The Economic Value of Improving the Reliability of Supply on a Bulk Power Transmission Network

by
Tim Mount*, Alberto Lamadrid, Surin Maneevitjit, Bob Thomas and Ray Zimmerman

Department of Applied Economics and Management and
School of Electrical and Computer Engineering
Cornell University

Abstract

The two basic criteria used by the North-American Electric Reliability Council (NERC) to measure reliability are 1) "Adequacy" (to ensure that the installed capacity of the supply system is sufficient to meet projected future loads and limit the Loss of Load Expectation (LOLE) to one day in ten years), and 2) "Operating Reliability" (to commit generating units with sufficient reserve capacity to withstand sudden disturbances such as equipment failures). Assuming that the transmission system meets the adequacy criterion, adopting a specific LOLE standard makes it feasible to determine the corresponding reserve margins for generating capacity that are needed to maintain operating reliability in real time. However, there is no established way of measuring the economic value of reliability, and as a result, regulators have adopted various proxy measures for meeting reliability standards such as specifying a required level of reserve generating capacity in a region. Although it is convenient to adopt such proxy procedures, by doing so, the underlying reasons for requiring a specific level of reliability are obscured, and more importantly, it is then virtually impossible to determine the true economic benefits of making additional investments in transmission, for example, that improve reliability. The objective of this paper is to present an analytical framework that makes it feasible to determine the economic benefits of both operating reliability and system adequacy in a mutually consistent way.

The first part of the paper demonstrates how "co-optimization" can be used to determine the correct nodal prices for an optimum AC dispatch that meets standards of operating reliability. The co-optimization criterion minimizes the expected cost of meeting load over an explicit set of credible contingencies. The corresponding nodal prices reflect the patterns of dispatch for the intact system as well as for the contingencies. In contrast, the procedures currently used by system operators typically determine an "optimum" dispatch for an intact system and comply with operating reliability indirectly by adding physical proxy constraints, such as minimum levels of reserve generating capacity. An additional distortion of typical dispatching procedures is that system operators use a DC dispatch. Since the true physical constraints on transmission are increasingly caused by limits on voltage, these non-thermal constraints are approximated by specifying "proxy limits" on the real power flows allowed on a network. Even if the resulting DC dispatch corresponds exactly to the optimum AC dispatch, the corresponding nodal prices are highly misleading, particularly when the system is stressed, but these are precisely the prices that must be determined correctly to measure the economic benefits of reliability. Getting the prices right is the main contribution of co-optimization. The second part of the paper uses the PowerWeb platform for case studies that illustrate how co-optimization can be extended to evaluate system adequacy. This is accomplished by allowing for load shedding as an expensive option to maintain the feasibility of the optimum dispatch as the system load is increased. The expected negative benefits from "energy not served" is the proposed economic metric for the reliability of supply.

* Corresponding author, tdm2@cornell.edu.

The Economic Value of Improving the Reliability of Supply on a Bulk Power Transmission Network

1. Introduction

It is clear that a major implication of the current scientific and public debate about climate change and sustainability is that our dependence on fossil fuels will and should be reduced substantially. Reductions of the global emissions of carbon dioxide into the atmosphere of at least 80% will be needed to stabilize the climate against global warming. For all intents and purposes, reaching this objective implies replacing the conventional uses of fossil fuels entirely from all but the most technically intractable applications (e.g. the military). Adapting lifestyles and production to a new low-carbon future implies a disruptive transformation of economies throughout the world that will be equivalent in importance to the first industrial revolution. Although the current administration in Washington has until recently been reluctant to address climate change effectively, all three Presidential candidates are committed to some form of new energy policy to reduce emissions of carbon. Hence, it seems likely that the nation will soon follow the lead of Europe and Japan and will initiate a transition to a low-carbon economy. The implication is that there will be many new programs that encourage the replacement of fossil fuels by renewable sources of energy such as wind and solar power. Enforcing Renewable Portfolio Standards (RPS) is an example of an existing type of program.

Most of the non-fossil sources of energy will involve the electric delivery system either directly (e.g. generation from wind) or indirectly (e.g. charging batteries for plug-in hybrid vehicles). Energy from liquid biofuels is one exception. Although the generation of electricity from fossil fuels is relatively inefficient, electricity actually delivers services very efficiently. Hence, solar power and wind power are intrinsically efficient because they avoid the losses that occur in a typical thermal power plant. In addition, using electric motors for transportation will avoid the losses due to the inherent inefficiency of internal combustion engines. Basically, one unit of electricity from a renewable source replaces three units of energy from coal in a typical power plant and five units from petroleum in a typical gasoline engine. This is the good news. We will become much more dependent on electricity for delivering energy services to customers in a low-carbon economy.

The bad news is that most renewable sources of energy are intermittent, and as a result, policies that simply focus on increasing the penetration of renewables may undermine the reliability of the electric delivery system. Major blackouts attributable to unexpected drops in wind generation have already occurred in Texas and Germany. If this type of disruption of electricity supply becomes more widespread, it is likely to lead to a backlash against renewables by system operators. Hence, policies for a successful transition to a low-carbon economy should include the following three essential components: 1) generating more electricity from solar, wind and other non-fossil sources of energy, 2) storing energy to cover periods when these sources are not available, and 3) decentralizing the delivery of electricity while maintaining the reliability of supply. Since it is practical to use small-scale production for many of the non-fossil sources of electricity, there will be many opportunities for improving the design of new buildings and communities to accommodate new energy and storage technologies. In a low-carbon economy, the current centralized sources of energy will not simply be replaced by new non-fossil sources using the existing forms of distribution. It is quite possible that most of the growth in electricity generation, for example, will come from local, distributed sources, and this change will require a disruptive transformation of the existing electric delivery system.

The objective of this paper is to present an analytical framework for evaluating the implications for the bulk power transmission network of increasing the dependence on renewable sources of energy in terms of the effects on 1) Operating Reliability and System Adequacy, and 2) the net-social benefits of making changes to the electric delivery system. This framework uses co-optimization to minimize the expected cost of meeting load with the intact system and a set of credible contingencies (equipment failures). This criterion is consistent with current practices for maintaining a standard of Operating Reliability. By allowing for involuntary load shedding to occur at a specified Value of Lost Load, it is also possible to determine how components of the network contribute to maintaining reliability. In simple terms, if the system load is increased by increments, eventually load shedding will be required to obtain feasible solutions. Usually load shedding occurs first in the contingencies. When this happens, the standard for Operating Reliability is effectively violated. It is then possible to determine where and how much additional capacity is needed on the network to avoid load shedding, and in this way, address the planning problem of maintaining System Adequacy. An important implication is that our co-optimization framework provides a consistent way to address both the real-time standard for Operating Reliability and the long-run standard for System Adequacy.

The discussion in Section 2 describes the current procedures used by regulators to maintain reliability, and explains why the use of proxy measures for reliability, such as reserve margins for generating capacity, obscures the true economic value of maintaining reliability. In Section 3, the co-optimization framework for determining an Optimal Power Flow (OPF) is presented. The new version of this software, the SuperOPF, uses a full AC representation of the network and includes a number of new capabilities, such as treating incremental and decremental real and reactive reserves explicitly, that will not be considered in this paper. Section 4 presents a case study using a 30-bus network. The network represents an urban center with relatively expensive local sources of generation and other regions with substantial amounts of relatively inexpensive sources of generation. Since the transmission capacity is limited, the urban center is in a load pocket. New wind capacity is added incrementally outside the load pocket, and this new capacity replaces some of the existing baseload capacity to represent a policy of switching from coal to renewables. The results of the case study are presented in Section 5 and are followed by the conclusions in Section 6.

2. Maintaining Reliability Standards

Federal legislators have formally recognized the importance of maintaining operating reliability in the Energy Policy Act of 2005 (EPACT05), and the major effect of this legislation is to give the Federal Energy Regulatory Commission (FERC) the overall authority to enforce reliability standards throughout the Eastern and Western Inter-Connections (see FERC [5]). The North-American Electric Reliability Corporation (NERC) has been appointed by FERC as the new Electric Reliability Organization (ERO), and NERC has been given the responsibility to specify explicit standards for reliability. Although it is still too early to know how well these arrangements will work, it is clear that the threat of paying penalties will be a tangible reason for state regulators to ensure that reliability standards are met.

In an electric supply system, the performance of the transmission network and the level of reliability are shared by all users of the network. Reliability has the characteristics of a "public" good (all customers benefit from the level of reliability without "consuming" it). In contrast, real electrical energy is a "private" good because the real energy used by one customer is no longer available to other customers. Markets can work well for private goods but tend to undersupply public goods, like reliability (and over-supply public "bads" like pollution). The reason is that customers are generally unwilling to pay their fair share of a public good because it is possible to rely on others to provide it (i.e. they are "free riders"). Some form of regulatory intervention is needed to make a market for a public good or a public bad socially efficient.

If a public good or a public bad has a simple quantitative measure that can be assigned to individual entities in a market, it is feasible to internalize the benefit or the cost in a modified market. For example, the emissions of sulfur and nitrogen oxides from a fossil fuel generator can be measured. Requiring every generator to purchase emission allowances for the quantities emitted makes pollution another production cost. Regulators determine a cap on the total number of allowances issued in a region, and this cap effectively limits the level of pollution. Independent (decentralized) decisions by individual generators in the market determine the pattern of emissions and the types of control mechanisms that are economically efficient. For example, the choice between purchasing low sulfur coal and installing a scrubber is left to market forces in a "cap-and-trade" market for emissions of sulfur dioxide.

Unfortunately, when dealing with the reliability of an electric supply system, it is impractical to measure and assign reliability to individual entities on the network in the same way that emissions can be assigned to individual generators. This is particularly true for transmission lines that are needed to maintain supply when equipment failures occur. The NERC uses the following two concepts to evaluate the reliability of the bulk electric supply system (see NERC [8]):

> *1. Adequacy — The ability of the electric system to supply the aggregate electrical demand and energy requirements of customers at all times, taking into account scheduled and reasonably expected unscheduled outages of system elements.*
>
> *2. Operating Reliability — The ability of the electric system to withstand sudden disturbances such as electric short circuits or unanticipated failure of system elements.*

Prior to EPACT05, the NERC standard of one day in ten years for the Loss of Load Expectation (LOLE) was widely accepted by regulators as the appropriate standard for the reliability of the bulk transmission system (i.e., this does not include outages of the local distribution systems caused, for example, by falling tree limbs and ice storms). Nevertheless, it is still very difficult to allocate the responsibilities for maintaining a standard of this type to individual owners of generating and transmission facilities because of the interdependencies that exist among the components of a network. This fundamental problem has not stopped regulators from trying to do it.

The basic approach used by state regulators in New England, New York and PJM is to assume that setting reserve margins for generating capacity (i.e., setting a standard for "generation adequacy") is an effective proxy for meeting the NERC reliability standard. This new proxy for reliability can now be viewed as the sum of its parts, like emissions from generators, and the task of maintaining generation adequacy can be turned over to market forces once the regulators have set a reserve margin. In New York State, regulators have gone one step further and passed the responsibility for purchasing enough generating capacity to meet the adequacy standard on to Load Serving Entities (LSE). Regulators decide what the amount of installed capacity should be in a region and the responsibility for acquiring this amount is prorated among the LSEs. An LSE that fails to comply would be fined (see NYISO [10] and [11]). In contrast, the ISOs in New England and PJM take the responsibility of purchasing the capacity needed in advance, and the cost is eventually prorated to LSEs using the actual load served in real time. This procedure identifies potential shortfalls of capacity in advance much more effectively than the NYISO procedure.

Even if the capacity markets are successful in maintaining generation adequacy, there are still important economic issues that are obscured when generation adequacy is used as a proxy for reliability. Changing a public good like reliability into a private good like installed capacity is a convenient sleight-of-hand for the advocates of deregulation because it then appears to be feasible to use market forces to maintain reliability standards. Nevertheless, this is not strictly correct because there is an implicit assumption that the transmission network is already adequate before decisions about generation adequacy are considered. It would be much more valuable for planning purposes to have a method of analysis that calculates the net-social benefits of generation and transmission assets in terms of both the delivery of real power to customers and the maintenance of reliability standards. This is particularly important for evaluating the role of renewables on a network because these sources are typically intermittent and require additional reserve capacity (or storage capacity) to maintain reliability. Before presenting the new analytical framework in the next section, some of the practical implications of adding an unreliable source of electricity are discussed.

The established reliability standard proposed by NERC is to limit failures to less than 1 day in 10 years. Is this standard too stringent, and therefore, more expensive to enforce than it should be? The answer is almost certainly no. The reason is that the Value of Lost Load (VOLL) when an unscheduled outage occurs is very high, particularly for large urban centers. A survey report published by the Lawrence Berkeley National Laboratory (LBNL) in 2004 (LaCommare and Eto [6]) concludes that the total cost of interruptions in electricity supply is $80 billion/year for the nation (op. cit. p. xi-xii), and 72% of this total is borne by the commercial sector (plus 26% by the industrial sector and only 2% by the residential sector). The frequency of interruptions is found to be an important determinant of the cost because the cost of an interruption increases less than proportionally with the length of an interruption. The costs of relatively short interruptions of only a few minutes are substantial.

The cost estimates in the LBNL report are developed from an earlier report prepared for the Office of Electric Transmission and Distribution in the U.S. Department of Energy (DOE, Lawton et al. [7]) that summarizes a number of different surveys of the outage costs for individual customers. For large commercial and industrial customers in different economic sectors, the average costs are reported for 1-hour outages in $/Peak kW (op.cit. Table 3-3, p.13). These average costs range from negligible for Construction to $168,000/MWh for Finance, Insurance and Real Estate, and the average cost for all sectors is $20,000/MWh. Although there is a lot of variability in the reported costs of an unscheduled outage, the overall conclusion is that the VOLL is very high for urban centers. The current NERC reliability standard of 1 day in 10 years corresponds to a VOLL of $33,393/MWh (60 + 80,000/2.4, based on an operating cost of $60/MWh and an annual capital cost of $80,000/MW for a peaking unit). Although this value is above the average value, it is still at the low end of the range of VOLL in the DOE report because the distribution of values is skewed to the right.

The key to deriving the economic value of maintaining a given reliability standard is to consider the benefits of avoiding unscheduled outages. In the empirical simulations discussed later in Section 4, a VOLL of $10,000/MWh is used. Consequently, reducing the probability of an unscheduled outage by 0.1%, for example, still saves $10/MWh. The analytical framework presented in the following section treats equipment failures (contingencies) explicitly. Some components of a network may only have a positive economic value when contingencies occur because they reduce the amount of Load-Not-Served (LNS). Other components, such as a new baseload unit, may reduce the cost of generation when the system is intact and have little affect on reliability. More generally, components will affect both operating costs for the intact system and reliability. For an intermittent source such as wind power, there is a fundamental tension between providing an inexpensive source of generation and making the existing network more vulnerable to outages. The solution to this predicament is to add new capabilities to the network that can compensate for the intermittent nature of wind power, such as load response and storage capacity. Evaluating the net-benefits of a portfolio of assets is the type of problem that can be evaluated using our new analytical framework.

3. The Analytical Framework

In a typical restructured market operated by an Independent System Operator (ISO), like the market in the New York Control Area, standards of Operating Reliability are met by requiring that minimum amounts of reserve capacity (spinning reserves) are available in different regions. These reserve requirements are the proxy measures of reliability discussed in the previous section. The generators submit price/quantity offers to sell energy and reserves into an auction, and the objective of the ISO is to determine the optimal patterns of generation and reserves by minimizing the total cost (the combined cost of energy and reserves) of meeting a forecasted pattern of load subject to network and system constraints and the specified amounts of reserves. The Last Accepted Offer is used to clear the market and set uniform market prices for energy and reserves. The market prices are adjusted for congestion and losses to determine the nodal prices for energy (i.e. Locational Based Marginal Prices (LBMP)). In addition, the auction determines the regional prices for reserves in a similar way.

Given the large number of nodes (over 400 in the New York Control Area) and the complexity of the network, it is computationally impractical to use a full AC representation of network flows to determine the OPF for a system of this size. As a result, a modified version of a DC OPF is used by the NYISO. For example, if the real flows on a transmission line are limited by a voltage constraint on a regular basis, the rated thermal capacity of the line is reduced in the dispatch to approximate this voltage constraint (an AC representation of network flows determines both real and reactive flows, but a DC representation determines only real flows). Hence, the lower thermal constraint on a transmission line is really another form of proxy limit that provides an additional distortion for determining the true shadow prices of transmission constraints. These distortions of the nodal prices are similar in effect to specifying minimum quantities of reserve capacity as proxies for reliability. The implications of using proxy variables in an OPF will be discussed in more detail in another paper. For this paper, the empirical analysis is based on an AC OPF using co-optimization to represent equipment failures (contingencies) explicitly in the objective function.

Fixed Reserve Requirements

To illustrate the specific differences between using co-optimization in an OPF instead of using the traditional fixed reserve requirements, it is convenient to start with the structure of an AC OPF using fixed reserve requirements. The objective criterion is to minimize the combined cost of energy, G_i, and reserves, R_i, needed to meet the forecasted pattern of load as follows:

$$\min_{G_i, R_i} \quad \sum_{\mathrm{i}=1}^{\mathrm{I}} [\, \mathrm{C}_{\mathrm{G}_\mathrm{i}}(G_i) + \mathrm{C}_{\mathrm{R}_\mathrm{i}}(R_i)] \tag{1}$$

subject to:

nodal power balancing constraints

$$F_j(\theta, V, G, Q) = 0 \quad \text{for } j = 1, \cdots, J \tag{2}$$

line power flow constraints

$$|S_l| \leq S_l^{\max} \quad \text{for } l = 1, \cdots, L \tag{3}$$

voltage limits

$$V_j^{\min} \leq V_j \leq V_j^{\max} \quad \text{for } j = 1, \cdots, J \tag{4}$$

real power limits

$$G_i^{\min} \leq G_i \leq G_i^{\max} \quad \text{for } i = 1, \cdots, I \tag{5}$$

reactive power limits

$$Q_i^{\min} \leq Q_i \leq Q_i^{\max} \quad \text{for } i = 1, \cdots, I \tag{6}$$

spinning reserve ramping limits

$$0 \leq R_i \leq R_i^{\max} \quad \text{for } i = 1, \cdots, I \tag{7}$$

unit capacity limits

$$G_i + R_i \leq G_i^{\max} \quad \text{for } i = 1, \cdots, I \tag{8}$$

Fixed Reserve Requirement for all N regions

$$\sum_{n=1}^{N} \sum_{i \in k} R_{ni} \geq \alpha \tag{9}$$

and **Fixed Reserve Requirements** for N* < N regions

$$\sum_{i \in n} R_{ni} \geq \alpha_n \quad \text{for } n = 1, \cdots, N^* \tag{10}$$

where i: generator index ($i = 1, 2, \cdots, I$)

j: bus index ($j = 1, 2, \cdots, J$)

l: transmission line index ($l = 1, 2, \cdots, L$)

n: regions ($n = 1, 2, \cdots, N$)

G_i / Q_i: real/reactive power output of generator i.

R_i: spinning reserve carried by generator i.

θ_j: voltage angle of bus j.

V_j: voltage magnitude of bus j.

S_l: power flow of line l.

$G_i^{\min}, G_i^{\max}$: minimum and maximum real energy for generator i.

$Q_i^{\min}, Q_i^{\max}$: minimum and maximum reactive power for generator i.

$R_i^{\max}$: maximum reserve for generator i.

$V_j^{\min}, V_j^{\max}$: voltage magnitude limits for bus j.

$S_l^{\max}$: power flow limit for line l.

$C_{G_i}(G_i)$: energy cost for operating generator i at output level G_{ik}.

$C_{R_i}(R_i)$: reserve cost for generator i carrying R_{ik} spinning reserve.

Equations (1) to (8) represent a standard OPF for an AC network, and (9) and (10) represent the mandated levels of reserve capacity needed in different regions to cover the unscheduled failure of equipment. In practice, determining the specified levels of reserves needed to meet the established standard of Operating Reliability depends on prior analyses, but it is likely that the actual mandated levels of reserve capacity are relatively conservative (i.e. high) to reduce the likelihood of facing the unpleasant political consequences of a blackout.

If Generator i with capacity G_i*, for example, is part of the optimal dispatch for the intact system, it could have an unexpected failure. In this case, Generator i would be eliminated and the OPF would be solved again using only the other generating units committed in the first optimal dispatch, after lowering the appropriate reserve requirements in (9) and (10) by G_i*. Hence, the actual dispatch and the prices paid could be substantially different from the optimal solution for the intact system if a contingency occurs. Furthermore, there is no guarantee that an optimal solution will actually be feasible in a given contingency. The feasibility of the dispatch is dependent on there being enough reserve capacity available in the right locations to cover the contingency, and in practice, the mandated levels of reserves are reset relatively infrequently as the characteristics of the system change over time.

2.2 Responsive Reserves Requirements (Co-optimization)

Chen et al. [3] have proposed an alternative way to determine the optimal dispatch and nodal prices in an energy-reserve market using "co-optimization" (CO-OPT). The new objective is to minimize the total expected cost (the combined production costs of energy and reserves) for a base case (intact system) and a specified set of credible contingencies (line-out, unit-lost, and high load) with their corresponding probabilities of occurring. Using CO-OPT, the optimal pattern of reserves is determined endogenously and it adjusts to changes in the physical and market conditions of the network. The initial motivation for developing the CO-OPT framework was to make the markets for reserves in load pockets less vulnerable to the exploitation of market power by generators. For this reason, the CO-OPT criterion is referred to as Responsive Reserve Requirements. If the offered prices for reserve capacity are high, the optimal solution will use fewer reserves by, for example, reducing the flow on a transmission tie line to reduce the size of the contingency if the tie line fails. This framework is equivalent to using a conventional n–1 contingency criterion to maintain Operating Reliability. In practice, the number of contingencies that affect the optimal dispatch is much smaller than the total number of contingencies. In other words, by covering a relatively small subset of critical contingencies, all of the remaining contingencies in the set can be covered without shedding load.

In the new SuperOPF, the CO-OPT criterion is modified to include the cost of Load-Not-Served (LNS) and also distinguishes between positive and negative reserves for both real and reactive power. Using this modified criterion, the System Operator determines the optimal dispatch for energy and reserves for the base case (intact system, k = 0) and for K different contingencies by minimizing the expected cost of meeting load in the (K + 1) states of the system as follows:

$$\min_{G_{ik},\, R_{ik},\, LNS_{jk}} \quad \sum_{k=0}^{K} p_k \left\{ \sum_{i=1}^{I} \left[C_{G_i}(G_{ik}) + C_{R_i}(R_{ik}) \right] \; + \; \sum_{j=1}^{J} VOLL_j \times LNS_{jk} \right\} \tag{11}$$

subject to (2) – (8) for each $k = 0,1,....,K$,

where $p_k > 0$ is the probability of contingency k occurring

$VOLL_j$ is the Value Of Lost Load for load at bus $j = 1,2,\ldots\ldots,J$

LNS_{jk} is the Load Not Served at bus j in contingency k

Implicitly, the constraints implied in (2) – (8) are replicated for $k = 0,1,....,K$ and adjusted to account for the failures of equipment that define each contingency. Note that the constraints (9) and (10) for Fixed Reserve Requirements are not included in (11). Given the structure of (11), the optimum criterion could be decoupled into $K + 1$ separate OPFs for $K + 1$ different systems. However, these solutions are coupled together through the following definition of reserve capacity:

Positive reserves

$$R_{ik+} = G_i^{*\max} - G_{i0}^{*} \quad \text{for } i = 1,2,....,I \text{ and } k = 0,1,....,K \tag{12}$$

Negative reserves

$$R_{ik-} = G_{i0}^{*} - G_i^{*\min} \quad \text{for } i = 1,2,....,I \text{ and } k = 0,1,....,K \tag{13}$$

where

G_{ik}^{*} is the optimal dispatch of energy from generator i in contingency k

$$G_i^{*\max} = \max_k (G_{ik}^*)$$

$$G_i^{*\min} = \min_k (G_{ik}^*)$$

Equations (11) – (13) represent a simplification of the SuperOPF because the real and reactive components of generation, reserves and load are not all identified explicitly, and for the remainder of the paper, the discussion only refers to real power (energy) even though reactive power is actually included in the case study. It would also be straightforward to include more flexible forms of load response in (11), such as interruptible contracts, but capabilities of this type were not used for this case study.

In (12) and (13), the reserve capacity for each generator in contingency *k* is defined as the difference between the maximum (minimum) of the $K + 1$ optimal levels of dispatch (which may be less than the true physical maximum) and the optimal level of dispatch for the intact system $k = 0$. In other words, the optimum quantities of energy and reserves for $k = 0$ are contracted ahead of real time and then the generators are also paid for the additional energy generated in real time. The maximum (minimum) dispatched capacity of every generator, $G_i^{*\max}$ $(G_i^{*\min})$, is needed for energy in at least one contingency. The level of reserve capacity for any generator is determined endogenously, and it responds to conditions on the network, such as the pattern of forecasted load. This feature is important for the case study due to the wide range of wind conditions that affect the actual generation from a wind farm and the difficulty in forecasting wind conditions accurately.

The regulated standard of Operating Reliability is maintained if load is met in all of the contingencies. Finding optimal values of $LNS_{jk} > 0$ is equivalent to violating this reliability standard, and it signals a failure of System Adequacy in a planning application that would be corrected by increasing the system capacity in some way. Since the *VOLL* is specified to be very large compared to typical market prices, it is important to note that a major part of the total benefit of many components of the grid comes from avoiding unscheduled load shedding when contingencies occur. When the system is Adequate, no failures of Operating Reliability will be observed, and therefore, it is no longer possible to use the observed market prices to determine the full net-benefit of an investment that was made to avoid unscheduled outages. These are the “Events that didn’t Happen” that should be considered when calculating the economic value of reliability in a planning model (see Mount et al. [8]).

One of the many useful capabilities of the SuperOPF is that the optimization can be considered in two stages. The first stage is the full co-optimization represented by (11) – (13) and it can be viewed as the optimum way to minimize the expected costs and maintain Operating Reliability when the system is Adequate (i.e. all $NSL_{jk} = 0$ for all credible contingencies). This stage determines the amounts and prices of energy and reserves contracted in advance of real time (e.g. one day ahead). The second stage corresponds to a real-time OPF when the actual state of the system is known and a contingency may have occurred. The objective cost is now to minimize the incremental cost of adjusting from the contracted amounts of resources from the first step to meet the actual system conditions.

The second stage of the SuperOPF treats the actual state as the new base ($k = 0$) and includes all of the remaining contingencies in the same way as before in (11) – (13). This implies that the optimum dispatch in the second stage still attempts to maintain Operating Reliability. However, if a major failure has already occurred, it may not be possible to meet the load in all situations if a second failure occurs. This would not be a violation of the typical standard of Operating Reliability assuming that the specification of the first stage covered all credible contingencies. For example, if the regulators define System Adequacy as the ability to cover all single failures, there is no guarantee that the system can cover the relatively rare event that two or more failures occur. Following any major contingency, bringing the system back into compliance with Operating Reliability would require adding existing resources that were rejected from the auction in the first stage of the optimization.

The current practices adopted in restructured markets are more in line with the optimization for Fixed Reserve Requirements in (1) – (10), and the expected cost of meeting the contingencies is not explicitly part of the objective function. In the New York Control Area, for example, a modified DC OPF minimizes the expected cost of meeting load for the intact system with specified levels of reserves included. If a contingency occurs, there is an ordered list of options, such as using reserve capacity and exercising contracts for interruptible load, with shedding load as the least desirable option. Since the contingencies are not considered explicitly in the optimization, it is virtually impossible to determine the true economic benefit of reliability from the market solutions, and meeting a given reliability standard is treated as a physical constraint rather than as an explicit economic component of the objective function as it is in the SuperOPF.

After a contingency occurs, the objective in the SuperOPF is still to minimize the expected cost over all contingencies even if this requires shedding some load in some contingencies. The amount and location of load shedding is determined optimally. For example, if the VOLL in an urban region is much higher than the VOLL in other regions, the solution will implicitly put more weight on avoiding the shedding of load in the urban area. In fact, the SuperOPF is consistent with the relatively successful market design in Australia.

In the Australian system, the market clears in real time every five minutes to meet load and to set the prices paid for the energy generated over the following five-minute period. These are the only prices used by the system operator to pay for energy. There are also forward markets, but these markets are financial and are not run or regulated by the system operator. The five-minute auction for energy includes a market for regulation and fast-responding reserves. These ancillary services receive payments for the reserve capacity contracted at the beginning of each five-minute period and for any energy that is actually generated. This is just like the first stage of the SuperOPF, but in the Australian market, the second stage never occurs. The five-minute auctions are like a continuous series of first stage optimizations. Capacity rejected in one period can still be entered into the next period's auction. Consequently, when a contingency occurs, the next market solution will bring new capacity into the market that was not needed (i.e. rejected from the auction) before the contingency occurred.

The incentive for ensuring that additional capacity will be ready to enter the market is provided to generators and loads by reporting forecasts of the prices a few hours ahead. These forecasted prices are determined by the existing offers and bids that have been submitted in advance but they are not binding for making payments. All payments for energy and ancillary services are made using the real-time prices. When the forecasted prices are high, and the price cap of $10,000A/MWh is relatively high in the Australian market, more generators are likely to enter the market and loads may adopt procedures for reducing demand in anticipation of the high prices. Another important feature of the Australian market is that the responses to a contingency before the next five-minute market clears are preset and automatic by, for example, using smart appliances as a fast way to shed load for a short period of time in response to a drop in frequency.

The following section describes the characteristics of the network used for the case study and the specifications of the simulations. The basic objective of the analysis is to evaluate the effects of increasing the load in a load pocket on Operating Reliability when the amount of installed generation and transmission capacity is fixed. The initial amounts of installed capacity are sufficient to meet the standard for System Adequacy and meet the load in all of the credible contingencies. As the load increases, standards of System Adequacy can no longer be maintained and some load has to be shed in some of the contingencies. In this case study, the focus is on showing how the analytical structure of the SuperOPF makes it possible to identify at what level of load and where on the network problems first occur. An important implication for regulators is that the high cost of shedding load is often localized in the sense that the high market prices are limited to a few nodes. As a result, the best way to fix a problem may be to add Distributed Energy Resources (DER) close to the affected loads rather than increase the capacity of the bulk power transmission network.

4. The Specifications for the Case Study

The case study is based on a 30-bus network that has been used extensively in our research to test the performance of different market designs using the *PowerWeb* platform. The one-line-diagram of this network is shown in Figure 1 below. The 30 nodes and the 39 lines are numbered in Figure 1 and this numbering scheme provides the key to identifying the locations of specific contingencies, constraints and shadow prices in the following discussion. In addition, the six generators are also identified. The network is divided into three regions, Areas 1 – 3, and Area 1 represents an urban load center with a large load, a high VOLL and expensive sources of local generation from Generators 1 and 2. The other two regions are rural with relatively small loads, low VOLLs and relatively inexpensive sources of generation from Generators 3 – 6. Consequently, an economically efficient dispatch uses the inexpensive generation in Areas 2 and 3 to cover the local loads and as much of the loads in Area 1 as possible. The capacities of the transmission tie lines linking Areas 2 and 3 with Area 1 (Lines 12, 14, 15 and 36) are the limiting factors. Since lines and generators may fail in contingencies, the generators in Area 1 are primarily needed to provide reserve capacity. The general structure of the network poses the same type of problem faced by the system operators and planners in the New York Control Area. Most of the load is in New York City (i.e. Area 1) and the inexpensive sources of baseload capacity (hydro, coal and nuclear) are located upstate (i.e. Areas 2 and 3).

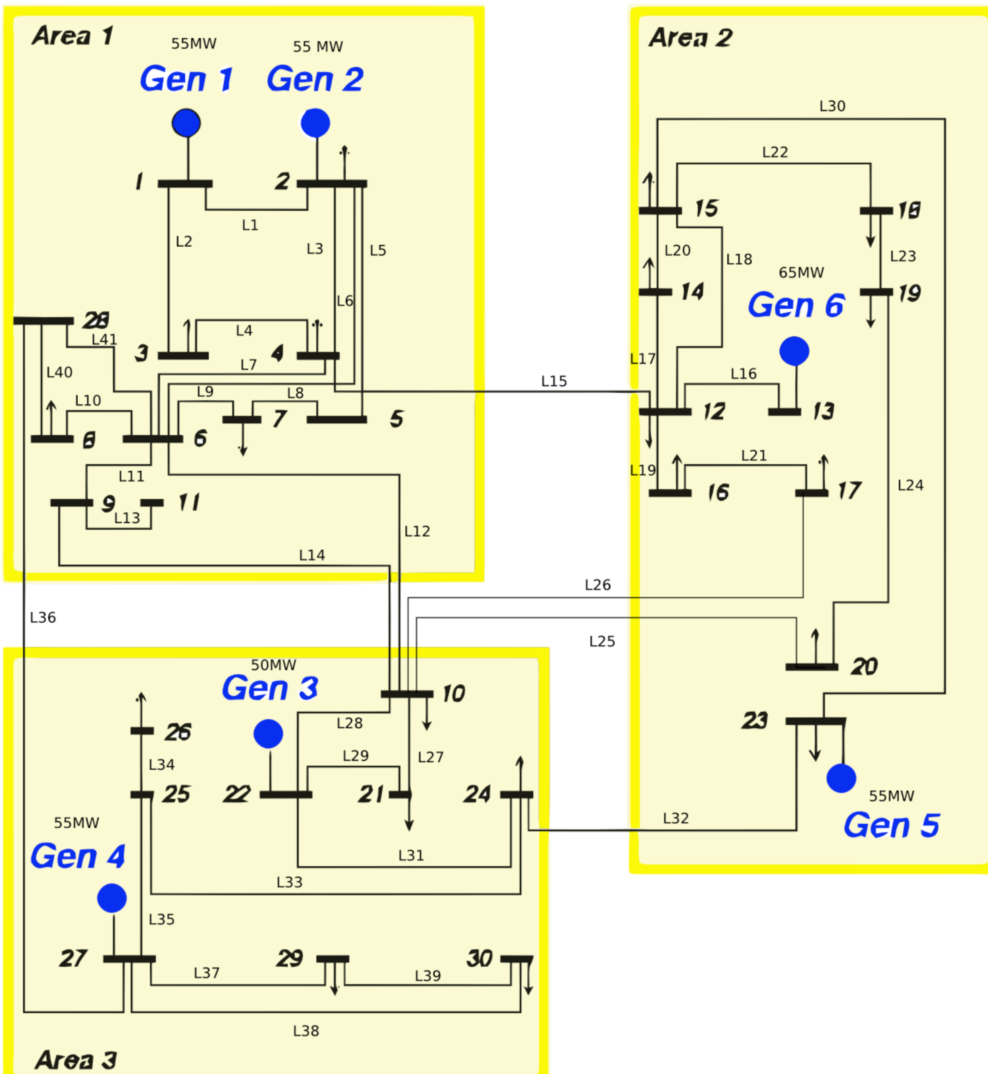

Figure 1: The One-Line-Diagram of the 30-Bus Network.

In this case study, the simulation increases the load in Area 1 in small increments until the capacity of the network is no longer able meet all loads in all contingencies. In other words, the standards for Operating Reliability and System Adequacy are eventually violated after the load has been increased sufficiently. The amounts and locations of the different types of installed generating capacity are shown in Table 1 together with the production costs. The levels of load and generation by Area are shown for the initial conditions (i.e. the lowest aggregate load) in Table 2. The total amounts of generating capacity in each Area are similar in Table 1, but the corresponding costs of production vary a lot and are much higher in Area 1.

Table 1: Installed Generating Capacity by Type and Location and the Production Costs

Area	Nuclear Hydro	Coal	Oil	Combined Cycle Gas	Gas Turbine	Total by Area
1	0	0	65MW	0	45MW	**110MW**
2	50MW	70MW	0	0	0	**120MW**
3	65MW	0	0	40MW	0	**105MW**
Total by Type	**115MW**	**70MW**	**65MW**	**40MW**	**45MW**	**335MW**
Production Cost	**\$5/MWh**	**\$25/MWh**	**\$95/MWh**	**\$55/MWh**	**\$80/MWh**	**-**

Table 2 shows that the initial system load is less than half the capacity of installed generating units, and under these conditions, the network has a lot of excess generating capacity. There is no generation in Area 1 in the base case (intact system) and transfers from Areas 2 and 3 are used to meet the load. However, 17MW are needed in Area 1 (33% of Load) to cover the contingencies (Gen. (max) – Gen. (base)). Exactly the opposite situation exists in Areas 2 and 3, and the levels of generation are substantially higher than the corresponding loads. The additional capacity needed to cover contingencies is smaller than the levels of generation in the base case, and the amounts of idle capacity (i.e. not used in any contingency) are relatively small in Areas 2 and 3 (21MW) compared to Area 1 (93MW). More of this unused capacity will be used as the load increases in Area 1, and the simulation covers a wide range of network conditions that illustrate clearly how the different types of constraint on network capacity affect nodal prices.

Table 2: Initial Patterns of Load and Generation by Area

Area	Load	Gen. (base)	% of Load	Gen. (min)	% of Load	Gen. (max)	% of Load	Idle	% of Load	Inst.	% of Load
	MW	MW		MW		MW		MW		MW	
1	50.7	0.0	0%	0.0	0%	16.8	33%	93.2	184%	110.0	217%
2	56.2	91.8	163%	73.6	131%	115.8	206%	4.2	8%	120.0	214%
3	48.5	65.0	134%	59.3	122%	87.1	180%	17.9	37%	105.0	217%
Total	**155.4**	**156.8**	**101%**	**132.9**	**86%**	**219.7**	**141%**	**115.3**	**74%**	**335.0**	**216%**

Gen. (base) - Generation for the intact system
Gen. (min) - Lowest generation in a contingency
Gen. (max) - Highest generation in a contingency
Inst. - Installed capacity

By maintaining Operating Reliability using the initial set of conditions on the network, there is an implicit assumption that the system is robust enough to meet all loads in all credible contingencies. The specific contingencies included in this case study are listed in Table 3. These contingencies include the failures of individual generators and transmission lines, and also the uncertainty about the actual level of load caused by the errors of forecasts when the optimum dispatch is determined a day ahead of real time, for example.

For generators and lines, there are only two possible outcomes. The first outcome is to perform as required in the optimum dispatch, and the second is to fail completely. However, the probability of failure is very small (0.2% for each failure in this case study), and as a result, the probability that each piece of equipment will perform as required is 99.8%. Since there are 15 failures identified in Table 3, the expected number of failures is 3 in 100 periods because the individual failures and periods are specified to be statistically independent. In other words, the system is expected to be intact 97% of the time. The last two contingencies correspond to errors in load forecasting, and there is a 1% probability that the system load will be substantially higher (lower) than the forecasted level. This capability to deal with uncertainty about load in the SuperOPF is even more useful when variable sources of generation such as wind turbines are part of the network.

Table 3: The Contingencies Used in the Case Study

0 = base case	95%
1 = line 1 : 1-2 (between gens 1 and 2, within area 1)	0.2%
2 = line 2 : 1-3 (from gen 1, within area 1)	0.2%
3 = line 3 : 2-4 (from gen 2, within area 1)	0.2%
4 = line 5 : 2-5 (from gen 2, within area 1)	0.2%
5 = line 6 : 2-6 (from gen 2, within area 1)	0.2%
6 = line 36 : 27-28 (main tie from area 3 to area 1)	0.2%
7 = line 15 : 4-12 (main tie from area 2 to area 1)	0.2%
8 = line 12 : 6-10 (other tie from area 3 to area 1)	0.2%
9 = line 14 : 9-10 (other tie from area 3 to area 1)	0.2%
10 = gen 1	0.2%
11 = gen 2	0.2%
12 = gen 3	0.2%
13 = gen 4	0.2%
14 = gen 5	0.2%
15 = gen 6	0.2%
16 = 10% increase in load	1.0%
17 = 10% decrease in load	1.0%

The basic structure of the simulation is to increase the five loads in Area 1 by proportional increments holding the pattern of loads constant in Areas 2 and 3. For each step in the simulation, the SuperOPF determines the optimal dispatch to meet load and maintain Operating Reliability using the 16 contingencies listed in Table 3. The expectation is that initially, as load increases in Area 1, generation in Areas 2 and 3 will increase until transmission limits on the tie lines are reached. When this happens, further increases in load will be covered by increases in generation from the expensive sources in Area 1. The market will fragment and the prices in the load pocket (Area 1) will be substantially higher than the prices in the Areas 2 and 3. Eventually, the capacity of the network will be insufficient to meet all loads in all contingencies, implying that the standard of Operating Reliability has been violated. This failure to meet all loads will occur first in one or more of the contingencies when equipment fails. Finally, the load will be so high that some load must be shed in the Base Case to obtain a feasible solution for the optimum dispatch. When this happens, the expected price is effectively at the VOLL at all nodes shed load.

In practice, it is difficult to predict exactly where on the network the high prices associated with load shedding will occur. Load shedding in a contingency may, for example, be caused by a voltage constraint on a specific transmission line. By incorporating the AC constraints on line flows in the optimization, this is exactly what the SuperOPF does well. The distortions caused by using proxy limits for network capacity in standard planning models tend to be more severe in these extreme situations when the network is stressed. By identifying the specific locations of the high prices where loads are shed, important information is provided for planning purposes to help determine exactly where on the network reliability has failed and what needs to be fixed. This is a necessary first step in determining whether the investment cost of upgrading the network to avoid load shedding can be justified in terms of the economic benefits from not shedding load. In simple terms, if the amount of load shed and the probability of this happening are both very small, the expected benefit may be smaller than the certain cost of financing the investment, and the investment would not be economically efficient. Even the most reliable network will fail to cover some very rare contingencies, such as cascading failures of equipment. The important point is that the structure of the SuperOPF makes it possible to evaluate the economic implications of meeting reliability standards instead of treating reliability as a set of additional physical constraints on network operations using, for example, minimum amounts of reserve generating capacity in different locations.

There are two different sources of uncertainty that need to be identified before evaluating the economic benefit of an investment in upgrading the capacity of a network. The first source comes from the inherent uncertainty about the state of the system in a co-optimization because the objective function in (11) determines the expected outcome over a set of different contingencies. It is not certain in a day-ahead market, for example, exactly what the state of the network will be in real time. The optimal dispatch determined by the SuperOPF represents a contracted pattern of generation and of upwards and downwards reserves, and the corresponding nodal prices (shadow prices) are the expected prices over the set of contingencies listed in Table 3. Maintaining Operating Reliability corresponds to having no unscheduled outages in any of the contingencies, and this is the case for the initial conditions summarized in Table 2. The main physical restriction on the choice of an optimal dispatch is that it must be possible to meet any one of the contingencies starting from the intact system ($k = 0$) without violating ramping constraints.

The second source of uncertainty is associated with the variability of the levels of load during a year. The expected economic value of an investment should consider the expectation over different contingencies and over different levels of load. The decrease in the expected annual cost of operating the system after making an investment in increased network capacity is the correct economic measure to compare with the annualized cost of financing this investment. In a planning application, the incremental increases in the loads in Area 1 discussed at the beginning of this section can be treated as the increases in the forecasted annual peak load for the system. Implicitly, it is assumed that if no load shedding is observed in any of the contingencies for a given peak load then Operating Reliability can be maintained for the other, lower levels of load throughout the year. In other words, the scheduling of the maintenance of generating units during the year is organized in a sensible way to avoid unscheduled outages, and for this case study, all generators are assumed to be available throughout the year to keep the analysis simple.

As soon as standards of Operating Reliability for a specified peak load are violated (because load shedding occurs in some contingencies), the network capacity is no longer Adequate and this is a signal that additional economic analysis is warranted to investigate whether violations also occur at lower levels of load. For this case study, the different levels of load that occur during a year are specified to be consistent with the patterns of loads in New York City and Long Island (for Area 1) and in upstate New York (for Areas 2 and 3) for a year with a relatively hot summer (2005). All loads within Area 1 and within Areas 2 and 3 vary proportionally to the aggregate loads in each of the regions. The simulation breaks a year into 100 equal steps and the corresponding scaled LDCs for each region are shown in Figure 2. Since both the LDC start at 100%, the relatively low values of the LDC for Area 1 imply that the loads in Area 1 are more affected by air conditioning in the summer than the loads are in Areas 2 and 3.

These procedures are used to complete the link between the annualized cost of an investment and the expected annual benefit from lower operating costs, and this link is essential for determining the net benefit of an investment. However, the procedures for specifying the pattern of loads during a year represent a rough approximation, and the loads used correspond to a single realization of the actual hourly time-series of loads in two regions. Incorporating the range of possible realizations of hourly load that could occur during a year will be the focus of future research.

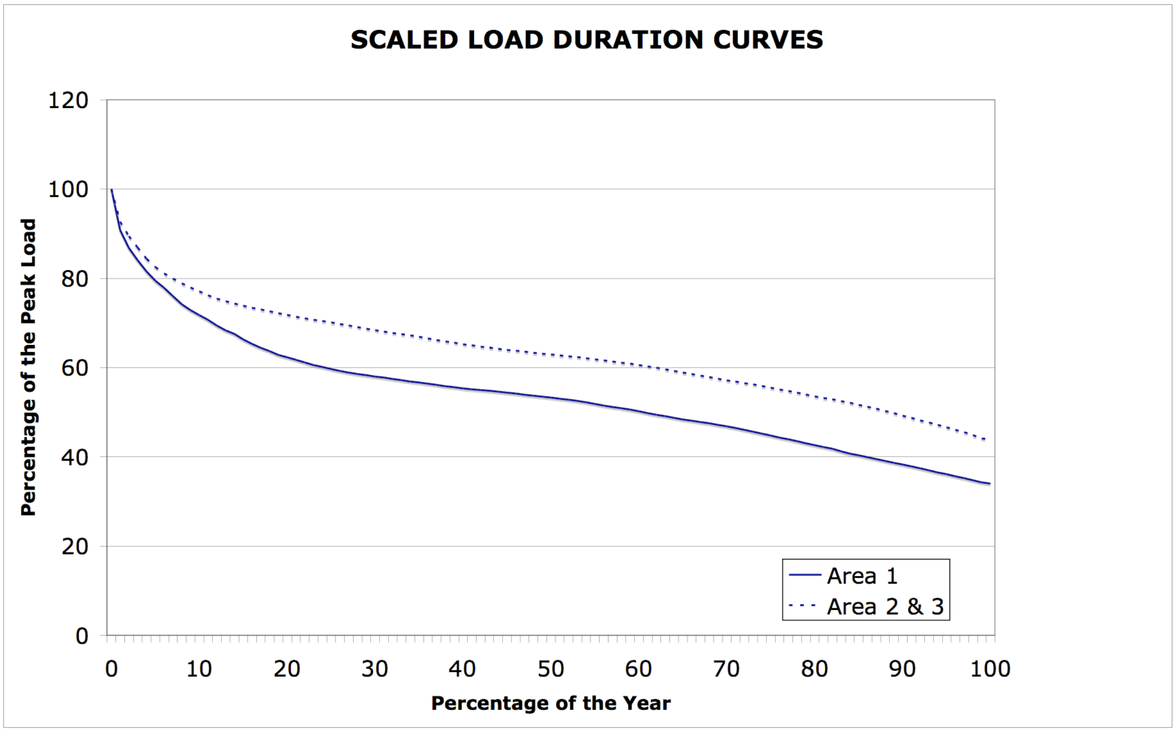

Figure 2: The Scaled Load Duration Curves for Area 1 and for Areas 2 and 3

5. Results of the Simulation

The simulation for this case study has two components. The first component starts with the network shown in Figure 1 and the initial conditions for installed capacity and costs summarized in Tables 1 and 2. The loads in Area 1 are then increased in proportional increments until the specified standard of Operating Reliability is violated. These loads represent the peak loads on the network as levels of demand increase over time. The second component of the simulation takes a specific level of peak load when the network is no longer "Adequate" and is unable to meet all loads in all contingencies. The simulation then determines the expected production costs for the annual pattern of loads described in Figure 2. This second component of the simulation makes it possible to determine the expected cost of violating the reliability standards.

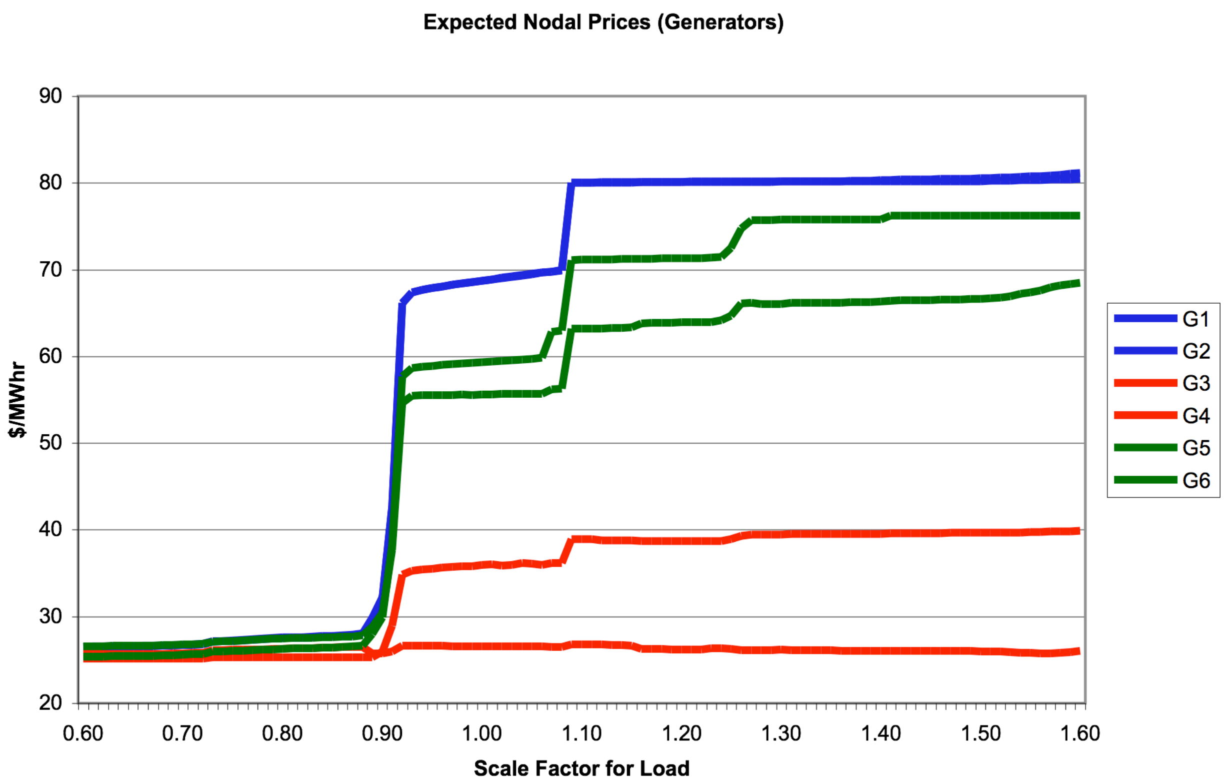

Figure 3: Expected Nodal Prices for Generators as the Peak Load Increases

The prices shown in Figure 3 represent the expected nodal prices for the six generators as the peak load on the network is increased (The "Scale Factor" measures the scale of the load in Area 1). Each nodal price is derived from the shadow prices determined by the SuperOPF in Equations (11) – (13), and it is the expectation of the shadow price over the list of contingencies shown in Table 3. Actual real-time prices would be different because more precise information about the state of network would be incorporated into determining these prices. For example, if a major piece of equipment has failed, the real-time prices in some locations may be much higher than they would have been if the system had remained intact. However, these high prices are weighted by a small probability when the expected nodal prices in Figure 3 are computed, and the biggest weight is put on the prices for the Base Case when the system is intact (Contingency 0 in Table 2).

At low levels of load (Scale Factor < 0.9), the nodal prices in Figure 3 are low and the price differences among the generators are small. Note that the nodal prices for Generators 1 and 2 can still be computed even though these units are only needed for reserve capacity when the system is intact. Under the low-load conditions, the network has a lot of excess transmission capacity and all loads can be met with generation from the baseload units. The coal units set the market prices at $25/MWh, and the small differences in the nodal prices reflect losses because there is no congestion on the network. When the system load increases sufficiently (Scale Factor > 0.9), congestion on the network occurs and the nodal prices for Generators 1 and 2 increase substantially to almost $70/MWh and finally to $90/MWh because the expensive units in Area 1 are needed to meet the load. The lowest prices in Figure 3 are for Generator 3 in Area 3 because the tie line from Area 3 to Area 1 (L36) and the adjoining distribution lines within Area 1 have limited capacity, and as a result, this unit is isolated by the network and cannot benefit from the high prices at other locations. Generators 5 and 6 in Area 2 do benefit from the higher prices even though the main tie line from Area 2 to Area 1 (L15) does get congested at high levels of load.

The expected nodal prices for the loads in the three areas are shown in Figure 4. Even though the behavior of these prices for most loads follows a similar pattern to the prices for the generators in Figure 3, the price for the load at Bus 8 is an anomaly and it increases to almost $10,000/MWh for the highest levels of load (Scale Factor > 1.30). This high price is the VOLL in Area 1 and it implies that some load at Bus 8 is being shed when the system is intact. In fact, all prices above $90/MWh for this bus are due to load shedding. When load shedding is limited to rare contingencies, the effect on the expected price is small and the expected price increases as load is shed in more of the contingencies. Eventually, load is shed in all contingencies including the Base Case when no equipment failures occur.

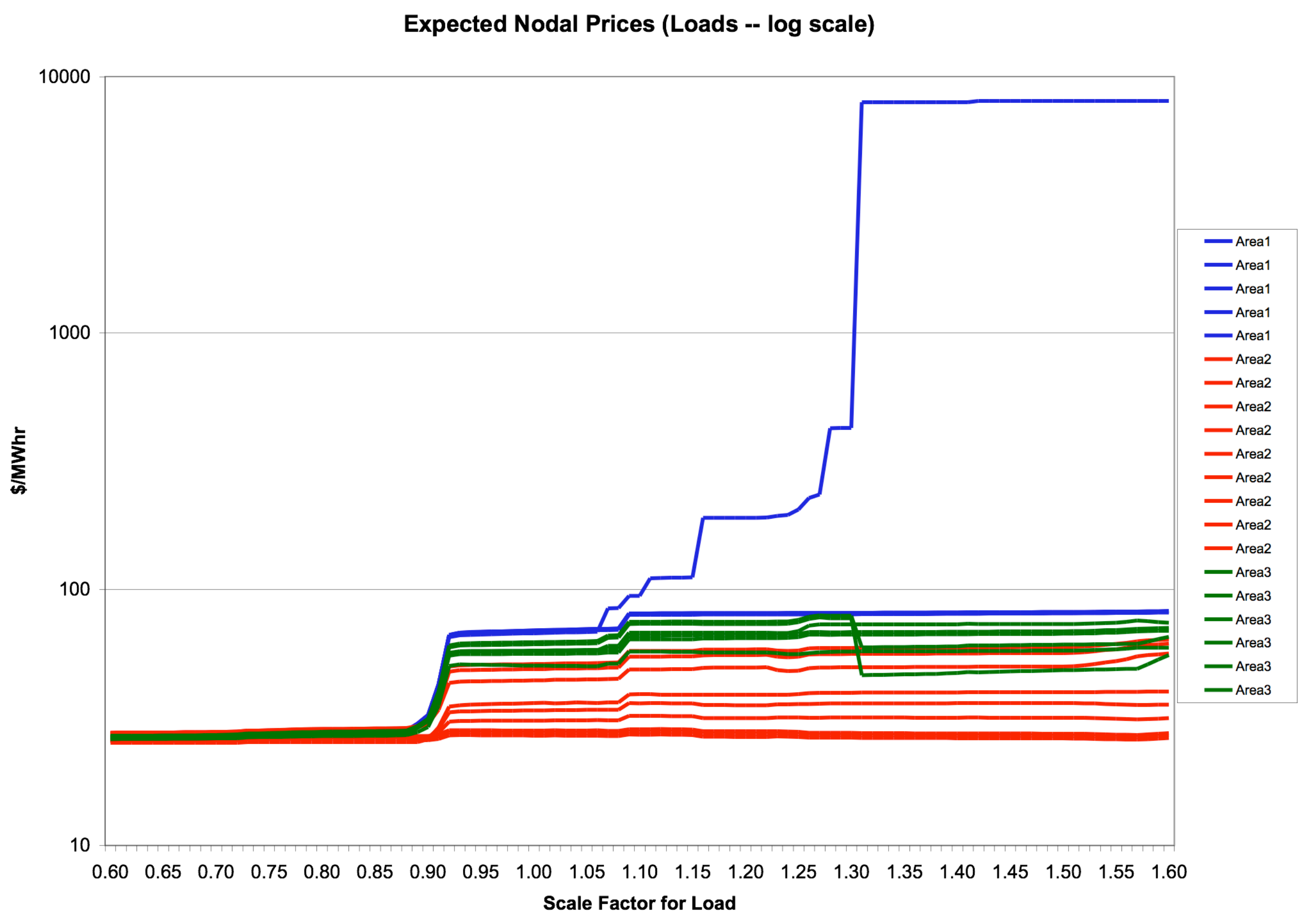

Figure 4: Expected Nodal Prices for Loads as the Peak Load Increases

There are two important implications for system planning that can be drawn from the prices in Figure 4. First, the results derived by the SuperOPF show when and where potential violations in reliability standards occur (i.e. load shedding at Bus 8), and as a result, they raise the important question, what is so special about the load at Bus 8? Why is the network able to meet the other loads in Area 1 and not meet the load at Bus 8? Second, the high prices at Bus 8 are very localized and have surprisingly little effect on the other nodal prices in Area 1. This suggests that the best solution for upgrading the network to meet reliability standards may be to add Distributed Energy Resources close to Bus 8 rather than to upgrade the tie lines into Area 1, for example.

A simple way to interpret the price changes in Figures 3 and 4 is to treat the initial increased differences in prices at different nodes as an indication of congestion on the tie lines into Area 1. The existence of a persistently large price difference between a low price region (Areas 2 and 3) and a high price region (Area 1) is the conventional rationale for upgrading transmission capacity on a tie line that is used by economists and advocates for merchant transmission projects. There is nothing basically wrong with this argument. The net benefits of a transmission upgrade should be evaluated if there are substantial amounts of inexpensive generation that could be built but could not be delivered to loads over the existing network. In fact, this was the main justification used by the FERC for encouraging merchant transfers and open access to the bulk power transmission network in Order 888. In fact, the type of analysis used to address network congestion is very similar to the typical economic analysis used to justify upgrading the capacity of a pipeline for natural gas.

There is a potential problem if economic analyses are limited to "pipeline' thinking when evaluating a transmission upgrade on an electric delivery system because it ignores the economic value of reliability. The large increase of the nodal price at Bus 8 in Figure 4 when the Scale Factor > 1.3 reflects the cost of having an unreliable network. It is quite possible in practice that the expected costs of these unscheduled outages are much larger than the expected costs of congestion because the VOLL in financial centers like New York City is so high. The overall conclusion is that a sound planning process should be able to evaluate the net economic benefits of both removing congestion and maintaining reliability standards, and in reality, it may be very difficult to allocate the cost of a specific upgrade in capacity to the "economic" component and the "reliability" component in a scientific way. A major benefit of using the SuperOPF is that both components are evaluated simultaneously as part of the standard optimization.

The expected costs of unscheduled load shedding are shown in Figure 5 for the different contingencies identified in Table 3. When the Scale Factor (referred to as the Load Factor in Figure 5) increases from 1.05 to 1.25, load shedding occurs in three different contingencies and the expected costs of load shedding are above zero, but when the Scale Factor is above 1.3, load shedding occurs in most contingencies. These high levels of load correspond to the jumps in the nodal prices in Area 1 seen in Figures 3 and 4. When the Scale Factor is above 1.35, load shedding occurs in the Base Case (Contingency 0), and as a result, the expected cost is close to the VOLL (note that the price scale in Figure 5 is logarithmic). As the system load increases in Figure 5, the following sequence of situations can be identified: 1) no load shedding occurs at low loads, 2) load shedding is localized to a small number of contingencies at slightly higher loads, 3) increasing the load further causes load shedding for almost all of the contingencies representing equipment failures, and finally 4) load is shed in the Base Case at the highest loads.

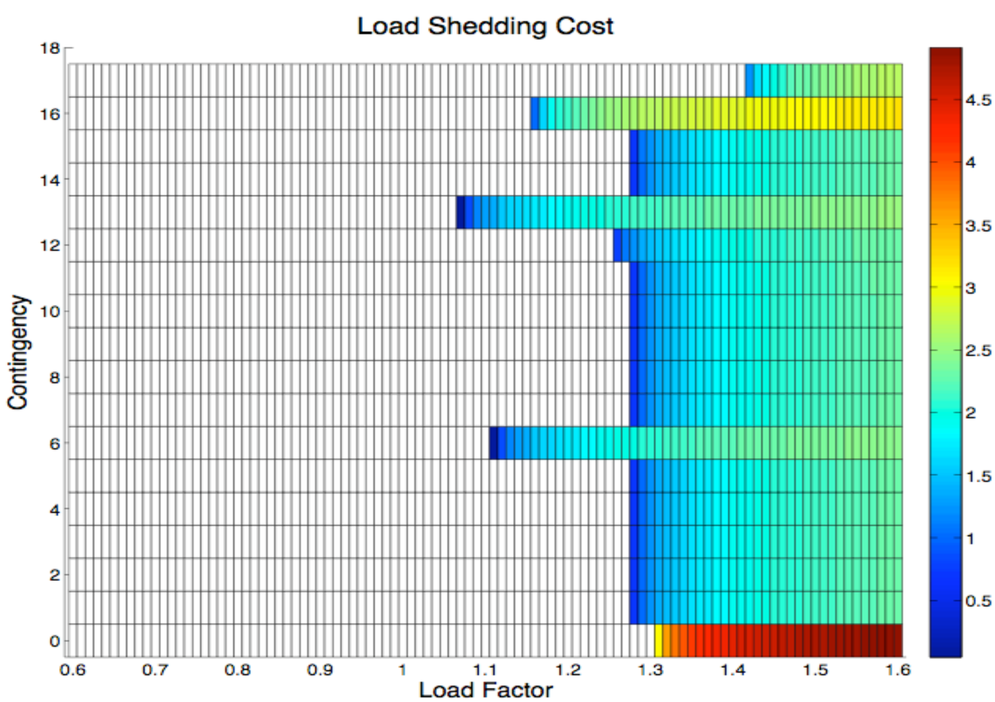

Figure 5: Expected Costs of Unscheduled Load Shedding in Different Contingencies

The discussion turns now to identifying what goes wrong when the load in Area 1 is increased sufficiently to cause load shedding at Bus 8, and which component(s) of the network should be fixed to maintain Operating Reliability and avoid shedding load at these higher levels of load. It should be noted, however, that the standard output from the SuperOPF computes shadow prices for all potential real and reactive constraints on the network for each one of the contingencies specified in Table 3. Hence the results that follow represent a highly selective sample that were chosen after screening all of the computed shadow prices to locate constraints that were binding (i.e. have non-zero shadow prices). In addition, there are many different ways to present these results. For example, the expected nodal prices in Figures 3 and 4 are the weighted averages of the shadow prices for real energy taken over all contingencies at the specific nodes for generators and loads, respectively. In contrast, the expected prices in Figure 5 are the weighted averages of the shadow prices for specific contingencies taken over the nodes for loads weighted by the amount of load shed. Furthermore, the underlying shadow prices for each node/contingency combination are determined before the actual network conditions are known using the probabilities of different contingencies occurring shown in Table 3. These shadow prices may not be exactly the same as the corresponding real-time shadow prices after a specific contingency has actually occurred and new information has been incorporated into the optimization. The shadow prices determined by the SuperOPF represent predictions for different contingencies based on the best information available at the time.

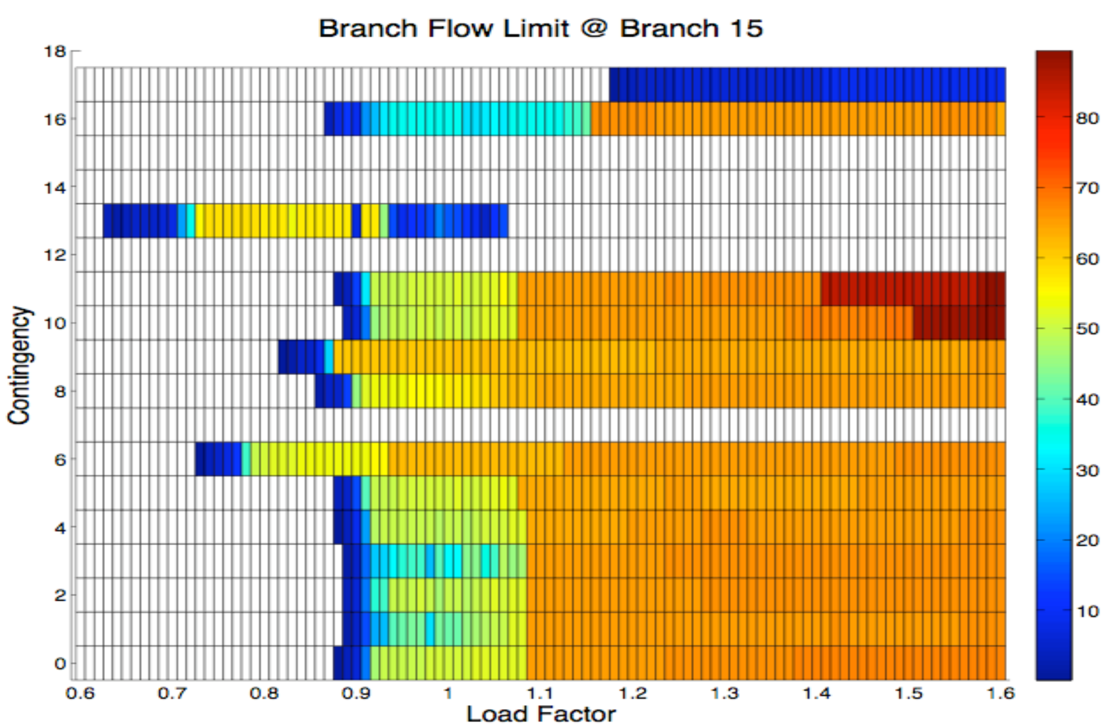

Figure 6: Shadow Prices for Line 15 Caused by Congestion

Figures 6 and 7 illustrate the shadow prices for specific transmission constraints that are caused by congestion and by load shedding, respectively. The shadow prices for different contingencies on the tie line from Area 2 to Area 1 (Line 15 in Figure 1) are shown in Figure 6 (note that Contingency 7 is the failure of Line 15, and therefore, the corresponding shadow prices cannot be computed). The non-zero shadow prices are caused by congestion that becomes apparent in some contingencies at relatively low levels of load (Scale Factor > 0.65), and congestion occurs for the Base Case (Contingency 0) when the Scale Factor is above 0.85. Nevertheless, the highest shadow price shown in Figure 6 is still relatively low (< $100/MWH) even when the system load is at its highest level.

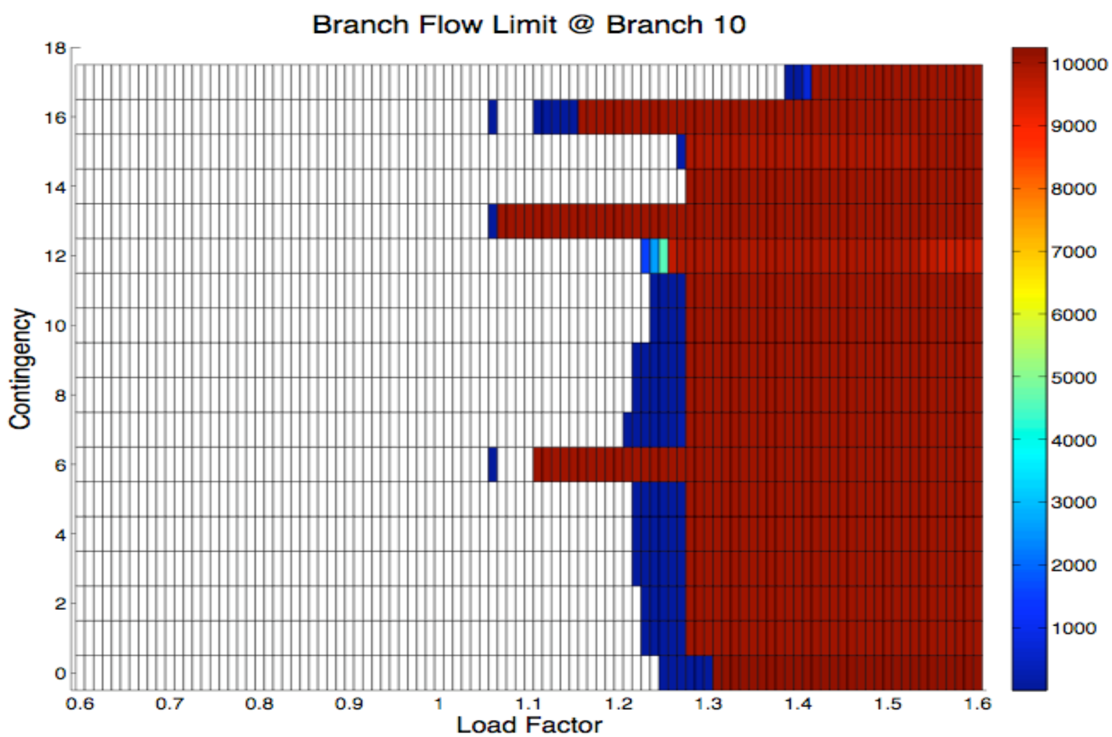

Figure 7: Shadow Prices for Line 10 Caused by Load Shedding

Figure 7 shows the shadow prices for Line 10 in the different contingencies. These prices reach $10,000/MWh in some contingencies when the Scale Factor is above 1.1, and in the Base Case (Contingency 0) when the Scale Factor is above 1.3. These levels of load correspond exactly to the levels of load in Figure 5 when load shedding occurs at Bus 8. It should be noted that Line 10 is not a major tie line into Area 1 but only a distribution line within Area 1 that links Bus 8 to Bus 6. The limited capacity of this line at higher levels of load is responsible for the load shedding at Bus 8 that violates the standard of Operating Reliability. An important implication for planning is that it is generally much easier to predict where congestion problems are likely to occur on a network than it is to predict the location of reliability problems.

Using knowledge of the levels of generation and costs of installed generating units, congestion occurs if some inexpensive units are not fully dispatched for energy or reserves. In practice, large differences in the nodal prices at different locations on the network indicate where this congestion is likey to be. On the other hand, reliability problems may be highly localized and hard to identify because they are associated with rare contingencies and may never actually be observed in a market. In fact, if standards of System Adequacy are maintained on a network, the high shadow prices associated with load shedding should not be observed. Furthermore, system operators often suspend market operations after major contingencies have occured so that the resulting high shadow prices are ignored. For these reasons, it is much better to deal with reliability issues before problems occur using appropriate analytical tools as part of an established planning process than it is to fix problems following an actual blackout. In the long run, maintaining reliabilty standards on an electric delivery system is just like maintaining other forms of instructure like bridges. It is very expensive and potentially dangerous to wait until things break before fixing them.

The final step in the analysis is to determine the expected annual cost of meeting load using the patterns of load shown in Figure 2. In this case study, the optimal dispatch for different levels of load that occur during a year can be computed in exactly the same way as they are for the different levels of load in Figures 3 – 7. Since the levels of load in Figures 3 – 7 represent the peak loads on the network, binding constraints are less likely to occur during the year at the lower levels of load, and as a result, the economic costs of congestion and load shedding will be smaller. At the lowest levels of load, the optimal dispatch will tend to be close to a merit order dispatch and the differences in nodal prices will be small, just as they are in Figures 3 and 4 at low levels of load.

Figure 8 shows how the averge shadow price charged to loads can be split into different components of cost for the ranked system loads in different hours in the year. The horizontal axis "Percentage of the Year" is consistent with the ranked loads in Figure 5. The underlying system load decreases moving from left to right. A peak system load of 200MW (Scale Factor = 1.1278) is chosen for the initial network conditions so that load shedding occurs in some contingencies (see Figure 5). Each plot line in Figure 8 is calculated as the expected revenue/cost over all contingencies summed over all relevant nodes (e.g. the nodes for each generator for the production cost) and divided by the total system load before any load shedding occurs. In this way, all of the variables in Figure 8 are measured in $/MWh and are calibrated in terms of the average price/cost per MWh demanded by customers with no outages.

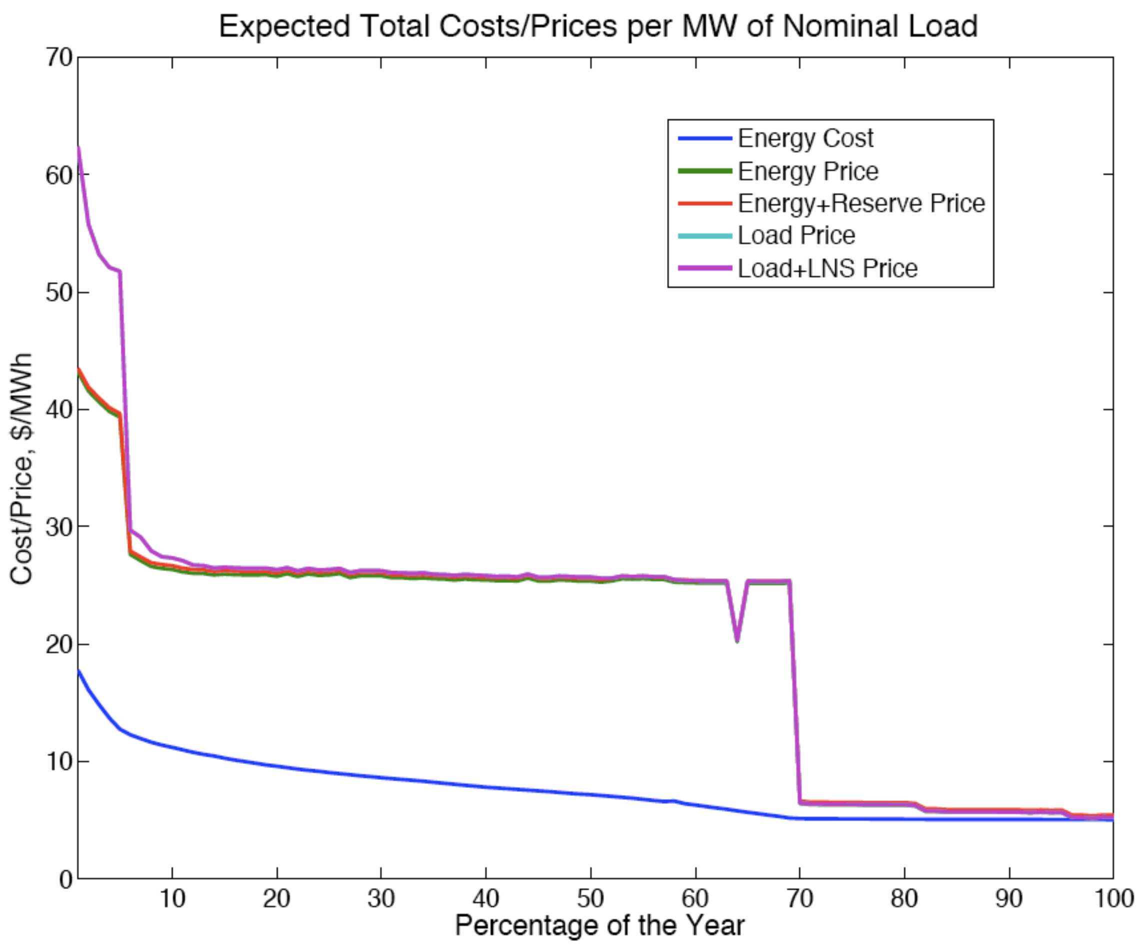

Figure 8: Components of the Average Price Paid by Loads During a Year

The two highest lines in Figure 8 represent the demand side of the market. The average price paid by loads (Load Price) corresponds to the expected nodal prices computed by the SuperOPF. The highest line (Load + LNS Price) includes the cost to customers of Load-Not-Served (LNS) (i.e. NLS x VOLL). In practice, the amount paid to the system operator by loads does not include the VOLL for customers who were served at a node where the load of some other customers was shed involuntarily. Furthermore, the customers at that node who actually faced outages are not compensated.

For the supply side of the market, the lowest line in Figure 8 represents the average production cost of meeting the system load (Energy Cost). Since generators are paid the shadow price at the appropriate node, the price actually paid for energy (Energy Price) is usually higher than the true production cost. Figure 8 also shows the average of payments for energy and reserves (Energy + Reserve Price). In this example, the cost of purchasing reserves is relatively small. It should be noted that the reported payments to generators include only the payments for real power and real reserves. However, the SuperOPF does actually compute the corresponding payments for reactive power and reactive reserves, but in this example, these payments are trivially small. In contrast, the nodal prices for loads include the cost of buying reactive power implicitly because each load is specified to have a fixed load factor.

Figure 9 shows the equivalent breakdown as Figure 8 in terms of the total hourly payments for different levels of system load during the year. The difference between the Energy Revenue and the Energy Cost measures the net revenue paid to generators above production costs that can be used to cover capital costs. This net revenue is large if expensive peaking units set a high price for baseload units with low production costs, and this is the situation when the Percentage of the Year is less than 70%. For lower levels of load with Percentage of the Year above 70%, the baseload units set the price and the net revenue is essentially zero. For low levels of system load (high values of Percentage of Load), customers pay prices that are very similar to the prices paid to generators. This implies that there is little congestion on the network. In contrast, loads pay a lot more than generators receive at higher levels of load, and this difference can be attributed to increasing congestion on the transmission network. For the highest levels of system load (Percentage of the Year < 10%), the true cost of generation increases and the payments to generators and by loads increases substantially more. This is when congestion on the network becomes severe, and finally when some load is shed, there is a spike in the average payments made by loads.

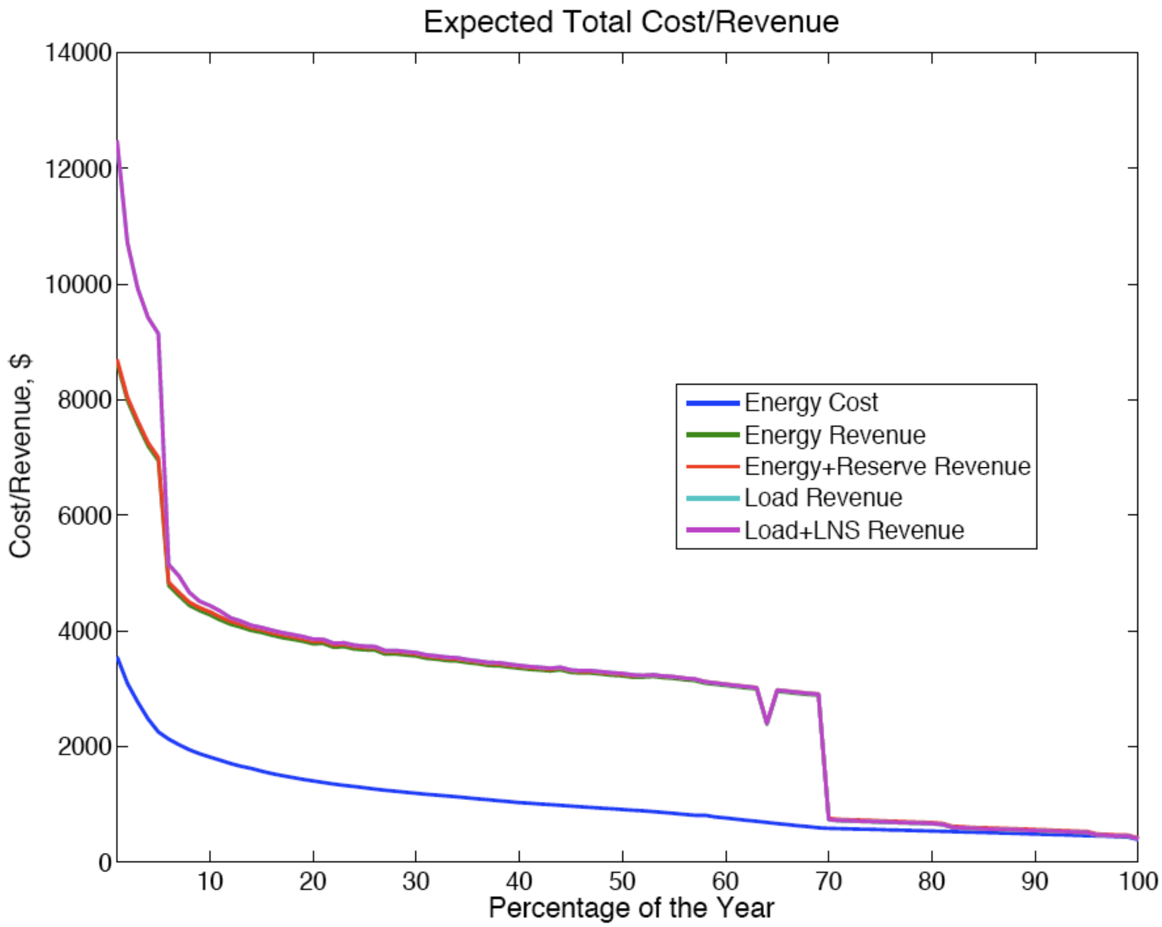

Figure 9: Components of the Hourly Payments Made by Loads During a Year

Aggregating the costs, revenues and payments in Figure 9 over all hours of the year provides the basic information needed to evaluate the effects of congestion and reliability on the aggregate market outcomes and the implications for the participants on the demand side and the supply side. Although the results presented are aggregated, this type of information can also be computed for individual nodes, and therefore, the implications for a load or a generator at any particular location can also be determined. In addition, by evaluating the nodal price differences and flows on a specific transmission line, the same type of information can be determined for transmission owners.

Table 4: Expected Annual Payments, Revenues and Costs ($/Year)

DEMAND SIDE		
1	Load Payment + Cost of LNS*	26,062,479
2	Load Payment	26,058,061
SUPPLY SIDE		
3	Energy + Reserves Revenue	24,792,467
4	Energy Revenue	24,506,897
5	Energy Cost	9,020,838
DERIVED VALUES		
6	Reliability Cost of LNS* (1 - 2)	4,418
7	Congestion Cost (2 - 3)	1,265,594
8	Net Revenue for Generators (4 - 5)	15,486,059

*NLS is Load-Not-Served

The aggregate annual values of the different components presented in Figure 9 are shown in Table 4 for the whole network. In addition, the corresponding costs of unreliability (shedding load), congestion and the net revenue above production costs for generators are computed. In this example, the expected cost of shedding load is very small (< $5,000/Year), and as a result, it is probably too small to justify fixing the problem. Using a conventional economic criterion, if the annualized cost of an investment to upgrade Line 10 is less than this amount, it would be economically beneficial to make the investment. Even though the VOLL is $10,000/MWh when load is shed, the amount of load shed is very small and the probability of the contingencies actually occurring in which load is shed is also very small. When these two features are combined, the implied amount of expected load shed is less than 0.5MW/Year for the system as a whole ($4,418 divided by the VOLL of $10,000). The NERC standard of limiting unscheduled outages to less than one day in ten years (< 2.4 Hours/Year) does not specify the quantity affected. However, it is probably realistic to interpret the NERC rule as stating that no load node should experience a complete outage for more than 2.4 hours in a year. In contrast, the SuperOPF allows partial outages to occur in which the loads for some customers at a node are shed but not for all customers at that node.

Reconciling the economic and NERC definitions of reliability for a bulk power transmission system will be the subject for future research using the SuperOPF.

For this particular example, the cost of congestion in Table 4 (> $1.2million/Year) is nearly 300 times larger than the cost of failing to maintain reliability. However, making an investment to reduce congestion requires the evaluation of a number of different candidate transmission lines for upgrading the network. This is exactly the type of analysis that can be done using the SuperOPF. In our experience, persistent high price differences on a transmission line are a necessary but not a sufficient condition for getting an economic return from upgrading the capacity of that line on a meshed network. However, the line with the phantom price differences are generally located next to the line that is really causing the congestion. This is another topic for future research.

The final derived cost in Table 4 is the net revenue above operating costs paid to generators. This cost (> $15million/Year) is over ten times greater than the cost to the system of congestion. In this example, most of the "excess" money paid by customers above the true operating costs goes to pay generators, and there is no mandate in a deregulated market about how this money should be spent. Even though this example is only a special case, the results raise the question of who is really benefiting from deregulation, and from the point of view of reliability, do the excess payments made by loads above out-of-pocket expenses provide the correct economic incentives needed to maintain standards of Operating Reliability? A casual answer to the question is no, and once again, this topic will be the focus of future research.

The costs presented in Table 4 illustrate, in aggregate, the types of information that can be derived from the SuperOPF. In terms of evaluating any proposed upgrade to an existing network, the capabilities of the SuperOPF could be used to calculate how the individual components of the annual costs of running this network change with and without a specified investment. This is exactly the type of capability that regulators should have available when evaluating the net public benefit of proposed changes to a network's capabilities. In practice, the conventional analytical procedures used by regulators fall far short of having this essential analytical capability. A basic criterion for judging planning models in the future should be that they can evaluate the economic consequences of congestion and reliability simultaneously for any AC network specification.

6. Conclusions

The main purpose of this paper is to illustrate how the new SuperOPF developed by PSERC researchers at Cornell can be used to determine the net social benefit of system reliability on a network with a specified pattern of loads. The important features of the SuperOPF are 1) failures of equipment (contingencies) are considered explicitly in the optimization, 2) load shedding at a high Value-of-Lost-Load (VOLL) is allowed in all contingencies, and 3) the optimization incorporates the nonlinear constraints of a full AC network. These three features make it possible to 1) determine the correct shadow prices for different components of the network under different operating conditions, 2) calculate the correct net social benefit of maintaining Operating Reliability, and 3) evaluate the net economic benefit of an investment that lowers expected production costs.

In contrast, most conventional algorithms for determining the dispatch of generators simplify the nonlinear computations by using proxy limits on network capacity, such as lowering the thermal limits of transmission lines. These proxy limits inevitably distort the shadow prices computed in the optimization. Furthermore, proxy measures, such as minimum reserve margins for generating capacity in different locations, are included in the optimization as additional constraints to represent the Operating Reliability. This procedure makes Operating Reliability a physical constraint rather than an economic requirement. In reality, the economic benefits of some components of a network are determined exclusively by avoiding the high cost of unscheduled outages when equipment fails in relatively rare contingencies. In the SuperOPF, the shadow prices and the level of Operating Reliability reflect the actual operating conditions, and high shadow prices tend to occur under adverse conditions when the network is congested due, for example, to high levels of load or equipment failures. These adverse situations are the most important for determining the true economic benefit of different components of a network, but these situations are exactly the ones in which the shadow prices are the most distorted using conventional algorithms.

Using conventional dispatching algorithms, it is potentially misleading to use the observed nodal prices of real power and ancillary services in a market as a guide for identifying what should be fixed on a network when standards of Operating Reliability are violated. The fundamental limitations of conventional planning tools are largely responsible for the attempt by many regulators to make a clear distinction between "economic" investments and "reliability" investments when planning capacity expansions. Although this is a convenient simplification, this practice completely ignores the true economic benefit of maintaining a high level of reliability. In reality, most upgrades of a network, particularly of transmission lines, affect both production costs and reliability. Determining how much of the capital cost of a specific upgrade should be treated as a reliability upgrade versus an economic upgrade is quite arbitrary using conventional planning tools.

Reliability should be treated as an economic decision that depends on the actual operating characteristics of the network. Using the SuperOPF, shedding load at specific locations in one or more contingencies is an explicit indication that the level of reliability has deteriorated and where on the network the problems have occurred. The basic planning decision is to determine whether an investment in upgrading capacity is justified by showing that the annual cost of this investment is less than the product of the high value of the Load-Not-Served times the small probability that the contingencies in which the outages happen actually occurs. Basically, it is not economically efficient or practical to avoid outages in all possible contingencies. The case study presented in this paper demonstrates how the SuperOPF can be used to address reliability questions using an analytical framework that links the short-run criterion of Operating Reliability with the long-run criterion of System Adequacy in a consistent way.

7. Acknowledgements

This research was supported by the US Department of Energy through the Consortium for Electric Reliability Technology Solutions (CERTS) and by the Power Systems Engineering Research Center (PSERC). The authors are responsible for all conclusions presented in the paper, and the views expressed have not been endorsed by the sponsoring agencies.

8. References

[1] Anderson, Lindsay and Judy Cardell, *Reducing the Variability of Wind Power Generation for Participation in Day Ahead Electricity Markets,* Proceedings of the 41st Annual Hawaii International Conference on System Sciences, January 2008.

[2] Chen, J., T. D. Mount, J. S. Thorp and R. J. Thomas, *Location-based scheduling and pricing for energy and reserves: a responsive reserve market proposal*, Decision Support Systems, 40, pp. 563-577, October 2005.

[3] Chen, J., J. S. Thorp and R. J. Thomas, *Time–space methods for determining locational reserves: a framework for location-based pricing and scheduling for reserve markets*, CERTS Report on Reliability Adequacy Tools, 2002.

[4] Chen, J., J. S. Thorp, R. J. Thomas and T. D. Mount, *Locational Pricing and Scheduling for an Integrated Energy-Reserve Market*, Proceedings of the 36th Annual Hawaii international Conference on System Sciences - Track 2 - Volume *2*, IEEE Computer Society, pp. 54.1, 2003.

[5] Federal Energy Regulatory Commission (FERC), *Energy Policy Act of 2005 (EPACT05)*, Washington DC, 2005.

[6] K. H. LaCommare and J. H. Eto, *Understanding the cost of power interruptions to U.S. electricity consumers*, Ernest Orlando Lawrence Berkeley National Laboratory, Berkeley, CA (US), 2004.

[7] Lawton, Leora, Michael Sullivan, Kent Van Liere, and Aaron Katz (Population Research Systems, LLC) and Joseph Eto (LBNL), *A Framework and Review of Customer Outage Costs: Integration and Analysis of Electric Utility Outage Cost Surveys*", Office of Electric Transmission and Distribution, U.S. Department of Energy, Washington DC, November 2003.

[8] Mount, Tim, Alberto Lamadrid, Surin Maneevitjit, Bob Thomas and Ray Zimmerman, *The Economics of Reliability and the Importance of Events that Didn't Happen,* Invited Presentation at the Fourth Annual Carnegie-Mellon University Conference on the Electricity Industry, March 2008.

[9] North-American Electric Reliability Corporation (NERC), *2007 Long-Term Reliability Assessment 2007-2016,* Princeton NJ, October 2007.

[10] New York Independent System Operator (NYISO), *2004 Load and Capacity Data*, Rensselaer NY, 2004.

[11] New York Independent System Operator (NYISO), *Comprehensive Reliability Planning Process Draft Reliability Needs Assessment*, Rensselaer NY, September 2005.

[12] New York Independent System Operator (*NYISO*), *Installed Capacity Manual Revision 5.1.1*, Rensselaer NY, September 2004.