A Nodal Capacity Market to Assure Resource Adequacy

Aleksandr M. Rudkevich
Charles River Associates
arudkevich@crai.com

Anatoly I. Lazebnik
CJSC “Energy Forecasting Agency”
lazebnikai@gmail.com

Igor S. Sorokin
CJSC “Energy Forecasting Agency”
isorokin@e-apbe.ru

Abstract

In power systems with no transmission constraints a simple relationship links resource adequacy criteria LOLP or LOLE with VOLL. This relationship justifies the choice of the LOLP/LOLE criteria as a guiding principle for making optimal investments in generation expansion. In constrained power systems LOLP/LOLE criteria do not lead to optimal investment decisions. Locational resource adequacy requirements presently used in known capacity markets are not economically justified and do not lead to optimal investment decisions.

We formulate the nodal capacity expansion problem in the form of a stochastic optimization based auction which solves for the optimal resource subject to resource adequacy criteria in the form of value of lost load. We explore the properties of this formulation, define locational capacity prices and a consistent system of load payments and generator compensations for reliability received and provided. Provided examples illustrate the importance of using the full network consideration in resource adequacy assessments.

1. Introduction

Assuring resource adequacy of a bulk power system is one of the main criteria of system planning. Traditionally, system planners follow a two-stage process. At first stage, resource adequacy of the system is studied using such probabilistic criteria as Loss of Load Probability (LOLP), Loss of Load Expectation (LOLE), Loss of Load Hours (LOLH), Expected Unserved Energy (EUE) and similar metrics [1-3]. One of the key outcomes of the first stage is the planning reserve margin requirement. Installed capacity equal or exceeding that requirement assures the desired probabilistically measured level of resource adequacy. The second stage is dedicated to capacity expansion. In a vertically integrated environment, capacity expansion is addressed as a long-term optimization problem. In deregulated systems, optimization problem is effectively replaced by capacity markets. Despite the innovative nature of the second approach, both essentially face similar challenges:

- what is the right indicator of resource adequacy? (see for example [1]);
- what is the right threshold for that indicator? (see [4] for a good discussion of this issue); and
- how to account for transmission constraints in assessing resource adequacy of the power system? (this challenge is particularly addressed in [5, 6, 7]).

As we demonstrate in this paper, these three challenges are intricately interrelated. The underlying economic principle of capacity expansion is very simple and well understood: more capacity should be built as long as its incremental cost is exceeded by the expected damage caused by the loss of load. Both the structure of the resource adequacy criteria and their threshold values should be consistent with this principle.

In an unconstrained power system, this principle is articulated expressed by a well-known equation [8]

\[
CONE = \text{LOLP} \times \text{VOLL}
\]

linking LOLP, Value of Lost Load (VOLL) and the annualized marginal cost of new entry (CONE). The appearance in this equation of LOLP, as opposed to other indicators of resource adequacy, is not arbitrary, but dictated by the particular structure of the optimization problem expressing the stated economic principle.

Transmission constraints change the structure of the optimization problem and renders the use of the LOLP criteria neither justifiable, nor correct. Unfortunately, this has not been recognized; not in theory and not in practice.

In theory, the impact of transmission constraints on resource adequacy planning is being actively studied [6,7]. However, a LOLP criterion is imposed without justification on the system as a whole in the form of a chance constraint. In contrast, as shown in this paper, the optimal capacity selection consistent with the stated economic principle leads to locational reliability criteria: specifically defined reliability standard should be met at each bus rather than for the system as a whole.

In practice, systems such as New York [9], PJM [10], Midwest ISO [11] and ISO New England [12]
use system-wide LOLE-type resource adequacy criteria of 1-day-in-10-years combined with ad-hoc approaches for determining locational installed capacity requirements. To the best of our knowledge, no economic analysis supporting these ad-hoc rules could be found in the literature.

To address identified shortcomings of theoretical foundations and current practices assuring resource adequacy, in this paper we explicitly consider a nodal formulation of the capacity expansion problem in the form of the stochastic capacity market auction. This approach embeds resource adequacy assessment into the auction selections process. Such a formulation leads to a coherent nodal theory of resource adequacy and capacity pricing.

2. Model Formulation

2.1. Representation of the Bulk Power System

We consider a DC linearized lossless\(^1\) representation of the power system such that the power flows \( f \) on transmission facilities and flowgates are represented by the following equation

\[
f(t, \omega) = f^o(t) + \Psi[p(t, \omega) - p^o(t) - (L(t, \omega) - L^o(t))]
\]

where the superscript \(^o\) indicates a base case power flow quantity, and \( p \) and \( L \) are the vectors of nodal power injections and loads, respectively. The power system is considered over time and under multiple stochastic scenarios.

For the purpose of this paper, we assume that the vector of power injections includes both actual physical injections by generating resources and virtual injections in the form of controllable by a system operator demand response measures. These physical generating resources and demand response measures are commonly referred to as generation resources, or resources for short, characterized by their capacities, availabilities and low bound operational limitations, as discussed below. Time spans over the planning horizon \( 0 \leq t \leq T \).

For the purpose of this paper, a finite number of stochastic scenarios numbered by \( \omega = 1, 2, ..., N \) is considered, where \( N \) is the total number of stochastic scenarios with probabilities \( \pi(\omega) \geq 0 \) and

\[
\sum_{\omega=1}^{\infty} \pi(\omega) = 1.
\]

Demand requirements, resource availability and transmission topology may change over time and these changes vary by stochastic scenario.

The base power flow parameters and the power flow itself caused by system conditions are dependent on time \( t \). The transmission sensitivity matrix \( \Psi \) gives the variations in flows due to changes in the nodal injections, with the reference bus assumed to ensure the real power balance. This matrix could also depend on time and be subject to random perturbations due to transmission line outages.

The system is considered under load conditions represented by a vector stochastic process \( L(\omega, t) \). Power flows represented by equation (2) must remain within security limits of the predetermined set of flowgates. Flowgate limits are also considered stochastic and time-dependent:

\[
f(t, \omega) \leq f \leq f^o(t, \omega)
\]

In general, the upper limits of all flowgates are non-negative, and lower limits of flowgates are non-positive:

\[
f(t, \omega) \leq 0, f^o(t, \omega) \geq 0
\]

Power injections by resources are bounded from above by the rated resource capacities and resource availabilities. The level of resource availability varies by stochastic scenario and over time. Power injections of resources could also be bound from below by operational constraints:

\[
S(t, \omega)X \leq p(t, \omega) \leq S(t, \omega)X
\]

where \( X \) is a vector of rated capacities. Stochastic resource availabilities and per unit low bound operational limitations are positioned on diagonals of matrices \( S(t, \omega) \) and \( S(t, \omega) \), respectively, while all other elements of these matrices are zeros. Upper limits of resource availability and its low bound operational limitation are in order:

\[
0 \leq \underline{s}(t, \omega) \leq \overline{s}(t, \omega) \leq 1
\]

In addition to resources and load, for each bus we define a non-negative value of unserved energy represented by vector \( u \) such that the system balance, security and load interruption feasibility equations take the following form, respectively:

\[
\frac{1}{2}(p(t, \omega) + u(t, \omega) - L(t, \omega)) = 0
\]

\[
F(t, \omega) \leq \Psi[p(t, \omega) + u(t, \omega) - L(t, \omega)] \leq \overline{F}(t, \omega)
\]

Here

\[
\overline{F} = f^o + \Psi[p^o - L^o]
\]

Assuming that the linearization is sufficiently precise, and due to (4), we can postulate that

\[
0 \leq u(t, \omega) \leq l(t, \omega)
\]

\[\text{Losses are assumed to be accounted for in the demand vector } \mathbf{L}.\]
and $e^T = (1 \ 1 \ \ldots \ 1)$. The load shedding limits in (11) are assumed to be known, albeit time and scenario dependent, parameters bounded by the total demand

$$l(t, \omega) \leq L(t, \omega) \quad (12)$$

### 2.2. Representation of a Service Interruption

For each realization of random variable $\omega$ the bulk power system is capable of providing non-interrupted services to all its elements if and only if a system of equations and inequalities (5)-(11) have a feasible solution such that $u = 0$. If such a solution does not exist, a service interruption in the system is unavoidable.

A service interruption could be detected by finding a set of resource injections and load shedding minimizing the total volume of unserved energy

$$\min_{t, \omega} UE(t, \omega) = e^T u(t, \omega) \quad (13)$$

subject to constraints (7)-(11). Service interruption will be unavoidable if and only if the minimum value of the objective function (13) is positive. The solution to this optimization problem could be called a reliability dispatch. While the reliability dispatch is not uniquely defined, especially in the absence of service interruption, locational dual variables could be uniquely determined

$$\mu UE(t, \omega) = \frac{\partial UE(t, \omega)}{\partial L} \quad (14)$$

These variables represent Locational Marginal Unserved Energy. By averaging equation (14) across all stochastic scenarios, we define a vector of Marginal Expected Unserved Energy (MEUE) as

$$MEUE(t) = \frac{\partial EUE(t)}{\partial L} \quad (15)$$

where

$$EUE(t) = E[UE(t, \omega)] = \sum_{\omega=1}^{\omega=\omega} \pi(\omega) UE(t, \omega) \quad (16)$$

$$MEUE(t) = E[\mu UE(t, \omega)]$$

Equations (14)-(15) deserve additional discussion. Let us assume that under all stochastic scenarios, no transmission constraint binds. Under this assumption, $\mu UE(t, \omega)$ will be the same for every bus in the power system, $\mu UE(t, \omega) = 0$ for all scenarios in which there is no service interruption

and $\mu UE(t, \omega) = e$ (increasing demand by 1 MW will increase the level of unserved energy by exactly 1 MW) for all scenarios with service interruption. From this, it is obvious, that in the absence of transmission constraints the Marginal Expected Unserved Energy at all locations will be equal to the probability to experience a service interruption, otherwise known as a Loss of Load Probability (LOLP). In other words, in the absence of transmission constraints, equation (15) becomes a well-known formula [8]:

$$LOLP(t) = \frac{\partial EUE(t)}{\partial L} \quad (17)$$

(In this case the derivative is taken by the total demand such that we have scalar variables in both sides of the equation). However, in a system with transmission constraints, equation (17) does not hold.

In section 3.6 we consider a more general formulation of reliability dispatch. As demonstrated in the second part of this paper, indicator MEUE is directly related to what we call a Locational Stochastic Reliability Price (LSRP) playing an important role in making optimal system expansion decisions and therefore should be considered an adequate generalization of the concept of LOLP for constrained systems.

### 3. The Optimal Capacity Selection Problem

We assume that the selection of generation capacity is conducted through an auction in which the Auctioneer in order to maintain an economically justified reliable operation of the power system solicits offers from existing and prospective generation owners and demand response service providers commonly referred to as resource providers.

#### 3.1. Resources

For each unit of offered resource capacity the following information is provided to the Auctioneer:

- Capacity location on the electrical grid (bus number $j$)
- Maximum capacity offered $Y_j$ (in MW)
- Per unit capacity payment required $c_j$

In addition, for each offered resource the Auctioneer obtains statistical parameters upon which it can develop availability and low bound operational limitation scenarios. These scenarios are combined in matrices $\bar{S}(t, \omega), \underline{S}(t, \omega)$. The Auctioneer selects a
mix of capacities that could be deployed over the entire planning horizon \(0 \leq t \leq T\).

For the purpose of this paper, we make the following simplifying assumptions:

- All offered resources could be placed into service at the beginning of the planning horizon;
- The Auctioneer is not required to select the entire size of the offered capacity \(Y_j\) and can choose any portion of it.

### 3.2. Economic Model of Load Shedding

In the simplest case we assume that for each location, the Auctioneer has an estimate of the instantaneous value of loss load (VOLL) which is assumed to remain constant over time, but could be different for different locations. The vector of VOLL values for all locations is denoted as \(V\). For example, locational VOLL values could be set administratively either uniformly for the entire system or vary by location based on particular economic and policy considerations.

The model of load shedding could be further expanded by assigning levels of VOLL increasing with the depth of service interruption at a given location. In this case values of VOLL are specified as a step function: load reduction at node \(j\) up to \(l_j^1\) is valued at \(V_j^1\), reduction between \(l_j^1\) and \(l_j^2\) is valued at \(V_j^2\) where \(l_j^1 < l_j^2\) and \(V_j^1 < V_j^2\) and so on. Such step-function might be represented by creating multiple nodes attached to the same bus, each with its own demand and VOLL. Similar models would emerge in the event of multiple load serving entities, each with its own demand profile and VOLL but connected to the same bus.

### 3.3. Auctioneer’s Optimization Problem

The Auctioneer’s objective function is to minimize the total cost of maintaining the economically justified reliable operation of the power system. This objective functions consists of two parts – the cost of procured mix of resources and the expected cost of unserved load.

\[
R = c^T X + \mathbb{E}[VUE] 
\]  

Here \(\mathbb{E}[VUE]\) represents the expected cost of unserved load equal to

\[
\mathbb{E}[VUE] = \sum_{\omega=1}^{N} \pi(\omega) \int_0^T V^T u(t, \omega) dt 
\]

The Auctioneer’s goal is

\[
\min R = c^T X + \sum_{\omega=1}^{N} \pi(\omega) \int_0^T V^T u(t, \omega) dt 
\]

The Auctioneer determines the optimal resource mix subject to the following five groups of constraints.

1) Maintaining the energy balance in each hour under each stochastic scenario, subject to potential service interruptions:

\[
c^T (p(t, \omega) + u(t, \omega) - L(t, \omega)) = 0 \quad \lambda(t, \omega) > 0 
\]

2) Maintaining the security of the transmission system:

\[
\bar{F}(t, \omega) \leq \Psi(t, \omega) \left[ p(t, \omega) + u(t, \omega) - L(t, \omega) \right] \\
\leq \bar{F}(t, \omega) \quad \lambda(t, \omega) \leq 0, \bar{F}(t, \omega) \geq 0 
\]

3) Operating each resource within the limits of capacity procured at the auction and subject to generator’s availability and low bound operational limitations:

\[
S(t, \omega) X \leq p(t, \omega) \leq \bar{S}(t, \omega) X \\
\lambda - \alpha(t, \omega) \leq 0, \bar{\alpha}(t, \omega) \geq 0 \quad \lambda(t, \omega) > 0 
\]

4) Unserved energy is non-negative and cannot exceed allowed load shedding limits:

\[
0 \leq u(t, \omega) \leq l(t, \omega) \quad \lambda - \bar{\theta}(t, \omega) \leq 0, \bar{\theta}(t, \omega) \geq 0 
\]

5) Procured capacity does not exceed capacity offered:

\[
0 \leq X \leq Y \quad \lambda - \beta \leq 0, \bar{\beta} \geq 0 
\]

### 3.5. Optimal Capacity Selection

**Proposition 1.** Capacity \(X_j\) will be:

selected at full offered volume \((X_j = Y_j)\) if and only if

\[
c_j < \sum_{\omega=1}^{N} \pi(\omega) \int_0^T \left[ S_j(t, \omega) \bar{\alpha}_j(t, \omega) - S_U(t, \omega) \bar{\alpha}_j(t, \omega) \right] dt 
\]

rejected \((X_j = 0)\) if and only if

\[
c_j > \sum_{\omega=1}^{N} \pi(\omega) \int_0^T \left[ S_j(t, \omega) \bar{\alpha}_j(t, \omega) - S_U(t, \omega) \bar{\alpha}_j(t, \omega) \right] dt 
\]

---

3. Theoretically it is conceivable, that the vector of VOLL values could be specified by Load Serving Entities (LSEs). However, this assumption requires further theoretical consideration which is beyond the scope of this paper.

4. In most cases for the purpose of the capacity procurement auction low bound operational limitations could be ignored: water on a hydro plant could be spilled, wind can be curtailed and excess thermal generator simply would not be committed. We, however, include these low bound constraints for the sake of generality.
accepted partially \((0 \leq X_j \leq Y_j)\) if and only if

\[ c_j = \sum_{\omega=1}^{N} \pi(\omega) \int_{0}^{T} \left[ S_j(t,\omega) \lambda(\omega) - \bar{S}_j(t,\omega) \alpha_j(t,\omega) \right] dt \tag{28} \]

The proof is provided in the Appendix.

Based on the complementarity properties (26)-(28), it is logical to define the resource capacity price (RCP) as

\[ RCP_j = \sum_{\omega=1}^{N} \pi(\omega) \int_{0}^{T} \left[ \bar{S}_j(t,\omega) \alpha_j(t,\omega) - S_j(t,\omega) \lambda(\omega) \right] dt \tag{29} \]

Based on the above definition, the resource capacity price should be applied to the total capacity of the resource procured at the auction.

### 3.6. Generalized Reliability Dispatch and Locational Stochastic Reliability Price

For any given set of selected capacities, generalized reliability dispatch for a given stochastic scenario is the dispatch which minimizes the cost of unserved load. At a given moment in time on each stochastic scenario, the objective function of reliability dispatch is to minimize

\[ \min V^T u(t,\omega) \tag{30} \]

subject to constraints (21)-(24).

This problem is a generalization of reliability dispatch defined in section 2.2. Optimal dispatch yields an optimal locational price \(-\) a reduction in the total cost of unserved energy in response to the marginal reduction in locational demand. The underlying mathematics of this locational price is identical to that of the Locational Marginal Price (LMP) resulting from the optimal economic dispatch. We will call the corresponding value of the locational price the Locational Stochastic Reliability Price (LSRP). This term reflects that the price is locational, that it corresponds to the reliability dispatch and that it is stochastic because it depends not only on time and location, but also on the stochastic scenario of system operation \(\omega\).

Similarly to the definition of LMPs, the vector of LSRPs is determined as

\[ LSRP(t,\omega) = \bar{\lambda}(t,\omega)e^{-\Psi(t,\omega)(\bar{\Pi}(t,\omega)-\mu(t,\omega))} \tag{31} \]

Using this definition, we establish the following relationship between LSRPs and Resource Capacity Price.

**Proposition 2.**

\[ RCP = \sum_{\omega=1}^{N} \pi(\omega) \int_{0}^{T} \left[ \bar{S}(t,\omega) \max \left[ 0, LSRP(t,\omega) \right] \right] dt \tag{32} \]

or, more generally:

\[ RCP = \int_{0}^{T} E\left[ \bar{S}(t,\omega) \max \left[ 0, LSRP(t,\omega) \right] \right] dt \tag{33} \]

Proof is provided in the Appendix.

The expression \( E\left[ \bar{S}(t,\omega) \max \left[ 0, LSRP(t,\omega) \right] \right] \) could be interpreted as the value of the option to use one unit of generating resource when it is needed. The expression \( E\left[ \bar{S}(t,\omega) \max \left[ 0, -LSRP(t,\omega) \right] \right] \) could be interpreted as the cost of the obligation to use one unit of the generating resource when it is not needed.

Formula (33) therefore determines resource capacity price as the difference between the total over time value of option to use the resource when it is needed and the total over time cost of obligation to use the resource when it is not needed. According to this formula, a resource is compensated for its contribution to system reliability only to the extent it is available at the time of need: for example if at the time of need (i.e. LSRP at resource’s location is positive) it is available in 50\% of cases, it will be compensated only by 50\% of the LSRP. If due to a low bound operational limitation the resource has to be used at the time when it is not needed (i.e. LSRP at resource’s location is negative) the resource will be negatively affecting system reliability by increasing the overall value of unserved energy and will be charged for that.

### 3.7. LSRPs and Nodal Economics of Load Shedding

Next we explore the locational economic properties of LSRPs. Because in the reliability dispatch the cost of each generator is effectively zero, in absence of energy shortage, i.e. if \( u(t,\omega) = 0 \), the LSRPs are equal to zero at all locations. Indeed, because of the zero price of marginal generation, the cost of congestion redispatch is zero and the system lambda must also equal zero. In contrast to LMPs, whose properties are primarily driven by the economics of nodal generation dispatch, the properties of LSRPs are primarily driven by the economics of energy rationing at times of shortage.
Thus, only the instances of shortages are of interest here.

Assuming that a supply shortage is observed at node \( j \) on scenario \( \omega \) at time moment \( t \), but the load reduction is properly within bounds, i.e. \( 0 < u_j(t,\omega) < l_j(t,\omega) \), then due to (61) both shadow prices \( \overrightarrow{\theta}_j(t,\omega) = 0 \) and \( \overleftarrow{\theta}_j(t,\omega) = 0 \). Combining this with (57) yields that \( LSRP_j(t,\omega) = V_j \)  

(34)

In other words, at the location of shortage where load is partially shed, the LSRP must equal VOLL. If the load at a given location has to be fully reduced \( (u_j(t,\omega) = l_j(t,\omega)) \), then the LSRP may exceed the locational VOLL \( LSRP_j(t,\omega) = V_j + \overrightarrow{\theta}_j(t,\omega) > V_j \)  

(35)

If the load is being shed at one location in the system, it will influence LSRPs at other locations, as follows from equation (31). In a special case when the shadow prices for all transmission constraints are equal to zero and therefore, as follows from (31), the LSRP will be the same at all locations. Under these conditions, load shedding will occur in order of increasing VOLL and the location with the highest VOLL among those where the load is being shed will set the LSRP for the entire system. Thus, in absence of transmission congestion \( LSRP_j(t,\omega) = \max_{f/u_j(t,\omega)>0} V_j \)  

(36)

An important case in which the value of lost load is assumed to be the same for all locations deserves special consideration. In this case equation (36) could be simplified \( LSRP_j(t,\omega) = VOLL \)  

(37)

In this case, the reliability dispatch problem is equivalent to earlier considered problem of minimizing the volume of unserved energy. Indeed, the value of unserved energy is simply the product of the volume of unserved energy and VOLL. Therefore, in this case LSRPs and \( \mu EUE \) are also proportional to each other: \( LSRP(t,\omega) = VOLL \times \mu EUE(t,\omega) \)  

(38)

In a more general case, when optimal reliability dispatch results in constrained transmission, values of LSRPs may significantly vary by location as illustrated by examples provided in section 5.

3.7. Reliability Implications of the Auction Outcome

Further assuming the single VOLL for the entire system, based on (38) resource capacity prices could be expressed as
\[
RCP_j = VOLL \times \left[ \mathbb{E} \left[ \max_{0, \mu EUE_j(t,\omega)} \right] dt \right] \\
- \mathbb{E} \left[ \max_{0, \mu EUE_j(t,\omega)} \right] dt \\
= VOLL \times \mu EUE_j \times EFORd_j
\]

(39)

Here, \( EFORd_j \) represents a generalized Equivalent Forced Outage Rate – demand (for definition of \( EFORd \) see for example, [10, 11]). The generalized \( EFORd_j \) is defined by the following formula
\[
EFORd_j = \mathbb{E} \left[ \max_{0, \mu EUE_j(t,\omega)} \right] dt \\
- \mathbb{E} \left[ \max_{0, \mu EUE_j(t,\omega)} \right] dt
\]

(40)

and thus reflects correlation between unavailability of the resource and the value of \( \mu EUE_j(t,\omega) \) at its location.

Auction clearing conditions (26) - (28) can now be expressed in reliability terms as the following

**Proposition 1a. Capacity \( X_j \) will be:**

selected at full offered volume \( (X_j = Y_j) \) if and only if
\[
MEUE_j > \frac{c_j}{VOLL \times EFORd_j}
\]

rejected \( (X_j = 0) \) if and only if
\[
MEUE_j < \frac{c_j}{VOLL \times EFORd_j}
\]

accepted partially \( (0 \leq X_j \leq Y_j) \) if and only if
\[
MEUE_j = \frac{c_j}{VOLL \times EFORd_j}
\]

This leads to an important conclusion that in the system bounded by transmission constraints, reliability criteria must be locational. The use of globally imposed reliability indicators, e.g. system-
wide LOLP in the form of chance constraints will lead to suboptimal investment decisions.

4. Financial Auction Outcome

In this section we consider the overall auction outcome not only in terms of prices but also in terms of resource receipts and load payments.

4.1. Resource Receipts

A resource at location $j$ receives the payment of

$$RR_j = X_j RCP$$

(44)

The sum of all resource receipts is equal to

$$RR_x = X^T RCP$$

(45)

4.2. Load Payments

At each load location $j$ load is assumed to be paying in accordance with the following formula:

$$LP_j = \sum_{\omega=1}^{N} \pi(\omega) \int_{0}^{T} [L_j(t, \omega) - u_j(t, \omega)] dt$$

(46)

or more generally

$$LP_j = E \left[ \int_{0}^{T} [L_j(t, \omega) - u_j(t, \omega)] dt \right]$$

(47)

and the total amount of payments collectable from all loads participating in the auction will be equal to

$$LP_x = E \left[ \int_{0}^{T} [L(t, \omega) - u(t, \omega)]^T LSRP(t, \omega) dt \right]$$

(48)

Let’s assume now that $D_j$ is the demand for location $j$ at the time of projected coincident system peak.

Using locational peak demands as a representation of load capacity requirements, we define Load Capacity Price (LCP) as

$$LCP_j = \frac{LP_j}{D_j}$$

(49)

$$= E \left[ \int_{0}^{T} LSRP_j(t, \omega) \left( \frac{L_j(t, \omega) - u_j(t, \omega)}{D_j} \right) dt \right]$$

and overall load capacity payments resulting from the auction could be expressed as

$$LP_x = D^T LCP$$

(50)

According to formula (49), payments to loads are assessed on the basis of demand realized in a given scenario. Even if two load serving entities receive power at the same location but have either different VOLLs or different demand patterns, they may see different capacity prices. Loads which are less interrupted (for example due to a higher VOLL) will be paying higher capacity prices than loads interrupted more often. If the level of VOLL at the load location is in agreement with load’s preference of being interrupted, its load is shed if and only if LSRP at load’s location equals VOLL and therefore interruption would not be objected by the load. If in some scenario, the load is interrupted fully, this scenario would not affect load’s capacity price at all. This formula provides a fair pricing structure for reliability – load pays for reliability only “at the time” when there are service interruptions and only to the extent it is not being interrupted. “At the time” here does not refer to real time of operation, but to the time period and scenario of simulated reliability dispatch.

This formula also reflects a correlation between LSRPs, demand and the level of interruption.

4.3. Congestion Rent

In each combination of time and stochastic scenario when load is being interrupted somewhere, LSRPs are either the same in all locations (in absence of transmission congestion) or vary by location (in the presence of binding transmission constraints). It is important to note that since physical reliability dispatch is often non-unique, the term “binding transmission constraint” is only meaningful in terms of dual variables, either $\mu(t, \omega) > 0$ or $\pi(t, \omega) > 0$.

**Proposition 3.** The difference between load payments and generator receipts is always non-negative and represents reliability congestion rent:

$$LP_x = RR_x + CR_x$$

$$D^T LCP = X^T RCP + CR_x$$

(51)

where

$$CR_x = E \left[ \int_{0}^{T} LSRP(t, \omega) - \mu(t, \omega) F(t, \omega) dt \right]$$

(52)

Components of congestion rent are always non-negative

$$CR(t, \omega) = \Pi^T(t, \omega) F(t, \omega) - \mu^T(t, \omega) F(t, \omega) \geq 0$$

and therefore

$$CR_x \geq 0$$

(53)

(54)

Proof in provided in the Appendix.

Along with (54), this identity proves that load payments in the auction are always in excess of, or equal to, resource receipts and the outcome of the auction will never leave the Auctioneer revenue deficient.
5. Examples

5.1. Examples of LSRP Calculation

All presented examples are developed for a three-bus network graphically depicted on Figure 1.

![Figure 1. An Example three-node Network](image)

The system depicted on the above figure is short on resources – for 540 MW total demand there is only 520 MW of total online capacity. In the ensuing examples we consider this system under different conditions in terms of locational values of VOLL, transmission constraints and low bound operational limitations on resources and observe how these conditions impact reliability dispatch and values of LSRPs.

![Example 1](image)

In Example 1 depicted on Figure 2 constrained transmission line increases actual load shedding to 35 MW and sets different LSRPs at all buses. Marginal load shedding of 35 MW takes place at bus B where LSRP equals VOLL of $10,000. Marginal resource is located at bus C where LSRP is equal to zero. LSRP at bus A is equal to $5,000. A 1 MW demand increase at bus A must come from 2 sources: 0.5 MW from generator at bus C and 0.5 MW additional load shedding at bus B (in order to keep flow on line C within limits). This results in an incremental VUE of $5,000. This example clearly demonstrates the nodal network impact on LSRPs and ultimately on RCPs. A shortage event like that renders generator G3 at bus C having zero value in its ability to resolve system deficit, while the generator G2 located at the bus where the load is being shed is twice more valuable than the generator G1 located at bus A.

Example 2 depicted on Figure 3 differs from Example 1 in one additional detail: load reduction at bus B is limited by 30 MW, which is 5 MW less than the optimal load shedding at that bus in Example 1. Due to constraint along line C-B, it is not enough to simply shed 5 MW at bus A, but in order to maintain system security load at A has to be shed by 10 MW. This results in LSRP at bus A rising to the level of VOLL, while LSRP at bus B is rising to twice the VOLL level! Load shedding limitation at bus B therefore doubles LSRPs at all locations (LSRP at bus A remains at zero).

![Example 2](image)

Example 3 depicted on Figure 4 differs from Examples 1 and 2 in two important elements. First, the flow on line C-B is now limited to mere 40 MW and at the same time the cost of load shedding at bus B beyond 30 MW is set at three times the VOLL. The result of imposing these two conditions is that now the overall demand reduction in the system increases to 140 MW, the generator at bus C is forced to shut down. LSRPs in this case are $10,000 at bus A, $30,000 at bus B and -$10,000 at bus C. Although the LSRP at bus C is negative, this will not impose a negative impact on RCP for that generator, because in this case no low bound operational limit is imposed on this resource.
Example 4 depicted on Figure 5 is different from Example 3 only in one instance – a 10 MW low operating bound imposed on Generator G3 at bus C. The LSRPs here are the same as in Example 3. The overall load reduction is actually 10 MW less than in Example 3: 20 MW less is shed at bus A but 10 MW more load is shed at bus B to maintain the security of line C-B and resulting in $10,000 higher value of unserved energy than in Example 3. But an important distinction of this example is that it produces a negative impact on RCP of generator G3. The per unit low operating bound for this resource is non-zero and this will reduce resource’s RCP in accordance with formula (33).

Figure 5. Example 4

5.2. Example of the Auction Outcome

Let us consider a stylized example of the auction outcome and assume that the solution of the auction problem results in full selection of the three generators considered and in exactly four instances of demand rationing represented by the above four examples considered. Each instance of demand rationing lasts one hour and has a probability of occurrence of 0.1/year. Computed RCPs for each generator is indicative of the maximum price each generator would have to offer at the auction in order to be fully selected. Reliability dispatch results under these scenarios are summarized in Table 1 below. Note that generator capacities shown in the first line of that table are above the generator capacity depicted in the examples. This discrepancy is reflected in the generator availabilities shown in this table. The auction outcome is summarized in Table 2. Due to transmission congestion, LSRPs on average significantly vary by location. Resource capacity prices differ from average LSRPs due to limitations on resource availability. Moreover, due to a low bound operational limitation of generator G3 at bus C, this resource receives overall negative capacity price. Load capacity prices also differ from average LSRPs and from resource capacity prices both at the same location and across locations (no load capacity price is defined for Bus C, because there is no load at this bus). Load payments exceed generator receipts due to congestion rent attributable to the constrained line C-B.

6. Conclusions and Further Research

The proposed capacity market auction based on the stochastic engine effectively combines in a single method the resource adequacy assessment and optimal capacity selection thus eliminating the need for a more traditional two-stage approach. Due to computational complexity, it is yet remains to be seen whether such an auction could be realized in practice in full nodal implementation. However, a simpler zonal implementation appears feasible.

The existing practice based on a system-wide LOLP-type resource adequacy criteria of 1-day-in-10-years combined with ad-hoc approaches for determining locational installed capacity requirements is neither optimal, nor supported by sound economic analysis.

The developed theory sheds the light on the locational nature of resource adequacy requirements. It also helps to define the process for determining locational (zonal) installed capacity requirements driven by reliability indicators aligned with capacity expansion optimality conditions. This is presently being implemented within the resource adequacy methodology developed for the Russian Non-profit Partnership "Council for Organizing Efficient System of Trading at Wholesale and Retail Electricity and Capacity Market" (NP "Market Council") [13].
In future research the authors intend to explore the computational feasibility of the nodal implementation of the defined stochastic auction. Further theoretical research is needed to better address the reliability tradeoff of generation and transmission expansion following ideas outlined in [14].

Table 1. Summary of Reliability Dispatch

<table>
<thead>
<tr>
<th></th>
<th>P</th>
<th>Bus A</th>
<th>Bus B</th>
<th>Bus C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generator capacities</td>
<td>240</td>
<td>220</td>
<td>130</td>
<td></td>
</tr>
<tr>
<td>Load capacity requirements</td>
<td>170</td>
<td>370</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>LRSPs by Scenario</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Example 1</td>
<td>0.1</td>
<td>5,000</td>
<td>10,000</td>
<td>0</td>
</tr>
<tr>
<td>Example 2</td>
<td>0.1</td>
<td>10,000</td>
<td>20,000</td>
<td>0</td>
</tr>
<tr>
<td>Example 3</td>
<td>0.1</td>
<td>10,000</td>
<td>30,000</td>
<td>(10,000)</td>
</tr>
<tr>
<td>Example 4</td>
<td>0.1</td>
<td>10,000</td>
<td>30,000</td>
<td>(10,000)</td>
</tr>
<tr>
<td>Loads Servable by Scenario</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Example 1</td>
<td>0.1</td>
<td>170</td>
<td>335</td>
<td>NA</td>
</tr>
<tr>
<td>Example 2</td>
<td>0.1</td>
<td>160</td>
<td>340</td>
<td>NA</td>
</tr>
<tr>
<td>Example 3</td>
<td>0.1</td>
<td>80</td>
<td>320</td>
<td>NA</td>
</tr>
<tr>
<td>Example 4</td>
<td>0.1</td>
<td>100</td>
<td>310</td>
<td>NA</td>
</tr>
<tr>
<td>Generator Availabilities by Scenario</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Example 1</td>
<td>0.1</td>
<td>83%</td>
<td>91%</td>
<td>92%</td>
</tr>
<tr>
<td>Example 2</td>
<td>0.1</td>
<td>83%</td>
<td>91%</td>
<td>92%</td>
</tr>
<tr>
<td>Example 3</td>
<td>0.1</td>
<td>83%</td>
<td>91%</td>
<td>92%</td>
</tr>
<tr>
<td>Example 4</td>
<td>0.1</td>
<td>83%</td>
<td>91%</td>
<td>92%</td>
</tr>
<tr>
<td>Generator low bound limitations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Example 1</td>
<td>0.1</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Example 2</td>
<td>0.1</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Example 3</td>
<td>0.1</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Example 4</td>
<td>0.1</td>
<td>0.0%</td>
<td>0.0%</td>
<td>7.7%</td>
</tr>
</tbody>
</table>

Table 2. Auction Outcome

<table>
<thead>
<tr>
<th></th>
<th>Bus A</th>
<th>Bus B</th>
<th>Bus C</th>
<th>System</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean LSRP ($/MW-yr)</td>
<td>3,500</td>
<td>9,000</td>
<td>(2,000)</td>
<td></td>
</tr>
<tr>
<td>LCP ($/MW-yr)</td>
<td>2,500</td>
<td>7,851</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RCP ($/MW-yr)</td>
<td>2,917</td>
<td>8,182</td>
<td>(76.92)</td>
<td></td>
</tr>
<tr>
<td>L.P. ($)</td>
<td>425,000</td>
<td>2,905,000</td>
<td>-</td>
<td>3,330,000</td>
</tr>
<tr>
<td>RR ($)</td>
<td>700,000</td>
<td>1,800,000</td>
<td>(10,000)</td>
<td>2,490,000</td>
</tr>
<tr>
<td>CR ($)</td>
<td></td>
<td></td>
<td></td>
<td>840,000</td>
</tr>
</tbody>
</table>

7. References


APPENDIX

Proof of Proposition 1.

Define the Lagrangian \( \Lambda \) of the Auctioneer’s problem

\[
(20) - (25).
\]

\[
\Lambda = c^T X + \sum_{t=1}^{N} \pi(t, \omega) \int_{0}^{T} V u(t, \omega) dt
\]

\[
- \sum_{t=1}^{N} \pi(t, \omega) \int_{0}^{T} \lambda(t, \omega) e^T (p(t, \omega) + u(t, \omega) - L(t, \omega)) dt
\]

\[
+ \sum_{t=1}^{N} \pi(t, \omega) \int_{0}^{T} \sigma^T(t, \omega) [\Psi^T(p(t, \omega) + u(t, \omega) - L(t, \omega)) - \bar{F}(t, \omega)] dt
\]

\[
- \sum_{t=1}^{N} \pi(t, \omega) \int_{0}^{T} \mu^T(t, \omega) [\Psi^T(p(t, \omega) + u(t, \omega) - L(t, \omega)) - \bar{F}(t, \omega)] dt
\]

\[
+ \sum_{t=1}^{N} \pi(t, \omega) \int_{0}^{T} \sigma^T(t, \omega) [\Psi^T(p(t, \omega) - \bar{S}(t, \omega) X)] dt
\]

\[
- \sum_{t=1}^{N} \pi(t, \omega) \int_{0}^{T} \mu^T(t, \omega) [\Psi^T(p(t, \omega) - \bar{S}(t, \omega) X)] dt
\]

\[
+ \sum_{t=1}^{N} \pi(t, \omega) \int_{0}^{T} \sigma^T(t, \omega) (u(t, \omega) - l(t, \omega)) dt
\]

\[
- \sum_{t=1}^{N} \pi(t, \omega) \int_{0}^{T} \mu^T(t, \omega) (u(t, \omega) - l(t, \omega)) dt
\]

\[
+ \beta^T (X - Y) - \beta^T X
\]

and state KKT optimality conditions

\[
\sum_{t=1}^{N} \pi(t, \omega) \int_{0}^{T} [\bar{S}(t, \omega) \bar{\pi}(t, \omega) - \bar{\pi}^T(t, \omega) \bar{S}(t, \omega)] dt - \beta + \beta
\]

\[
- \left[ \bar{\pi}(t, \omega) - \theta(t, \omega) \right]
\]

\[
\lambda(t, \omega) e - \Psi^T (\bar{\pi}(t, \omega) - \mu(t, \omega))
\]

\[
\lambda(t, \omega) e - \Psi^T (\bar{\pi}(t, \omega) - \mu(t, \omega)) = (\bar{\pi}(t, \omega) - \mu(t, \omega)) (58)
\]

\[
[p(t, \omega) + u(t, \omega) - L(t, \omega)]^T \Psi^T - \bar{F}^T(t, \omega) \bar{\pi}(t, \omega) = 0
\]

\[
- [p(t, \omega) + u(t, \omega) - L(t, \omega)]^T \Psi^T - \bar{F}^T(t, \omega) \mu(t, \omega) = 0
\]

\[
[p^T(t, \omega) - X^T \bar{S}(t, \omega)] \bar{\pi}(t, \omega) = 0
\]

\[
-p^T(t, \omega) - X^T \bar{S}(t, \omega) \mu(t, \omega) = 0
\]

\[
\bar{\pi}^T(t, \omega)(u(t, \omega) - l(t, \omega)) = 0
\]

\[
\bar{\pi}^T(t, \omega)u(t, \omega) = 0
\]

\[
(X^T - Y^T) \beta = 0
\]

\[
- X^T \beta = 0
\]

The proof of Proposition 1 immediately follows from (56) and (62)

Proof of Proposition 2.

From (58) and (60) we derive that if \( \bar{\pi}(t, \omega) > 0 \) then \( \bar{\pi}(t, \omega) = 0 \), resource \( j \) is dispatched at full level of available procured capacity \( p_j(t, \omega) = X_j \bar{S}_j(t, \omega) \) and \( \bar{\pi}(t, \omega) = \text{LSRP}_j(t, \omega) \).

Similarly, if \( \bar{\pi}(t, \omega) > 0 \) then \( \bar{\pi}(t, \omega) = 0 \), the resource is forced to operate at the lowest point \( p_j(t, \omega) = X_j \bar{S}_j(t, \omega) \) and \( -\bar{\pi}(t, \omega) = \text{LSRP}_j(t, \omega) \).

Finally, if both \( \bar{\pi}(t, \omega) = 0 \) and \( \bar{\pi}(t, \omega) = 0 \), then the resource \( j \) is dispatched as marginal and the LSPR at that resource’s location equals zero. Based on this discussion, we can conclude that

\[
\bar{\pi}(t, \omega) = \max \{0, \text{LSRP}_j(t, \omega)\}
\]

(63)

\[
\bar{\pi}(t, \omega) = \max \{0, -\text{LSRP}_j(t, \omega)\}
\]

(64)

Indeed, if \( \text{LSRP}_j(t, \omega) > 0 \) it is equal to \( \bar{\pi}(t, \omega) \). If \( \text{LSRP}_j(t, \omega) = 0 \), then both \( \bar{\pi}(t, \omega) = 0 \) and \( \bar{\pi}(t, \omega) = 0 \), and if \( \text{LSRP}_j(t, \omega) < 0 \) then \( \bar{\pi}(t, \omega) = 0 \) and \( \bar{\pi}(t, \omega) = -\text{LSRP}_j(t, \omega) \) and therefore (63) and (64) always hold.

Proof of Proposition 3.

First, assume that transmission constraints do not bind (\( \Theta(t, \omega) - \mu(t, \omega) = 0 \)). Since the system is in balance in the sense of equation (21) and since all LSRPs are the same, the following identity holds:

\[
\left[ L(t, \omega) - u(t, \omega) \right]^T \text{LSRP}(t, \omega)
\]

(65)

\[
= p^T(t, \omega) \text{LSRP}(t, \omega)
\]

However, it is not difficult to demonstrate that

\[
-p^T(t, \omega) \text{LSRP}(t, \omega)
\]

(66)

\[
\bar{\pi}^T(t, \omega) \text{LSRP}(t, \omega)
\]

Indeed, in all instances in which injection of resource \( j \) is within the open interval \( X_j \bar{S}_j(t, \omega) < p_j(t, \omega) < X_j \bar{S}_j(t, \omega) \) stochastic reliability
price equals zero, \( LSRP_j(t, \omega) = 0 \) and identity (66) holds. If \( X_j S_j(t, \omega) = p_j(t, \omega) \), then \( LSRP_j(t, \omega) < 0 \) and
\[
-\max\left[0, -LSRP_j(t, \omega)\right] = LSRP_j(t, \omega).
\]
Similarly, if \( p_j(t, \omega) = X_j S_j(t, \omega) \), then \( LSRP_j(t, \omega) > 0 \) and
\[
\max\left[0, LSRP_j(t, \omega)\right] = LSRP_j(t, \omega).
\]
This proves the identity (66) for all cases.

By combining (65) and (66), we obtain that
\[
\left[ L(t, \omega) - u(t, \omega) \right]^T LSRP(t, \omega)
= X^T \left[ S(t, \omega) \max\left[0, LSRP(t, \omega)\right] \right]
\]

Identity (67) indicates that in the absence of binding transmission constraints total load payments are exactly equal resource receipts.

If some transmission constraints bind, equation (65) does not hold, but instead the following is true:
\[
\left[ L(t, \omega) - u(t, \omega) \right]^T LSRP(t, \omega) - p(t, \omega) LSRP(t, \omega)
= \left[ L(t, \omega) - u(t, \omega) - p(t, \omega) \right]^T \lambda(t, \omega) e
- \Psi^T(t, \omega)(\bar{p}(t, \omega) - \mu(t, \omega))
\]
\[
= \left[ L(t, \omega) - u(t, \omega) - p(t, \omega) \right]^T \lambda(t, \omega) e
- \left[ L(t, \omega) - u(t, \omega) - p(t, \omega) \right]^T \Psi^T(t, \omega)(\bar{p}(t, \omega) - \mu(t, \omega))
\]
\[
= \left[ \bar{p}(t, \omega) - \mu(t, \omega) \right]^T \Psi(t, \omega) \left[ p(t, \omega) - (L(t, \omega) - u(t, \omega)) \right]
\]
\[
= \left[ \bar{p}(t, \omega) - \mu(t, \omega) \right]^T f(t, \omega)
\]
\[
= \bar{F}(t, \omega) \mu(t, \omega) \bar{F}(t, \omega) - \mu(t, \omega) \bar{F}(t, \omega)
\]

Since shadow prices for binding constraints are always non-negative and due to condition (10), we conclude that \( \bar{F}(t, \omega) \bar{F}(t, \omega) - \mu(t, \omega) \bar{F}(t, \omega) \) is always non-negative and will be positive if and only if the reliability dispatch results in binding transmission constraints. This value represents a non-negative stochastic congestion rent in time period \( t \) realized in a given scenario \( \omega \). This value represents a non-negative stochastic congestion rent in time period \( t \) realized in a given scenario.

\[
CR(t, \omega) = \bar{F}(t, \omega) \mu(t, \omega) \bar{F}(t, \omega) - \mu(t, \omega) \bar{F}(t, \omega) \geq 0 \quad (68)
\]

Congestion rent arising in the outcome of the auction will therefore equal to
\[
CR_{\text{c}} = \mathbb{E} \left[ \int_0^T \bar{F}(t, \omega) \mu(t, \omega) \bar{F}(t, \omega) - \mu(t, \omega) \bar{F}(t, \omega) \, dt \right] \quad (69)
\]

and because of (68) congestion rent is always non-negative.