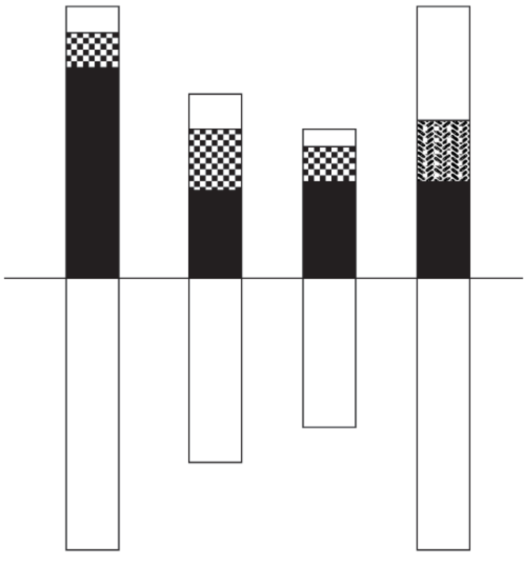


Figure 4. The operating condition flows on four potentially limiting transmission lines are illustrated. The solid portion illustrates the actual flow, and the unshaded region illustrates the capability of the line ("transmission reserves"). Flows can be bidirectional even though one limit is usually more likely than the other. In this example, the positive limits on flows are closer to being reached because all flows are in the positive direction.

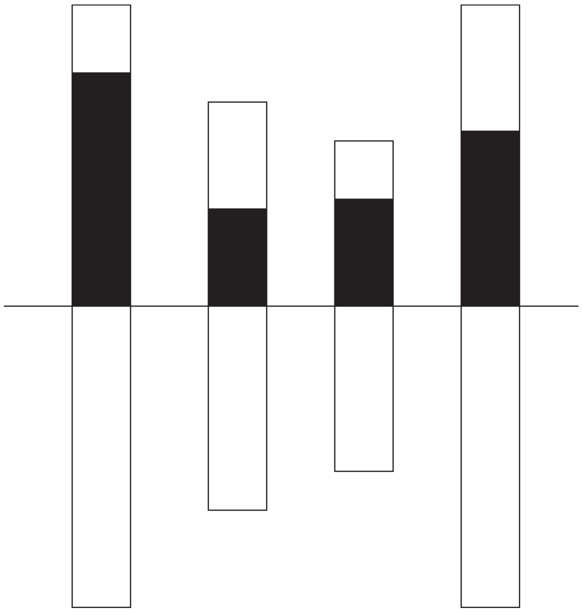


Figure 6. Flows after the outage of a generator. Under the conditions illustrated, the system does not satisfy the N-1 criterion because the outage of the generator results in an overload of the third transmission facility (the new flow is above the limit). Thus, although generation reserves may be adequate, the transmission system is unable to support the contingency flow.

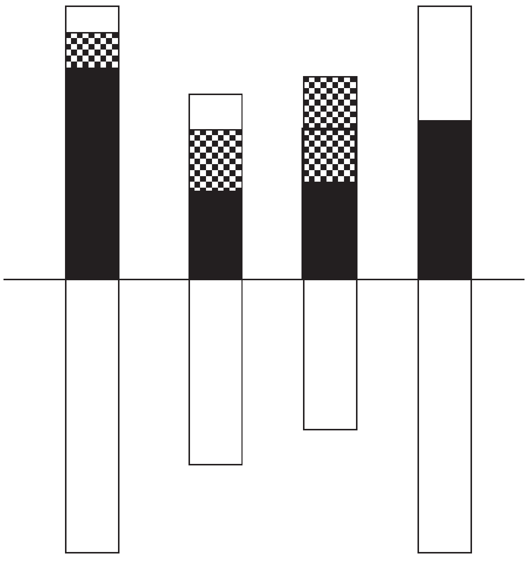


Figure 7. Disconnected (or weakly coupled) two-area system. In standard operational terminology, marginal operating costs are referred to as the system λ. The bold and unbold segments are used to designate distinct suppliers and to emphasize the lumpy nature of the supply curves. Marginal cost of production and actual conditions can give rise to significantly different marginal operating costs (and consequently significantly different marginal costs of delivering power to customers) in the two systems.

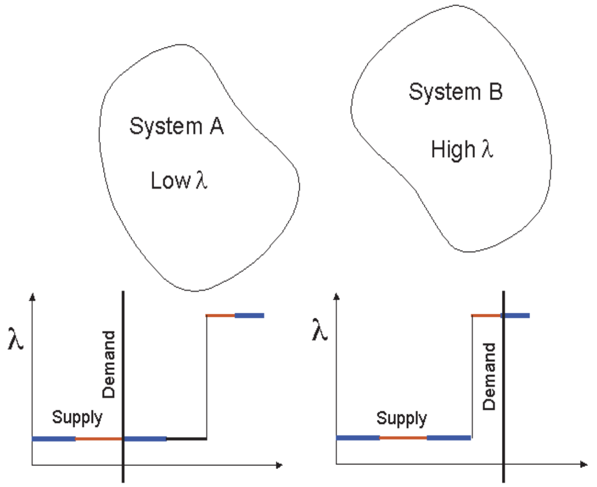
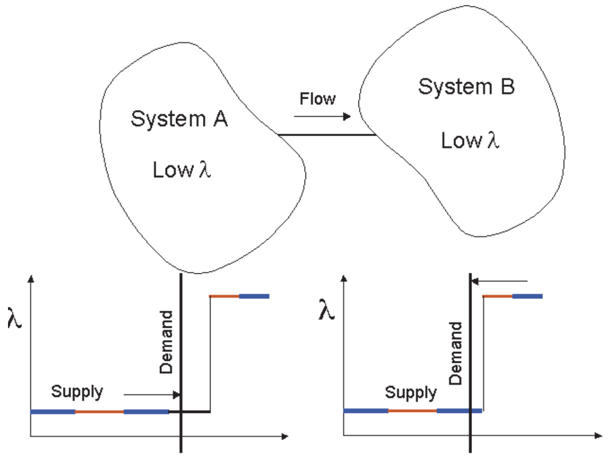


Figure 8. Construction (or expansion) of interconnection makes it possible to operate the system at lower marginal costs, resulting in cheaper production costs (that translate into end-user electricity rates) for system B without appreciably affecting the cost of system A.



an outage would be to have instantly available local generation or an energy storage unit at every load. The alternative is not to connect loads radially, which would be costly and complex.[4] The need for redundancy of transmission capability is the main reason that power systems are generally networks.

The N-1 criterion is often modified when increased security is important. In situations when it is obvious that the loss of a line or corridor is likely (for example, when a weather pattern makes it very likely that at least some lines will go out of service), the system can be operated in a more secure mode than usual, and an N-2 or -3 criterion can be used. This means that the system could sustain two or three arbitrary and simultaneous disruptions without a blackout or similar failure. These operating criteria are, not surprisingly, more expensive to satisfy than the N-1 criterion. A higher-order criterion is often used when high security is demanded or when recovery to N-1 reliability after a particular event would be particularly difficult—that is, if the N-1 event were to occur, returning the system to an N-1 secure state would be expensive or time consuming.

[4] Hospitals and other loads that are considered critical are often connected to the grid at multiple points, so they are nonradial. Even in these cases, however, instantly available dispersed generation is usually provided; most loads of this type have emergency generating units on their premises.

Interactions between Generation and Transmission for System Security

The role of generation/transmission interactions in determining a secure operating point is not always recognized in traditional operating environments. In particular, there is some substitutability between transmission and generation. For example, the system can at times make up for the outage of a generator by using another generator in the same control area, but at other times, by using a generator or generators elsewhere in the system if sufficient transmission capacity is available. The outage of any generator is initially felt everywhere in the system as a twofold impact. First, a downward frequency drift begins everywhere, and, second, the control area that contained the failed generator will start to see a large ACE. Frequency drift is stabilized initially because all generators and many loads are sensitive to frequency changes. The export (or import) from (into) a control area is by definition a regulated quantity. However, as a result of the loss of a generation and the fact that all generators participate in maintaining frequency, the immediate net effect on the area that lost the generator will be a reduction in its exports. Only after the area with the lost generator can readjust its remaining generators' output to accommodate for these reduced exports is the power that supplied was by the lost unit replaced with additional local generation (local reserves). Conversely, the loss of a transmission line will generally cause a redistribution of flows in the system. If the newly redistributed flows lead to overloads or other transmission system problems, a different generation pattern that is appropriate to the newly limited transmission conditions must be established.

Reliance on demand options during traditional system operations has generally been limited to certain interruptible load programs that meet reserve requirements when conditions are tight. Although in recent years there has been an increase in the use and creativity of voluntary customer load response programs, there remains much unexploited potential for using demand response to help manage system security. Some new trends in this area include the disclosure of real-time prices to loads, the more aggressive use of interruptible load programs in several regions, and the design of entire new ways of compensating customers for the willingness to be interrupted or curtailed. The emergence of alternatives to "all-or-nothing" power service is another possibility, e.g., allowing customers to sign up for guaranteed levels of service with anything beyond those levels subject to curtailment. Improvements in metering and metering technology will permit the system to take full advantage of voluntary load response. However, local regulatory barriers and some consumer groups' opposition to the notion that electricity should be treated as a commodity have resulted in some parts of the country being either forbidden or reluctant to adopt these demand response solutions.

System Losses

System losses must be taken into account in system operations. Most system losses are associated with series losses in the conductors; because conductors have resistance, every line, transformer, and generator loses some power as it delivers or transmits energy through them. There are also "shunt" losses in some cable systems and in overhead line arrangements, but these losses are generally far less important than series losses.

Consideration of system losses during operations is quite important for system efficiency. Average losses can account for two to five percent of total system energy, and, on the margin, losses can be considerably greater. Incremental losses in the ± six percent range are quite common, with incremental losses that exceed 10 and even 15 percent possible when there is significant reactive power flow. Ignoring losses or simplifying the accounting of them can lead to substantial economic inefficiencies as a result of some generators being chosen as “most economical” when in reality, as a result of their marginal losses at the time, they are far from economical. Likewise, desirable generators that do not increase (and in fact may reduce) losses may not be chosen because their marginal cost or bid may be slightly above some other less desirable generator. Thus, in almost all traditional system operations environments, some mechanism is used to account for losses.

Because losses vary significantly over the range of system operating capabilities, only correctly computed marginal losses should be used. A marginal loss refers to the change in losses due to the change in an injection of power at a given location. Marginal losses do not increase linearly with system demand, as some would like to assume. The best method for computing marginal losses during operations is based on the use of a transposed Jacobian matrix[5] of the system evaluated at the current operating point (Alvarado 1979). All that is required for this method is a good model of the system and a knowledge of the system’s state.

An accurate way to incorporate marginal losses (used in many systems) is the use of carefully computed penalty factors. These factors are multipliers that are applied to the marginal cost (or bid) of every generator (and in principle also to every demand) to correct for the losses attributable to the power injection or the demand. A penalty factor of 1.1 applied to a generator at one location versus a penalty factor of 1 applied to another generator means that, under the postulated system conditions, 10 percent of the power injected by the first generator will be dissipated as losses and none of the power from the second generator will be lost. A penalty factor of 0.95 would indicate that for every MW injected at a location, a reduction of 0.05 MW in losses occurs, making this location desirable. In an efficient environment, credit is due to injections with penalty factors less than 1 although credit for loss reductions is not always given in traditional system operations.

An Optimal Power Flow (OPF) properly incorporates the consideration of losses, so a separate analysis of losses is not required when one relies on an OPF.

A word of caution is in order with regard to approximation of losses. In traditional systems, marginal losses have sometimes been computed by utilities based on “B coefficients,” in which an approximate formula is established for the losses as a function of generation injections. Although such a formula can work reasonably well for some systems (particularly systems with known and established generation injection patterns), the results of such approximations can be quite wrong for systems with trading and unusual flow patterns or when conditions change. In many cases, using penalty factors obtained from B coefficients can be less accurate than using no penalty factors at all. Thus, the use of B coefficients is not recommended.

Competitive Operation: The Reliability-Driven Viewpoint

In the “reliability-driven” model of competitive market operation, trade is enabled by the posting of available capacities and requirements of the system for those engaged in commercial activities and the setting up of

[5] A matrix with the derivative of every system equation with respect to every system variable. It is used for many purposes, including efficient solution methods for the load-flow problem (Newton’s method), and for accurate determination of incremental losses and power transfer distribution factors.

some form of reservation system to allocate and approve permissible trades. The operator retains the authority to perform administrative overrides of trades (including previously approved trades) when they impair system reliability.

Reliability and Unit Commitment

In a competitive environment, the problems of unit commitment and the necessity of having an appropriate available excess (reserve) generation can be resolved in a number of ways. Because of the time lags associated with the start-up and shutdown protocols for many generating units (and also the start-up and shutdown costs for these units), an appropriate organizational structure is needed for deciding what units should be in service ("committed") at a future time. The unit commitment problem can be addressed in various ways, from "command and control" measures (also called "administrative solutions") for ensuring reliability to purely market and contractual means for ensuring sufficient reliability. From an operations viewpoint, reliability depends on the units, loads, and transmission equipment that are available in real time, so it is impossible to entirely separate the problem of operations planning and unit commitment from the problem of providing reserves. Likewise, the ability to provide sufficient reserves or reserves from certain locations is entirely dependent on the ability of the transmission system to support the transfers that would be required under contingency conditions.

In order to assure an adequate level of reliability (including a sufficient number of in-service units) several choices (and combinations choices) are possible:

- Special contracts and/or rules can be put in place to designate certain units as "must run" (or "reliability must run"), allowing the system operator to require particular units to be available for security reasons under all or certain conditions. This type of "outside the market" rule is useful for addressing potential emergencies, unusual conditions, and possibly some market power situations when the number of choices available to an operator is otherwise limited.

- In some markets (the Pennsylvania-New Jersey-Maryland Interconnection or PJM, among others), generators may if they wish bid multi-step cost curves as well as startup and shut down costs into a centralized market, allowing the market maker to "take over" commitment decisions. This means that the bidder relies on the methods and algorithms of the central dispatcher to decide when to operate. Under these conditions, however, the solution chosen can change drastically with very small changes in either the bids or the parameters of the solution method used to choose the winning bids. There is a tendency for the method to, for example, identify and choose one particular solution that strongly favors one market participant and disfavors another based on what amount to trivial differences. Furthermore, once a particular participant has been "shut out" of the commitment, it may tend to continue to be shut out even though subsequent changes in system conditions would have clearly favored the participant at the onset (Johnson et al. 1997). In the systems of the past where both units were likely owned by the same owner, this made little difference, but, in a competitive environment, it invites disputes, and in the end there is no assurance that the path taken did indeed result in the lowest possible costs because changes in system conditions subsequent to the initial decision may have favored the path that was not chosen.

- A reserves market can be created in which generators can bid units into the market that will not necessarily be used to supply energy but rather will exclusively provide reserves. This forces these units to come on line (or have the capability to come on line sufficiently fast) in order to provide the contracted for reserves if these are offered into the market.
- Generators may self-commit generating units based on their own assessment of what real-time energy prices will be. In an extreme case, there would be no reserves markets and all self-commitment would be done in anticipation of real-time prices (including the anticipation of real-time price spikes). However, system operators would understandably be quite uneasy about such a design because the slightest insufficiency in supply would require a decision to, in effect, switch to an emergency mode of operation and curtail load. Thus, it is more likely that self-commitment would be based not only on the anticipation of real-time prices but also on prices that are paid explicitly for having a given amount of reserve available (i.e., a reserves market).
- Demand can be allowed to bid into the supply market so that the operator has demand control as needed, presumably at an agreed-upon price. There are various mechanisms for compensating demand for assisting with reliability. Offering uniform cheaper tariffs at all locations for the right to disconnect a load is one way. More appropriate designs are possible, including programs where users are paid per incident and location-specific tariffs and programs are created. To the extent possible, these programs should be voluntary and based on needs of the market rather than rigidly predesigned by states or regions.
- Demand can also provide reliability services in the form of voluntary interruptible contracts. In a few cases, these contracts have an option to permit the customer to “buy out” of the interruption, something that makes the reliability service less valuable.

Of particular interest is a “combination approach” that has been working quite successfully for PJM. This approach provides for self-scheduling but then uses a centralized commitment process to meet requirements not satisfied by self-scheduled units. This strategy helps mitigate the perceived problem of potential “unfairness” of centralized commitment (mentioned above) because a generator can self-schedule if it feels the commitment results are not economically consistent with the generator’s expectations.

Administrative approaches to reliability management

The power system’s physical requirements remain the same before and after restructuring: it must still be able to survive the outage of any component, it must still balance load and generation, and it must still maintain and manage voltage profiles and frequency. In a competitive environment, however, parties are free to engage in energy transactions, that is, to purchase power for delivery elsewhere in the grid. As part of this purchase process, it is necessary to secure the transmission network “rights” that will permit the transaction.

No transaction (or combination of transactions) is allowed to violate the system’s security requirements. The “reliability-driven” viewpoint of grid operations presumes that markets are, for the most part, separate from reliability requirements. That is, it presumes that system operations do not interfere with markets until and unless there is a reliability problem, and, when a reliability problem occurs, the system operator has an overriding ability to intervene in the market with actions that include (but are not restricted to) the right to not approve, to disallow, or to terminate previously approved transactions. When this viewpoint is taken, the enforcement of system security limits (when and if required) is direct: transactions that cause the violation of a system constraint are not permitted, or, if a condition that warrants curtailment develops after a transaction has been authorized, the transaction is subsequently curtailed.

The scenario for congestion management in a competitive environment based on the reliability-driven viewpoint (and widely used within the Eastern Interconnection) goes more or less as follows:

- For every intended transaction, an Available Transfer Capability (ATC) on the transmission grid is determined. The process for determining what transmission capacity is available is based on rules created by NERC. These rules incorporate an initial determination of Total Transfer Capability, from which some "reserve transmission capacity" is deducted. Most regions within the Eastern Interconnection as well as WSCC have established and published protocols, based on NERC guidelines, by which this calculation is performed. For a comprehensive listing of these protocols, refer to www.wscc.com, to www.nerc.com, or to the various individual regional reliability council home pages. Because these rules depend on system characteristics and are administrative in nature, a process for continual updating and revision is required.

- The transfer capabilities are posted in an Open Access Same-Time Information System (OASIS) so that market participants can be aware of them. The system is updated as system conditions and scheduled transactions change. Currently, this posting of transfer capabilities is required by Federal Energy Regulatory Commission (FERC) order 889 and is available to all participants that wish to trade on the market. As described elsewhere, however, the notion of ATC without consideration of interactions among transfers and the impacts of transfer capacity on economics is of limited validity.

- Reservations for access for specific transactions are made and approved. Tariffs apply to the transmission capacity required by each transaction. Rights to transmission are reserved for designated period(s) at the desired "firmness level." The transmission tariff is not coupled to system conditions but to an established access tariff for each transaction. Transmission tariffs are generally based on revenue requirements related to the investment costs associated with transmission. Transmission tariffs for open access are generally filed with and approved by FERC based on nondiscriminatory ("just and reasonable") access principles, as required by FERC order 888.

- ATCs are updated regularly (typically at intervals of a few minutes) and re-posted in the OASIS as the scheduling of some transactions alters the ATC for other transactions. NERC requires that approvals of transactions reflect system capabilities and actual conditions. NERC does not, however, have the authority to truly enforce its requirements.

- Transactions that violate security rules are not authorized.

- During actual operation, the system is monitored for possible security violations. If security violations occur or become imminent and cannot be resolved by other means, a method for curtailment of transactions (called Transmission Loading Relief or TLR) goes into effect. The objective of TLR is to curtail (or threaten to curtail) transactions to relieve congestion and restore secure operating conditions. The determination of what transactions to curtail and by how much is based on formulas that take into consideration the size of each transaction and its relative impact on the congested flow(s). These formulas do not take into consideration the economics or the specific value of the congested transmission facility.

- As an alternative (or an extension) to TLR, anyone scheduling a transaction can also specify alternative dispatch that can be used to relieve a congestion condition in lieu of facing a TLR curtailment. This is called Market Redispatch (MRD). MRD resolves some but not all problems associated with the TLR approach to congestion management (NERC 1999).

TLR is only one example of an administrative solution to the problem of reliability assurance. Although we have chosen to focus on TLR as an example, other administrative solutions are possible and are used in other regions. However, it is a common feature of all administrative solutions that price and the actual value of transmission are not principal considerations in the process of reliability assurance, so these solutions share many of the features of the TLR process. Some other administrative alternatives are discussed below.

It is necessary to keep track not only of all transactions (this is done in NERC by means of a "tagging" system) but of the impact of every transaction on every potentially congested flow. This tracking is done in the NERC system by means of Power Transfer Distribution Factors (PTDFs), which track the unilateral impact of every system injection on every flow of interest with respect to a "reference location." PTDFs can be used to find the impact of every bilateral or multilateral transaction on every flow of interest. The tagging system is cumbersome and difficult to manage well, but some such system is necessary for administrative solutions to reliability management. The alternative to administrative reliability management that is embodied in nodal or flowgate pricing systems makes the tagging system mostly irrelevant. This alternative is discussed below.

As indicated above, TLR relies on a formula that curtails transactions based on several factors, such the impact of the transaction on the problematic flow, and the size and firmness of the transaction. The formula does not take economic factors into consideration. The formula used by for TLR curtailments is implied by the tabular curtailment procedure developed by NERC (available at www.nerc.com). This formula was made explicit in Rajaraman and Alvarado (1998).

The TLR formula and methodology are, in theory, fully capable of addressing any congestion problem in an effective manner. In practice, there are a number of problems with TLR, no matter how well implemented it is. From the reliability viewpoint, the problems of using the TLR paradigm for transmission congestion management are numerous and have been documented elsewhere (Rajaraman and Alvarado 1998). Some of the problems are:

- The TLR system can be "gamed" in a variety of ways, including overscheduling of transactions. Gaming has the potential of detracting from system reliability and does not contribute to the economic efficiency of the system.

- The curtailment process does not generally take into consideration desirable counterflows. It discourages putting together packaged multilateral transactions that would prevent congestion, simply because no credit is given for counterflows.

- The MRD method implemented by NERC allows market participants to avoid TLR transaction curtailments proposing pairs of generators that can be redispatched to relieve congestion in lieu of curtailing a trade. Even when MRD is used, it is hard to optimize system operation; when there is one congested facility, a single degree of freedom afforded by MRD may be sufficient, but when multiple facilities may congest simultaneously, a single MRD cannot resolve the problem optimally. Even for a single congested line, there is no assurance that the specific MRD offered for a given transaction will be optimal.

- Another issue with TLRs is their somewhat limited ability to control a transmission problem in view of the limited data that are used in the TLR analysis. For example, once a TLR is called, the mitigating effect it provides can be undone by the individual response of various system operators attempting to replace the energy that was being supplied by the TLR. In other words, the uncoordinated redispatch that is performed in response to a TLR can cause the same problem to repeat or a new problem to surface.

In spite of the many problems associated with TLR, it has been used with relative success to ensure secure operation of electrically interconnected but administratively and organizationally dissimilar systems. The most distinctive feature of TLR is its "command and control" flavor, which allows operators to deal with congestion in a reasonably effective manner irrespective of market considerations. Many "traditional" operators feel more comfortable with a system of command and control although many have come to realize that merely having the ability to command an action does not, by itself, ensure that the problem being addressed will be solved without creating other (perhaps worse) problems.

Another key point for any administrative solution is that, as a basic principle of efficiency, the model used by system operations ought to agree with the model used for the underlying market. TLRs or any other administrative solution violate this principle because the model used to curtail does not accurately reflect the manner in which the system operates, which creates gaming opportunities.

A well-designed and well-implemented transmission expansion plan should reduce the number and severity of TLR occurrences. There are, however, two problems with using TLRs as an indicator for assessing the adequacy of a particular transmission grid and the possible arguments for expansion of the grid:

- Care must be taken that a congestion problem is not simply "moved" to a different location. Consider the case of two transmission lines in series, 1 to 2 and 2 to 3, with capacities of 100 and 101 respectively. and no net load at location 2. Congestion can occur in the 1 to 2 segment, leading to the observation that, if congestion occurs frequently enough, a second parallel 100-MW line along the 1 to 2 corridor could alleviate the problem. More likely, however, this expansion would only move to the 2 to 3 segment the congestion that now takes place in the 1 to 2 segment of the line. The moral of this story is be careful where and how you expand the system or else you may spend a lot of money to simply move the problem somewhere else.

- The TLR system misprices the real value of congestion. For a radial system in which the value of power at one end of the line is \$30 and the value of power at the other end is \$70 and if the transmission tariff is \$20, there will be a tendency to overschedule transactions because all participants will want to capitalize on the price difference. Conversely, if the prices were only \$30 and \$40, respectively, and the cost of moving power was \$20, no transactions would take place even though there is value in scheduling transactions. The moral of this story is that before we can determine the value of a transmission expansion plan, it is important to have the prices right throughout the system. Prices can be set appropriately using a form of Locational Marginal Pricing (LMP).[6] A pure TLR system is not likely in practice to converge to optimal prices everywhere (even though this might be possible in theory).

Uninstructed deviations are another concern of traditional systems. In a reliability-driven system, the traditional "deferred payback" approach to uninstructed deviations can be implemented although it should be clear that it opens the door for intentional or unintentional abuse.

Competitive Operation: The Market-Driven Viewpoint

In the market-driven view of system operations, the operator makes relatively aggressive use of market signals and prices and uses markets as much as possible to assure reliability.

Nodal Spot Prices

In an ideal competitive market everything is, in theory, priced on the margin. In the electricity market, this principle should apply not only to generation but also to transmission. Reality is, of course, different from the ideal. To look at transmission from a market viewpoint, we begin by ignoring market power and assuming that every generator everywhere will bid its marginal cost into the market. We then proceed to define the nodal (spot) price of electricity at any system location as the cheapest way to deliver one MW of electricity to the location in question from among the available generating units while respecting all constraints and security limits in effect. If we define a nodal spot price this way, there can be no argument as to whether a different market structure could in principle lead to a cheaper set of prices: by definition, the nodal spot prices are the least expensive.

Spot prices can be attained in a variety of ways. The most direct is centralized calculation of prices by instantaneously and simultaneously "clearing" the market at all times and in all locations. This market clearing

[6] A methodology for pricing the energy at every node in the system at the cheapest possible marginal price of delivery consistent with available generation and with congestion conditions in effect. LMPs can be determined as by-products of an OPF although this is not the only way; knowledge of marginal unit locations (the location of the next cheapest generator in the system with available capacity to supply the load), congestion conditions in effect, and the characteristics of the transmission system are sufficient to determine the LMPs.

(not unlike the clearing process for markets such as the NASDAQ except that locational constraints must be respected in the transmission market) results in the establishment of (usually unique) prices for every node each time the market clears. (The technology is not yet at the point of permitting instantaneous real-time market clearing, but it is close.) Under this scheme, it is possible to define a system of property rights to the transmission system using financial contracts. The right to send power from one node to another can be acquired by buying a financial instrument, often referred to as a Financial Transmission Right (FTR) that is denominated in MW and entitles its holder to collect (or obligates its holder to pay) the difference between the prices at the two nodes. This financial right will exactly offset the deficit or gain resulting from selling the power at the injection point and repurchasing it at the withdrawal point at the respective nodal spot prices. In an idealized nodal market, in which generators have no market power and reveal their true marginal costs, an operator may have to do nothing to ensure security under most conditions.

The problem of determining nodal spot prices requires attention to a number of crucial details. First, there must be agreement by all parties about the system model that will be used to establish the prices. This model must be capable of producing unique, reproducible (auditable) results. Nonlinear models may be more "accurate," but they can lead to more than one solution, and this can be a serious problem. One way to avoid some difficulties of more accurate models is to use a simpler, slightly less accurate "linear" model. Linear models lead to reproducible prices. However, the underlying system model must take into consideration issues of voltage and reactive power flow if it is going to be credible. AC models are notorious for the fact that sometimes slight changes in the model can lead to important changes in the solution and thus to different price patterns. Thus, the best compromise is the use of linearized nonlinear models for all purposes of price determination.

For any given model the resulting nodal prices are sensitive to subjective security criteria determined by the dispatcher. When the dispatcher is also in charge of settling congestion, and paying off transmission rights whose value is determined by the nodal prices, the central calculation of prices puts the dispatcher (like PJM) in a monopoly position (Joskow and Tirole 2000). Being non-profit does not guarantee efficiency or equity and in the absence of market contestability to the dispatcher governance and monitoring of the dispatcher becomes an issue of concern.

There is always the possibility that the market will fail to clear and no set of valid resulting prices can be obtained. The only way in which this failure can be resolved is to permit much greater participation of the load. If all load is curtailable in principle, there will almost always be a valid solution for the system at zero generation and zero load. In addition, there is the issue of timing. Unless prices are determined and communicated promptly to participants, prices cannot steer the market to a reliable operating point. In short, there is no "perfect" way to establish prices. All we can hope for is reasonable approximations with good attributes and characteristics, as mentioned herein.

The nodal price patterns that result from the onset of congestion naturally create incentives to redispatch the system in a manner consistent with security. All that is needed in most cases is to produce the price signals sufficiently rapidly and then to patiently wait for the market to respond. In some cases, the market can respond based on the exercise of presubmitted inc/dec bids at individual locations. In other cases, it can respond as a result of independent action by generators observing a price signal. Only in cases where the market fails to respond because of either lack of clear price signals or insufficient available inc/dec capacity at critical locations does it become important to take "command and control" actions to direct the market. Experience with actual operation of this type of market (PJM and New York are two examples) has shown that, in spite of the current limitations on who sees the price signal (most loads do not) and in spite of "hourly averaging" effects (that is, prices tend to be for a whole hour period, even though in reality they should vary instantaneously as conditions change), markets that are purely nodal do in fact respond in a way that tends to ensure real-time security based almost entirely on price signals. An important caveat for such a system (particularly when spot prices are not correctly determined) is that, for short periods of time, prices may rise extremely high (prices reached $7,000 per MW in the midwest during the summer of 1998) because the market may fail to clear. This pricing situation prompted most systems throughout the U.S. to impose prices caps in some form or another.

Transmission Rights (Physical Rights, FTRs, FGRs)

In some cases, diversity in nodal prices can be traced to a relatively small number of constrained facilities. To the extent that the congestion status of these facilities is predictable, it is possible to directly set the price for their use at the corresponding shadow prices (the price associated with expanding the capacity of the facility by 1 MW) and define property rights associated with all potentially congested facilities, either singly or in combination, in terms of flow through these constrained facilities. Such a scheme can be implemented using a "physical rights" approach where the flows that a scheduled transaction produces through a constrained facility are determined according to the PTDFs, and the transaction must be backed by the appropriate portfolio of rights for accessing the congested facilities. In theory, it can be shown that the value of the FTRs should converge to the value of the portfolio of physical rights (also known as flowgate[7] rights, or FGRs) that are necessary to support a specific transaction under congestion conditions. A physical rights approach may require the acquisition of rights on many potentially congested paths, so a simplification is made requiring that only a predefined "commercially significant" subset of flowgates be addressed. Moreover, the exercise of physical rights requires much last-minute maneuvering to assure that rights are used and not lost, which adds to the complexities of system operation. In addition, many of the actions required by the physical rights may be in conflict with actions that the operator may wish to take in order to ensure real-time security; in other words, the physical rights approach to markets may be incompatible with the operator's need to control physical assets for security reasons.

Another approach to pricing transmission is based on flowgate rights (FGRs), in which parties acquire financial rights to specific flowgates. This approach represents a midway point between the physical rights and the FTR approaches. As in the case of physical rights, the acquisition of FGRs is based on distribution factors for flowgates that have been determined to be commercially significant. However, settlements are based on the actual marginal value of capacity on the flowgate at the time of congestion, i.e., the shadow prices on the constrained facilities. Under this setup, last-minute scheduling and operation are left to the operator, and all scheduled transactions are charged a transmission fee for the flows they induce on the congested facilities; this fee equals the corresponding shadow prices on these facilities. A transaction that is backed by the proper portfolio of FGRs will collect settlement revenues that will exactly offset its transmission fees. In practice,

[7] A flowgate is any line, transformer, or collection of lines and transformers where there is a restriction on the total power that may flow through. More generally, a flowgate is any system constraint.

however, a transaction would be covered by only a limited set of flowgate rights that only approximately track changing distribution factors, which would leave some residual congestion risk exposure unhedged. Such flexibility decouples operational decisions from the settlement issues associated with transmission rights.

Zonal Approximations

A fourth alternative for pricing in a real-time electricity market is to use zonal approximations. A zonal approximation is motivated by the intention to enable decentralized forward energy markets (that is, allowing market participants to freely trade with each other in the futures market without the need to involve transmission system considerations that require central coordination) by homogenizing the traded commodity (electric power) through deliberately ignoring transmission constraints within the zones. In cases where the only limits are on radial lines or where stability limits are translated into limits on the flow of the sum of power across several parallel lines, zonal approximations can be quite reasonable. Forward energy markets can result in infeasible schedules, so the operation of a zonal market requires that the operator have administrative tools available to force rescheduling of generators; these tools are generally presubmitted and accepted inc/dec bids for generator dispatch. In principle, if the operator has the authority to redispatch all the scheduled transactions and all generators are required to submit default inc and dec bids, then the operator retains full control capability when and if required.

The key issue is how the cost of redispatch is being covered. In some zonal markets, that cost is spread among all participants in a zone (in the form of an uplift charge levied on a load-share basis within the zone). This approach creates gaming opportunities that motivate some market participants to overschedule transactions in the zonal forward market and then be paid in the real-time adjustment market to essentially solve the congestion problem that they have created. This strategy results in a net profit when the cost of relieving the congestion is spread among all market participants. Unfortunately, such overscheduling is not only financially unfair to some participants, but it also creates serious problems for the operator who must anticipate the adjustment bids that it needs to procure. In any case, the problems attributed to the zonal approach are a result of the spreading of intrazonal congestion costs through uplift charges. To the extent that intrazonal constraints are rare and unpredictable, zonal aggregation in the forward market may have some merit in facilitating liquidity and forward energy trading. It is widely accepted that early commitments through forward trading and multi-settlements (that is, payments for forward contracts and payments for real time spot market transactions can be based on different prices that are locked in at the time that the transactions takes place) tend to mitigate market power by reducing the incentives of generators to manipulate spot prices. It is essential, however, that all transactions (forward and spot) be charged the correct ex-post congestion charges for the congestion they induce (either through a real-time nodal price mechanism or through flowgate fees on induced flows).

Can Pricing Alone Eliminate Transmission System Congestion?

A fundamental question about the relationship of markets and transmission system reliability is whether it is always possible to come up with a pricing pattern that, even within an ideal nodal pricing system, can eliminate congestion by means of pricing signals alone. The answer, partially provided in Glavitsch and Alvarado

(1998), is that this is not possible in every case. Although eliminating congestion by prices alone is possible in most cases, the lumpiness in the response capability of this type of pricing system when linear or roughly linear costs are the norm means that it is not possible to rely on price signals alone to relieve any possible congestion. Nevertheless, the system seems to work quite well in practice as evidenced by the successful real-time operation of PJM, where congestion is managed almost entirely by price signals alone. In fact, management of congestion by price signals alone can be quite effective (Ott 2000).

Under some conditions, it may be impossible to operate a system based on prices alone (particularly when demand is not exposed to or is unable to respond to the price signals). Thus, it is quite likely that a backstop would continue to be needed in such a system. The TLR approach can be viewed as a "command first, prices second" approach whereas the locational pricing approach can be better described as a "get the prices right first, then rely on command and control as a last resort." Even under the best pricing system, however, an administrative alternative (such as TLR) will continue to be necessary to ensure reliability in extreme cases.

The issue of market power is a bit more complex in this context. In a nodal pricing system it is simply not possible to study market power by observing locational price differences. A more complex analysis is necessary that takes into consideration the natural price differences that are the result of congestion and that does not ascribe these to market power. In other words, we cannot conclude *a-priori* that all price spikes are the result of the exercise of market power by some market participant. There are also potential interactions between market power in generation and ownership of financial transmission rights (FTRs or FGRs). A generator with market power in a so-called "load pocket" that imports power across congested transmission lines can profit from raising prices as long as the increased revenues exceed the lost sales. However, such a generator will have an incentive to raise prices even further if it also owns a large share of the transmission rights into the load pocket. By raising prices in the load pocket it can profit both from the sale of power in the load pocket and from the increase in the value of its transmission right that are based on the nodal price difference between the import node and export node.[8] Such a situation cannot, however, occur in the absence of physical generation ownership. Someone that owns solely financial rights or, more generally, pure financial positions, and has no control of assets in real time cannot have market power.

Other important considerations are the operational problems associated with day-ahead markets and activities. In most existing markets, there is a day-ahead market for energy. The rationale for having a multiple-settlement market (in this case day ahead and real time) is that clearing a market on a day-ahead basis is simple. Furthermore, there is an opportunity in this time frame for bidders to start and/or shut down generating units and time to arrange for transmission rights when necessary. However, when the day ahead market rules are not properly coordinated with those of the spot market participants may profit by gaming these discrepancies (for instance, by deliberately submitting incorrect schedules in the day-ahead market and deviating from them in the real time market). Such gaming may cause serious problems for the operator which must anticipate the deviations from schedules and procure sufficient reserves to maintain system reliability.

[8] This observation was made by Joskow and Tirole (2000). It has led in Texas to a restriction on the shares of transmission rights that any single market entity can own on any commercially significant constrained transmission line.

Impact of Congestion on Prices

As indicated above, in the absence of congestion (and losses), prices in a strictly nodal system are all identical. Although there may be (and in reality there always are) temporal variations in prices, there is no spatial variation in an uncongested, "lossless" market. Congestion, however, leads to differentiation of prices by location so that every node acquires, in effect, a unique and distinct price. The reason that prices are unique is simply that the PTDFs for every node with respect to a particular congested flow are unique. As an example of the price variations resulting from congestion, Figure 9 illustrates the congestion price pattern for the PJM system from 5 A.M. to 10 A.M. on October 13, 2001. For simplicity, zonal aggregate values are given rather than individual bus prices (individual prices are available from the PJM site at www.pjm.com). The purpose of this illustration is to emphasize that although they may seem somewhat "random," these prices are auditable and verifiable provided that the security criteria have been agreed upon. The prices are, however, sensitive to the implementation of the security criteria by the dispatcher. The fact that the dispatcher (PJM in this case) is nonprofit does not ensure efficiency or equity in the absence of market contestability to the dispatcher governance and monitoring structure.

Price signals give rise to behavior changes in real time. Observation at five-minute intervals of the prices and the supply at a few locations within PJM reveals the manner in which suppliers react to changing prices. The response observed depends greatly on the nature of the congestion and on the ability of the various generators on both sides of the congested facility to react in a timely manner.

Figure 9: Sample PJM locational prices for October 3, 2001, starting at 5:00 A.M. This table illustrates the rich variation in prices that is possible not only as a function of time but also as a function of location within the grid. Note: This table does not include nearly enough locations to be useful for operations and congestion management. A much larger number of locations is necessary to use pricing for efficient congestion management.

For the acronyms in this table, refer to the FERC website and report.

Region/node	5:00 A.M.	6:00 A.M.	7:00 A.M.	8:00 A.M.	9:00 A.M.	10:00 A.M.
PSEG	16.76	20.37	31.50	29.47	27.85	28.63
PECO	16.76	20.10	31.21	28.74	27.35	29.18
PPL	16.76	20.20	31.37	29.06	27.57	29.04
BGE	16.76	19.64	27.72	24.91	24.30	27.00
JCPL	16.76	20.30	31.50	29.35	27.78	28.82
PENELEC	16.76	26.58	48.13	54.91	47.36	26.90
METED	16.76	20.06	30.75	28.28	26.96	28.80
PEPCO	16.76	19.52	24.86	22.25	22.11	24.38
AECO	16.76	20.11	31.23	28.77	27.35	29.14
DPL	16.76	20.05	31.47	35.20	46.76	48.15
GPU	16.76	22.47	37.10	37.98	34.30	28.17
EASTERN HUB	16.76	20.08	31.61	36.39	50.57	51.71
WEST INT HUB	16.76	18.51	14.72	11.73	13.71	17.18
WESTERN HUB	16.76	20.67	26.34	25.63	24.59	22.33

Transmission System Expansion

Although in general transmission expansion in a pure market system will reduce congestion and improve reliability, there are quite notable exceptions. We illustrate just one: consider again a two-area system, as illustrated in Figure 10. Assume that prices are, as in most market systems, based on marginal costs of production. Under these conditions, one system (system A) sees low marginal costs and therefore low marginal prices. System B sees high marginal production costs and therefore high prices. The supply curve is, for purposes of this example, a three-tiered, steep-fronted curve. Expansion of the transmission system would lead to the possibility of trade between the systems, which would tend to equalize the prices on both sides. Assume that this equalization takes place at the "intermediate" price, as illustrated in Figure 11. The changes in consumer surplus are illustrated as the shaded areas in these two figures, which show that the decrease in consumer surplus is lower than the increase in consumer surplus for system A, leading to the seemingly counterintuitive conclusion that transmission expansion can, in fact, be counterproductive for consumers at large in a pure market situation. Although this would not normally be the situation, it is a cautionary example against assuming that transmission expansion is always beneficial. A similar example could be outlined which shows that expansion of the transmission grid can actually increase congestion.

Uninstructed deviations

Uninstructed deviations are the differences between intended or contracted amounts of energy delivery and actual deliveries. These occurs for a variety of reasons that include normal time delays in the response of units, metering and control errors as well as deliberate "price chasing" by generators who increase their out-

Figure 10: A two-area system with a proposed transmission expansion project. System A sees low prices, system B sees high prices, and a new line seems to make sense.

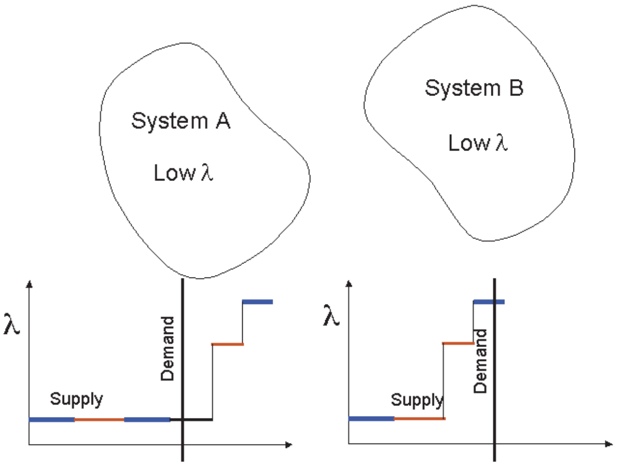
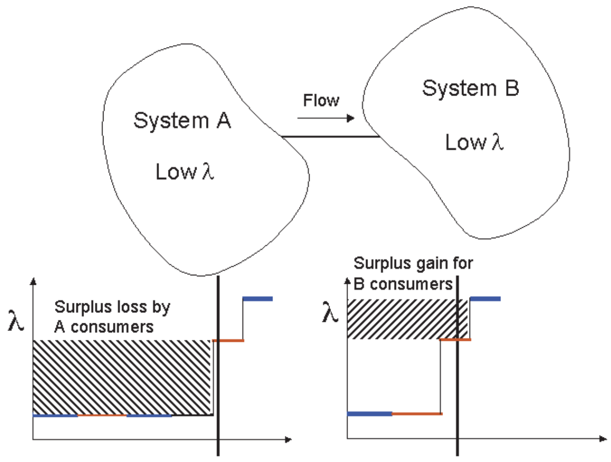


Figure 11: Construction of the line can result in a greater loss of consumer surplus than the resulting benefit to those consumers who benefit. The greater loss occurs in this case because there are fewer customers that benefit than those that see a price increase. This is not to say that surplus is not improved overall by the addition of the line, but that, in a market-pricing situation, there can be this unintended and important side-effect as viewed from the customers' vantage point.



put after learning that the price is high. In theory, in a market-driven approach to reliability and operations, uninstructed deviations can be addressed extremely simply, provided that a sufficiently robust and accurate metering system is in place. Such a system would price any uninstructed deviation at the market price for the given moment and location in which it occurs. Thus, any party having specified contracts for delivery (or consumption) of power could, in effect, hedge most of the intended amount by writing a contract with another party that specifies times and amounts for any desired price, with the understanding that settlement of any differences between contracted amounts and delivered amounts would take place on the locational real-time spot market. In effect, therefore, there would be no uninstructed deviations, only a real-time settlement market for the differences. You pay actual real time prices for any amount you have not contracted for. The reality, however, in all the currently operating markets is that real time prices are precalculated based on forecasted demand and supply bids and fixed for finite time periods (e.g., 15 minutes in Texas and 5 minutes at a subset of nodes in PJM). When the prices are announced at the beginning or prior to the time interval in which they apply (this is often justified on the ground of allowing time for load to respond), uninstructed deviations will occur and they require that the operator take mitigating actions such as dispatch of up or down regulation reserves in order to maintain frequency control.

Evolutionary Directions

The ideas and possible new directions in this section suggest what remains to be done regarding power system operations in a competitive market. In some cases, these ideas are not entirely developed, and in other cases they do not follow directly from the analysis in the previous sections.

Ensuring Reliability through Price Signals

As indicated above, one of the most effective trends in system operations has been the use of market price signals to operate the system. Using price signals and without the need for centralized controls, it is possible to induce behavior from generators (and loads) connected to the system based on the prices that the generators see at their respective locations. Instead of “commanding” that a particular generator produce more power, one can simply increase the price offered to that generator at that location at a given time. If the offer is above the marginal cost of the generator, the generator will naturally respond with an increased output. Likewise, to discourage production, the price can be lowered (to negative values if necessary) to encourage reduction in supply. Further progress in the direction of real-time node-specific pricing will go a long way toward ensuring power system reliability within the operational timeframe. To support this progress, FERC must “stay the course” of integrating the grid and unifying the rules to attain greater economic efficiency but should also recognize the need for more accurate system models whenever these models are to be used to set prices.

Ensuring Reliability by Connecting Transactions and their Flows

The second direction in which real-time reliability assurance is evolving is toward a more precise connection between transactions and flows (that is, an effort is under way to try to link or “tag” every transaction so that the flows that are “caused” by the transaction can be accounted for). The objective of this effort is to more closely focus command and control strategies on resolving specific problems. In addition, making command and control strategies more responsive to market signals (as is the case in Market Redispatch) is an important step for those cases where this strategy for security assurance is preferred.

Ensuring Reliability by Voluntary Load Response

Incorporation of load response is another important trend in reliability assurance. Load response can take many forms, ranging from alternative contractual arrangements between suppliers and load to real-time pricing at the retail level to new versions of interruptible load programs and other incentives. If an outage is seen as only a reliability problem when it is involuntary, much can be done toward improving load responsiveness. It is essential to enroll the creativity of the marketplace in the design of demand-response programs by enabling the types of customer/provider contracts that makes demand-response possible.

Ensuring Reliability through Improved Information on System Status

One of the most important preconditions for good system security is knowledge of the precise state of the system at any given time. Knowledge of system status is a precondition to establishing better flow and other types of system limits. It has become imperative to consider the evolution of a single seamless view of the system in real time. Concerns that the physical security of the system may be jeopardized because flow and other similar information is too widely available must be balanced against the security concerns that result when only partial, incomplete, and in some cases erroneous information is available about the state of the system. The creation of a single integrated view of the status of the entire interconnected grid will have great value for reliability assurance by giving those in charge a global picture of system status. More than one major blackout has been traced to operator actions based on a narrow view of the system and a focus on resolving a specific problem rather than a global view of system status and potential impact on the entire grid of the actions taken. Whether making information available to the marketplace in addition to the operators contributes or detracts from security is debatable. Recently, concerns about the physical security of the grid have suggested that publishing information about physical flows and system status represents a greater risk than would be associated with withholding such information. This view is not universally shared, however.

The complexity of determining system state increases significantly with system size. Some would argue that the growth of this complexity could be exponential, which would mean that determination of the state of the whole grid would be nearly impossible in practice. However, in the opinion of the authors (confirmed by Energy Management System experts), determining the system-wide state is possible. Nevertheless, to the best of our knowledge, no production-grade software is available to consolidate a regional SCADA system into a validated regional security model.

Ensuring Reliability through Grid Expansion and Energy Efficiency

Policies can be developed to support creation of excess capacity in the transmission grid to reduce regional market power, promote additional market activity, and increase reliability assurance. This intentional expansion of the grid by means other than reliance on the market can be encouraged either by specific policies for transmission expansion (for example, by direct or indirect government actions) or by the creation of “capacity markets" in transmission that can improve reliability margins and lead to construction of additional transmission facilities. These types of incentives for grid expansion are better explained in the companion paper, *Alternative Business Models for Transmission Investment and Operation,* by Oren, Alvarado and Gross.

Over time, energy efficiency may also be an effective tool for improving power system reliability by reducing total consumption and making any problem that occurs easier to solve.

Alternatives to Transmission System Expansion

All "outside the market" solutions (such as government- or quasi-government-sponsored expansion projects and taxes or subsidies to encourage expansion) should be considered in terms of their contribution to the "common good" of greater system security and simpler system operability. One must be careful, however, not to expand transmission in cases where expansion of generation would be more effective. This is particularly important when the best solution to physical security threats may be alternatives to transmission, such as distributed generation along with distributed fuel or energy-storage technologies.

It is also important to take into consideration the costs of fuel transportation (in recent years this refers mainly to gas pipelines) versus the cost of transmission system expansion. This tradeoff can only be taken into account accurately if the proper locational price signals are used for energy, with consideration given to transmission congestion. Only the "right" price signals will give rise to appropriate tradeoffs between the possible expansion of the transmission grid versus expansion of the fuel-supply system.

Intentional overexpansion of transmission for reliability and avoidance of market power should not be confused with the problem of free riders associated with transmission system expansion. The free rider problem (the fact that once a line is built, many parties who did not share the expense of construction will benefit from it) can lead to underexpansion of the grid. This is akin to the classic economics problem of the commons where an asset that has societal value (in this case the transmission network) will be utilized to the fullest extent by all parties, and any party investing in any improvement to the commons will be at a competitive disadvantage because it will bear the added burden of the cost of investment. This is a problem even in cases where economic expansion of the system can be justified. One of the primary purposes for the creation of Regional Transmission Organizations (RTOs) is to help resolve the free rider issue by creating mechanisms that will permit a wider view of the benefits of transmission expansion. Even prior to the creation of RTOs, the western states have adopted policies intended to take a more regional view of transmission and transmission expansion benefits. It is the role of the regulatory structure to deal effectively with the free rider problem so that otherwise economically desirable transmission system expansion takes place.

Concluding Remarks

The fact that "time and location matter" is fundamental to operations. There needs to be widespread recognition that the value of energy to an operator can have quite strong locational and temporal components associated with it. Thus, fixed tariffs do not reflect operational realities and are useless as a tool that can facilitate market-based operations.

All alternative methods for reliability assurance contemplated in this paper rely on a combination of incentives, load, and availability of resources. The resources are both transmission and generation. Reliability requires redundancy (that is, generation and transmission resources in excess of those necessary to satisfy the needs of an intact system). Providing reliability through generation alone may mean that a large amount of excess generation will be required, with a great deal of redundancy. The transmission system makes it possible to share generation resources in the provision of reliability. However, transmission redundancy is also required. A strong transmission grid is necessary for the practical sharing of reliability resources. Any expansion of the transmission grid to solve a problem may result instead in the problem moving to another location. Policies dealing with transmission system expansion must also address the lumpiness of transmission investment.

There are a number of options for transmission system expansion from the perspective of system operation and interconnection. All of these options present a possible system operations scenario followed by the related possible transmission expansion scenario.

- Option 1 (consolidated operations, market-driven government-incentivized transmission expansion): Establish a uniform criterion for system operation and interconnection, in consultation with parties who are knowledgeable about system operation. Establish uniform business practices that properly value transmission (much as FERC is doing at the moment—TLR and related approaches will not be sufficient). Although consolidation is called for, actual implementation of this model should use distinct coordinated regional control centers with protocols and policies appropriate to their regions. Base all decisions about transmission system expansion on a coordinated assessment of the anticipated benefits of expansion projects, but only after the pricing system has been given the opportunity to reflect the true costs of transmission and after generation options have been assessed. Encourage a sufficient voluntary demand response capability. Address all issues of system reliability by reliance on market forces up until the last minute, and rely on administrative solutions only under highly unusual or emergency circumstances.

- Option 2 (coordinated regional controls, government-directed transmission expansion): Operate the system much as it is operated now, retaining some form of control areas, but improve coordination. Use government incentives and resources to help create a "designed" expanded transmission infrastructure intended to relieve interregional bottlenecks. This approach must consider that, as a result, there will likely be regions where, under normal market operations, customers will see an increase in their electricity prices. Means for ensuring that proportionate benefits are derived by all parties (or at least that some parties are not adversely affected) should be part of the incentive system.

- Option 3 (government-assisted merchant transmission): Operation of the system takes place based on proper locational pricing, with single or coordinated set of distinct control centers, each with expanded demand-management options. Uniform business practices are required. The model for operations is to be based on the "market-driven" model described above in Section 4. Transmission expansion is primarily in the form of merchant lines. However, because of free-rider problems, governmental protocols are used to ensure that at least a fraction of the benefits that accrue (if any) are directed toward investors and not free riders.

- Option 3a: Same as option 3, but the model for operations is to be based on the "reliability-driven" model described above in Section 3 (e.g., a TLR model).

- Option 4: (utility-led government-assisted evolutionary alternative): Allow several operational and business models to coexist. Have the government operate reactively, approving and monitoring actions proposed by traditional industry participants and being proactive only in situations where a market power or major reliability issue becomes a concern. Otherwise, the government acts as a catalyst for action.

The authors' opinion, based on comments received as part of the DOE public-input process and the facts presented in this document, is that the most desirable method for system operation is Option 1: operate the grid as an integrated whole. Points that require further consideration are:

- Whether the grid should be organized into regions, RTOs, or some other structure, and

- Whether some form of control area should be retained.

Specific recommendations are:

- A unified business environment must be fostered for ensuring reliable operations, preferably based on the "market-driven" approach to congestion management. In order for market-driven approaches to reliability to be effective, they must be fast (i.e., operate in or close to real time) and have a sufficient number of nodes (or flowgates) available to permit correct "steering" of the system.

- Some form of administrative backstop to a purely market-driven approach must remain for extreme cases. When and how the administrative rules should "trump" the market is an issue that will require considerable additional discussion and investigation by knowledgeable parties that are sensitive to marketplace needs. In general, the answer to this question should be "only under highly unusual or emergency conditions" and not as a routine part of system operations.

- Voluntary demand management options to help achieve reliability should be expanded. The precise manner in which this can be done should not be prescribed; it should be driven by system needs and market opportunities, but the government can and should facilitate the consideration of these demand options to system operation and reliability.

- Transmission grid expansion should be based on the considered judgment of a non-partisan authority that addresses need from both the viewpoints of the operational characteristics of the system and the expanded trade opportunities the new transmission capacity will afford. Furthermore, the entity assessing expansion options must consider mitigation measures for the almost inevitable adverse impacts of increased interconnection on the customers in certain regions. Such mitigation should not render the market less efficient or more cumbersome to operate, however, but should be pursued by means of temporary financial structures (such as side-payments to adversely affected parties) that help spread the benefits of expansion.

- Study should be undertaken of drastically different transmission structures and organizations, including the possibilities of much greater use of HVDC transmission, greater system separation and islanding by means of DC converters, and active separation of the grid into separate areas. Because of their higher investment cost, these strategies should not be implemented unless studies clearly indicate their superiority for a particular situation.

- As a means of increasing operational accuracy leading to greater existing system utilization, methods should be used that permit the system to operate closer to its limits. Examples of these concepts include the use of dynamic line ratings; that is, where the flow limit of the line is not a precalculated number, but a value that depends on conditions such as the temperature of the line, its sag, wind conditions and more. Another example is the establishment and adjustment of stability limits based on actual operating conditions. This latter approach will require a continuing investment in methods and techniques for grid analysis and operation because such methods are not possible with today's state-of-the-art technology.

- Performance-based regulation (PBR) should be used to create incentives for transmission construction and efficient transmission operation and maintenance practices. Appropriate PBR methods should be developed in consultation with those who have experience with this type of regulation, both inside and outside of government. However, this topic is outside the purview of this issue paper.

- Study and enabling of a system-wide, real-time State Estimator should be undertaken, to provide information about the actual status of the system to both operators and market participants. Creation of this type of tool will require no new fundamental research but will require enlisting experts on large-scale computation and encouraging deliberate development and incorporation of new and expanded metering technology including sufficient metering redundancy, throughout the grid.

- Reserves markets should be made locational, ideally aligned with the main energy market, and nodal if necessary. This issue should be studied carefully before implementation.

- Incentives for technical engineering personnel must be compatible with the type of talent and capability that is required for grid operation and design. It is imperative that all limits and decisions relating to grid operations be compatible with sound engineering practices.

- System operators should remain independent and not be direct market participants. Nevertheless, some form of PBR should compensate and motivate system operators.

- Proper consideration of losses is essential. The correct way to handle losses is through use of penalty factors obtained from the system Jacobian matrix. Alternatively, an Optimum Power Flow can be used.

References

Alvarado, F.L. 1979. "Penalty Factors from Newton's Method." *IEEE Transactions on Power Apparatus and Systems.*

Glavitsch, H. and F.L. Alvarado. 1998. "Management of Multiple Congested Conditions in Unbundled Operation of a Power System." *IEEE Transactions on Power Systems.* August: 1013–1019.

Johnson, R.B., S.S. Oren, and A.J. Svoboda. 1997. "Equity and Efficiency of Unit Commitment in Competitive Electricity Markets." *Utilities Policy.* Vol. 6. No. 1:9–19.

Joskow, L.P. and J. Tirole. 2000. "Transmission Rights and Market Power on Electric Power Networks," *The Rand Journal of Economics,* Vol. 31, No. 3, pp. 450–485.

North American Reliability Council (NERC). 1999. "Market Redispatch." www.nerc.com

NERC. 1995. "Transmission Transfer Capability." May. www.nerc.com

NERC. 1994. "Regional Reliability Criteria." November. www.nerc.com.

NERC. 1993. "NERC 2000" (Policies for Interconnected Systems Operation and Planning) September 30. www.nerc.com

NERC. 1992. "Control Area Concepts and Obligations." www.nerc.com

NERC. 1989. "Electricity Transfers and Reliability." ww.nerc.com

NERC. 1985. "Reliability Concepts." www.nerc.com

Ott, A.L. 2000. "Can Flowgates Really Work? An Analysis of Transmission Congestion in the PJM Market from April 1, 1998–April 30, 2000." *PJM Report,* www.pjm.com

Rajaraman, R. and F.L. Alvarado. 1998. "Inefficiencies of NERC's Transmission Loading Relief Procedure." *The Electricity Journal,* Elsevier Publishers. October: 47–54.

Suggested additional reading

Alvarado, F.L. 1999. "The stability of power system markets." *IEEE Transactions on Power Systems.* Number 2, Volume 14, May: 505–511.

Chao, H. and S. Peck. 1996. "A Market Mechanism for Electric Power Transmission." *Journal of Regulatory Economics.* Volume 10, Number 1, January.

Chao, H., S.S. Oren, S.A. Smith, and R.B Wilson. 1986. "Priority Service: Unbundling the Quality Attributes of Electric Power." EPRI Report EA-4851. Palo Alto CA.

Hogan, W.W. 2000. "Flowgate Rights and Wrongs." *Harvard University Report,* August 20.

NERC. 2000. Planning Committee, “Report of the Available Transfer Capability Coordination Task Force.” July.

NERC. 1996. “Available Transfer Capability Definitions and Determination—A Framework for Determining Available Transfer Capabilities of the Interconnected Transmission Networks for a Commercially Viable Electricity Market.” June.

Tseng, C.L., S.S. Oren, A.J. Svoboda, and R.B. Johnson. 1999. “Price-Based Adaptive Spinning Reserve Requirements in Power System Scheduling.” *International Journal of Electrical Power & Energy Systems.* Volume 21, Number 2, February:137–145.

Western Systems Coordinating Council (WSCC). 2000. *Determination of Available Transfer Capability Within the Western Interconnection.*

Wood, A.J. and B.F. Wollenberg. 1996. *Power Generation, Operation and Control.* Second Edition, New York: Wiley and Sons.