Open Transmission Access: An Efficient Minimal Role for the ISO

Fernando L. Alvarado, Robert J. Camfield* and Rajesh Rajaraman*

Department of Electrical and Computer Engineering,
University of Wisconsin-Madison,
Madison, WI 53706
Phone: (608) 262-8900

* Laurits R. Christensen Associates, Inc.
Economics and Engineering Consultants
4610 University Avenue
Madison, WI 53705-2164
USA
Phone: (608) 231-2266; Fax: (608) 231-1365
email: rajesh@lrca.com

Abstract
This paper presents a vision of an efficient minimal role for an independent system operator (ISO) in a competitive market for electric power. The ISO framework could be achieved during the final stages of the transition to the deregulated environment. The chief responsibility of the ISO will be to maintain system reliability at all times. To achieve this end, ISO only requires information about the expected status of the system and scheduled transactions. There is no need for the ISO to engage in transactions in the bulk power market except for comparatively small amounts of power required to maintain reliability. The ISO will coordinate the actions of various market participants using technically sound rules, classified according to time frames. These rules include procedures by which the ISO can estimate congestion costs and allow markets to efficiently ration scarce flow capability. The paper also touches upon the ISO’s key role in the ancillary services market. Finally, the paper discusses the regulated transmission company, and concludes that ISO-determined congestion costs should be collected by the transmission company.

1. Introduction

This paper presents a vision of an efficient minimal role of independent system operator (ISO) in competitive power markets. Specific procedural mechanisms and modus operandi of the ISO are presented. While we advocate a minimalist ISO role in a fully competitive market, we recognize that to carry out a smooth transition from the regulated to the competitive market environment, the ISO may have to play a more active role, e.g., by initially operating under a Poolco setup [10,11]. As the electric services industry becomes more competitive, the ISO could become correspondingly less active.

The paper sets forth boundaries of minimal intervention for an ISO, thus facilitating the maximum independence among all market participants while also addressing the concerns of system security. The role of the ISO could be classified according to three broad time-frames: the real time operating time scale, the ex-ante planning stage in various possible time scales, and the ex-post stage. The paper discusses in detail the ISO’s role in each of these stages.

A minimalist role of the ISO can be achieved by mandated means (such as the method advocated in [19]), or by the use of efficient transmission pricing mechanisms, as advocated in this paper. Fundamentally, both approaches rely on an estimate of the binding constraints in the system, and both approaches characterize the operational boundaries of the system in a simple manner for ease of use by buyers and sellers of power. However, [19] relies on a pre-determined curtailment scheme to be used by the ISO to force power schedules to remain within the operational boundary. On the other hand, the present paper imposes an ex-ante congestion surcharge, determined in a transparent manner, to achieve the same aim.
efficient system operation, the congestion revenues collected by the ISO should be applied toward the expenses of the transmission network, including expansion costs.

The paper is divided into six sections. Section 2 discusses the complex, highly integrated nature of power systems. Section 3 discusses our vision of an efficient transmission access model that could be achieved in the final stages of market restructuring. Section 4 describes how the proposed transmission access model would work in practice. Section 5 discusses congestion pricing issues and the activities of the regulated transmission company. Section 6 concludes the paper.

2. Power system operation and constraints

We summarize the essential requirements for continued, reliable operation of power systems and the questions that they pose for market restructuring plans.

2.1. Energy balance must be maintained

For reliable operation, power systems must satisfy energy balance requirements in near real-time periods. There are several reasons why it is difficult for participants, by themselves, to maintain energy balance without the help of a central coordinator:

- The practical operation of power systems implies transmission losses. These losses are unique to location and are nonlinear with regard to the load level and reactive power consumption. The system losses depend on the system state, or current operating point which cannot be known or estimated with only the limited information available to individual agents.
- Load varies in a stochastic manner and thus cannot be forecasted precisely. While an electric system cannot maintain energy balance exactly, it can maintain energy balance within an acceptably narrow range. Maintenance of this balance within these limits, however, requires central coordination.
- The likelihood of constraints and congestion within the transmission network may require that the system be organized in some rational fashion into areas where the power balance can be maintained. This requirement is embodied in present systems in the form of control areas.

The size of an interconnected system has a dramatic effect on system frequency. In a large interconnected system, frequency can be maintained within very tight much more pronounced. Thus, frequency alone is not always the best indicator of power balance. In order to have a more universal method for the control of power balance, it is necessary to organize the system into control areas, where the responsibility for regulating frequency is split among the various control areas by assuring that each area maintain a tight balance between load and generation.

The fact that in the past the control areas have often been along company boundaries has also made it possible to use the notion of the control area boundary for the purpose of negotiating and establishing transactions among parties. In the future, it may be necessary to separate the control areas’ responsibility to provide load balance and frequency regulation services, from bulk power transactions. That is to say, there is no sustained justification for restricting transactions to be transactions solely among control areas.

2.2. The operating point must be feasible at all times

A feasible operating point is a level of supply and demand that lies within the limits, or boundary, of the system capabilities. Even if an operating point is energy balanced, the transmission network may limit the amount of electricity services that can be transferred among various parties. This limit can be described as the Operation Limit Boundary (OLB) [3].

Within any power system, operational constraints arise because of generator limits, transmission constraints caused by line overloads, voltage collapse problems, transient stability limits, etc. The OLB can be conveniently viewed as a union of several feasibility constraint surfaces in a multi-dimensional space. Each axis represents either a contract transaction or the demand for electric power. If an operating point lies within the OLB, it can be physically realized, and is said to be feasible. Operating points projected to lie outside the OLB cannot be physically realized and must be curtailed to positions within the OLB.

Figure 1 illustrates the concept of the OLB. The OLB is composed of the intersection of several constraints. The horizontal and vertical axes can represent the electric power demand of two loads A and B, or two bulk power transactions/contracts (in general, the “dimension” of this “space” will be much greater than two; there will be as many dimensions as there are independent transactions or demands).

---

1 The system capabilities can be described as: 1) the quantity of real and reactive power that can be supplied; 2) the limits of transport of power to loads; and 3) and speed with which the system can respond to changes in system conditions.
This figure demonstrates that the feasibility of an operating point depends upon the activities of both A and B. For instance, operating point “a” is feasible and is allowed by the OLB. Operating point “b,” however, lies outside the OLB and is infeasible. There are several ways to “ration” operating point “b” to a feasible position within the OLB.

In practice, it is also essential to employ a conservative OLB in order to allow for demand uncertainty in the operating point for ex ante periods. Moreover, contingency events affect the position of the OLB. Specifically, outages of lines and/or generators result in a new OLB position. It may be appropriate to define an OLB so that the system survive one or more “credible” contingencies. Whatever the criteria for its determination, an OLB exists and the power system must operate within its confines at all times.

2.3. Ancillary services

Besides real energy balance and feasibility, the operation of power systems requires the regulation of voltage by means of reactive power support, the availability of sufficient operating reserves, the ability to regulate frequency, and several other “services” besides bulk power. These services have been called “ancillary services” [1,9,15]. Supply of these ancillary services is closely integrated with the maintenance of energy balance and the feasibility of operating points.

We discuss the provision of reactive power in some detail. Reactive power is rarely a commodity of direct interest to users, but without it the provision of active real power is impossible insofar as the transmission network is inherently inductive. The effect of reactive power is to provide sufficient voltage for current flows in networks. The need for reactive power in a power system arises for two distinct reasons [1]:

- power factors less than unity. This requirement arises from the reactive power consumption that can be directly observed at the end-user’s site.
- The maintenance of voltage and thus the real power flow within the transmission network requires reactive power. Unlike real power, a key feature of reactive power is that it does not travel very far. Thus, reactive power sources, including capacitor and reactor technologies, must be dispersed throughout the system.

In addition to the above two requirements, it is important to recognize the implications of rapidly changing (dynamic) power loads. Reactive power support plays a major role in providing system security in light of very fast changing loads. For example, arc furnaces need dynamic reactive power support, often supplied by static VAR compensators (SVC), synchronous condensors and generators designed for rapidly changing loads.

Ancillary services also include operating reserves. Reserves are insurance, or options, to protect against contingency events. Operating reserves can be differentiated by the speed of response to contingency events. In near-real time, reserves are also differentiated by type of response: load-following (regulation) reserves continuously respond to small forecasted variations in system supply and load, and spinning reserves are continuously available to respond to larger variations and unexpected contingency events. Adequate amounts of all types of reserves are needed for reliable power system operation.

3. A transmission access model

3.1. Market participants

As a premise for later discussions, the list of market participants implicit in most discussions regarding market restructuring is summarized as follows:

- A regulated independent system operator (ISO), unaffiliated with consumers or suppliers that assumes two major responsibilities:
  1. Coordinate the real-time operation of the energy marketplace, including maintaining reliable operation, and possibly arbitrate contract disputes. The ISO also contracts for the minimal amount of energy services in the competitive market for reliable system operation.
  2. Charge efficient transmission pricing to all power exchange transactions.
1. Maintenance of lines.
2. Expanding the network appropriately (the ISO, power exchange, and transmission owners will operate under the auspices of fairly strict protocols assembled by regulators and/or a consortium of interests of parties to the power market).

- Distribution companies that convert power from the transmission network level to the end-user level.
- Electric power suppliers that sell and buy services including real power service, reactive power support, and operating reserves. The suppliers are unaffiliated with the ISO (and transmission companies).
- Energy exchanges where electric products and bundled services are bought and sold. The exchanges will likely facilitate both primary and secondary packages of services, and will be organized into spot, futures and forward markets instruments [17].
- Loads or end-users that buy or sell (e.g., interruptible service) power services.
- Merchants such as brokers or power marketers who act as middlemen between buyers and sellers of individual products or bundled services.

3.2. Contracts for electric services

While the full extent of the prospective contractual arrangements for power services cannot be known, a view of the general direction can be anticipated. Contracts, of course, can contain any package of services, provisions, and arrangements for compensation. However, the specialized provisions of forward contracts, which are currently more common, limit transferability. The loss of generality in provisions inhibits their secondary market participation. Generally, we can expect that while custom forward contracting will remain, more fluid futures market instruments will become prevalent and be actively traded on exchanges.

Merchants have clear incentives to actively pursue the development of standardized bundles of electric products to be sold as futures. As with other commodities, bundled electric service futures will be standardized for delivery at specific locations within transmission networks, such as buses or zones, at various points of time. For an excellent discussion about pricing unbundling electric energy services, see [12].

needed

The organization and packaging of electric products, however, need not be a major concern of the ISO. The ISO will operate under the auspices of a governance structure delineated by a strict protocol that defines the ISO’s activities and how such activities are carried out.

These delimited activities will be guided by the global elements or dimensions of restructured markets. The goals of restructuring initiatives may embrace the following elements as far as the ISO is concerned:

- Charge transactions efficient prices for transmission services
- Contract for the appropriate services for system reliability.

In order to carry out activities that meet these broadly defined requirements, however, the information and knowledge required by the ISO is rather modest, and include the following components:

- Expected transaction schedule quantities -- i.e., power injections and withdrawals and their locations. Knowledge of schedules allows the ISO to develop projections of likely flows along lines of the network. Transaction schedules associated with contracts will be registered with the ISO (almost certainly in an automated manner).
- Near real-time information regarding contemporary system conditions, as gathered through state-estimation techniques operated within the energy management systems.

These two items along with a knowledge of the network parameters are sufficient for the ISO to perform its duties.

3.4. The dimension of time

The activities of the ISO can be organized into ex-ante periods, real-time or contemporary periods (current system operations), and ex-post periods.

1. Ex ante Periods. Contracts for ex-ante (future) services” will take place among contracting parties for

---

2 There are some services that the ISO must provide in real time operation for reliable system operation. However this does not preclude other participants from offering these services themselves. For example, the ISO must provide load following service on a system wide level, since in practice there will not be perfect match between load and supply at all times. Suppose that a market for reliable load following services develops and that all the contracts purchase this capability from vendors of the load following service. Then the load following that the ISO provides at the system wide level becomes minimal. Hence the ISO’s costs of providing the system wide load following service become minimal, and the participants pay only a minimal price to the ISO for this service.
These contract schedules (not the pricing information) are registered with the ISO, and provides the ISO with knowledge of the major components of the expected flows of power system within each prospective period. Based on this information, the ISO will notify each contract its share of system losses, and starts a market clearing process for transmission constraints. This may result in alterations in existing contracts and/or registration of new contract schedules. The ISO will also contract for the appropriate minimal services to maintain system reliability in real time operation (e.g., if all parties have already made arrangements for reliable load-following service, the ISO need only purchase a minimal amount of load following service).

This process may be viewed as a coordinated feedback process between the ISO and the participants to help maintain system reliability. This stage will end when it is no longer possible to involve all the participants in the coordination process. The exact rules and time-frames that govern operative actions of the ISO and behavior of participants will be established beforehand. The rules, -- i.e., rules of the road -- will be largely determined by the marketplace, as it evolves, as well as the technology and information processes used to maintaining contemporary system security. Institutional safeguards must be built in to prevent participants from gaining unfair competitive advantage by using undesirable gaming behavior.

2. **Real-time period.** The contemporary stage ("present") starts when transaction schedules can no longer be affected via price signals. In this stage, transactions are frozen and effective control to alter flows is given to the ISO, who is unilaterally managing system security. Of course, in this stage, participants can physically execute contracts involving real time operation (e.g., load following).

3. **Ex post period.** This third stage ("the past") provides participants with an opportunity to settle contracts after the fact. Participants to transactions will be charged a transmission fees for use of transmission facilities. Transmission fees will be based on congestion prices and usage which are determined exactly as specified in the protocol under which the ISO operates, their scheduled contract quantities and actual usage of the system capabilities. The ISO’s actions during the ex-ante and real-time stage will also be subject to occasional ex-post external audits.

---

4 Examples of contracts are: power delivered from one supplier to one or more consumers, operating reserves provided by a merchant to a municipality, static VAR provided by a marketer to a wind tunnel, etc.

---

4 Generally, the better the ISO is able to anticipate market behavior, the smaller will the number of market clearing iterations be, and the less volatile and more stable will the market clearing process be.
planned energy imbalance in contracts is kept to a practical minimum.

2. Real time operating stage. As appropriate, the ISO will operate one or more marginal units or “slack buses,” with Automatic Generation Control (AGC) features [5,18]. Energy balance is automatically maintained by tracking small variations in frequency and area control error (ACE) in the various areas of the system. If at any time net load is slightly less than net generation, the ACE increases, and the AGC serves to offset this imbalance by decreasing generation. Conversely, if net load becomes slightly more than total generation, the frequency decreases, and the AGC increases generation to restore the frequency to its normal value. Thus, generating units under AGC provide instantaneous operating reserves, often referred to as load following.

AGC serves to correct minor variations in energy imbalances. However, significant energy imbalances in contracts may arise as a result of large load variations, or due to contingencies such as the failure of a generation unit or a transmission line. The ISO will also have available appropriate amounts of “spinning” operating reserves (we discuss operating reserves in the Ancillary Services section of this paper).

3. Ex-post stage Because of the random nature of events such as load variations and unexpected contingencies, there may be need for participants to make adjustments to contracts in an ex-post manner. The settlements for losses and energy imbalance between the parties within a contract will be based upon departures between actual services provided and amounts contracted for. These settlements will be based on the power flows in the power system as measured by the ISO.

4.2. Monitoring contracts for feasibility

It is assumed that all contracts are “energy balanced.” This includes the losses associated with each contract. The roles of the ISO in assuring feasibility of all contracts differ depending on the time-frame.

1. Planning Stage for a prospective period The ISO must coordinate the proposed actions of various participants to ensure feasibility of contracts. To help solve this problem, the ISO notices participants of expected congestion cost. The participants can be expected to reduce their load (transactions) in response to the high prices. The price response is efficient insofar as those loads which respond the most are those that forego the least consumer value or surplus in reducing their loads. The steps of this market clearing process are as follows:

feasible, the ISO allows it. If the proposed operating point is infeasible, the ISO computes and communicates relevant information about the infeasible operating point to all the parties and starts a market clearing process to ensure feasibility [16]. (If an operating point is infeasible as a result of a single contract transaction only, the ISO could curtail that contract sufficiently to achieve a feasible operating point, and a market clearing process is not necessary. In practice however, the reason for an infeasible operating point can be attributed to several contract schedules, cf., Fig. 1.)

- For each transmission constraint violation, the ISO computes and makes available knowledge to all participants: the normal vector \( \mathbf{n} \), evaluated at a point on the OLB which is “closest” to the proposed infeasible operating point; and a measure of violation of the constraint (the “distance” [7] between the proposed infeasible operating point and the OLB). The vector \( \mathbf{n} \) can also be viewed as a first order sensitivity vector, where \( n_i \) is the incremental increase in violation of the constraint in response to an incremental load/generator increase by the \( i \)th participant [19].

- For each constraint violation, the ISO imposes a congestion cost \( \lambda m_k \) on the \( k \)th contract, where \( m_k \) is the incremental increase in violation of the constraint in response to a small, incremental transaction increase by the \( k \)th contract, and \( \lambda \) is the market price for the constraint. \( \lambda \) will be updated (increased) by the ISO regularly until it reaches the market clearing level and the constraint clears\(^5\). Based upon the response of the participants, both \( \mathbf{n} \) and the measure of violation of the constraint will also be updated in the process. The market clearing process may be done electronically in an automated fashion.

- The combination of the ISO’s congestion pricing approach and information about the security constraints, as communicated to the participants, will induce the participants to restructure contract

\( ^5 \) When computing \( \mathbf{n} \), it is necessary to make the assumption of a slack bus (or distributed slack). See [7,8] for computing the vector \( \mathbf{n} \) for voltage collapse or oscillatory instabilities, and see [18] for computing the vector \( \mathbf{n} \) for the case of line overloads. The sensitivity of a contract on the constraint can be obtained by algebraically adding up each contract participant’s effect on the constraint. It is possible that \( m_k < 0 \) which means that the contract \( k \) has a beneficial effect on the security constraint (e.g., a “must run” generating unit). This implies that contract \( k \) receives a congestion payment. For a practical system the net congestion payment by the participants is expected to be positive.
security constraint, perhaps because of lack of time to respond, the ISO will be forced to ration contracts in the real-time operating stage.

Many aspects of power system operation, such as load variation, are stochastic. For this reason, the ISO will impose congestion costs based on the expectation that congestion will occur (e.g., whenever the particular set of contracts is “close” to security constraints). In the case of soft security constraints -- e.g., line flow constraints on low voltage transmission lines -- the violation can occur for short periods of time.

2. Real time operating stage. The market may not, however, clear a security constraint in the planning stage. In such instances, the “final round” of updated congestion costs imposed by the ISO will be communicated to the participants in the scheduled planning stage. The ISO would have to curtail contracts in the real-time operating stage, based on established and pre-determined criteria, in order to maintain feasibility.

The curtailment criteria could possibly be based on each contract schedule’s effect on the security constraint, and also on the advanced notice time given to the ISO by the contract. Depending on the rules agreed upon, a contract schedule that registered late in the planning stage with the ISO may be given less priority, and hence curtailed more, than a contract that provided the ISO more notice. An alternative option is that contract schedules that registered late in the planning stage are charged a higher transaction cost than those contract schedules that registered earlier; then the curtailment criteria would be each contract schedule’s effect on the system constraint.

3. Ex-post stage Based on actual observed power flows, the ISO computes the congestion costs incurred by all contracts. The congestion cost incurred by contract k is $\sum \lambda_j m^k_j x_k$ (over all constraints j), where $\lambda_j$ is the market clearing (or shadow) price for constraint j, $m^k_j$ is the sensitivity of contract k to constraint j, and $x_k$ is the contract transaction. See the section “Pricing Issues and Role of the Transmission Company” for a discussion of how revenues from congestion charges can be “allocated.”

4.3. Ancillary services

1. Planning stage for a prospective period

Operating reserves. The amount of instantaneous operating reserves required for a prospective period will depend on a number of factors, such as size of load, probable contingency events, weather, degree of network congestion, registered contract schedules (including perhaps contracts for operating reserves among participants), historical deviation from schedule, etc. Industry reliability standards, like NERC guidelines, and other factors, such as historical reliability of the system, will determine the required amount of the instantaneous (AGC and spinning) reserves. The ISO will also similarly determine the size of the non-instantaneous (but sufficiently short term) operating reserves. The ISO will contract in the competitive market for the appropriate amount of operating reserves (see also footnote 2).

Non-instantaneous operating reserves consist of supply side reserves such as generators that are currently idle, but may come on line on short notice, and demand-side reserves such as interruptible load contracts. The ISO need not buy longer term operating reserves since there will be sufficient time to curtail those transactions that cause the need for the use of these reserves, if the transactions themselves have not arranged for the longer term reserves.

The costs of having the non-instantaneous and spinning operating reserves available are not fixed. Different time periods and different system locations may have different costs of reserves. One way of recovering the availability costs of the operating reserves is to charge each contract with an “insurance premium” based upon criteria such as size of contract, historical reliability of the supply, and historical deviation from its planned schedule. For instance, a contract which deviates frequently from its planned schedule or needs non-instantaneous reserves called in frequently, will require resources, and would pay a high premium. The pricing of operating reserves will be presented in a separate publication.

Reactive power support. Based on the registered contract schedules and the nature of the loads, the ISO determines the amount of reactive power support required (Static VAR compensator, switched capacitors, shunt devices, etc.) so that the contract transactions can be executed reliably. This information is then communicated to the participants. To the extent that the reactive power requirements can be attributed to individual contracts, the contracts will be responsible for the necessary reactive power dispatch. Provision of the needed reactive power at the required locations and within the required time frames can be done by the ISO and charged to the individual transacting parties. To the extent that there is a “market” for reactive power [6,13], the participants could arrange for

---

6 The market clearing process of a transmission constraint is conducted with complete transparency of the congestion prices. The sensitivity vector helps participants to restructure contracts as follows. Participants with $n^k=0$ may find it beneficial to enter into contracts with participants whose $n^k<0$ [19]. If no participants exist such that $n^k=0$ (a case not considered in [19]), then contracts will have to self-curtail.
arrange for their dispatch by buying the reactive power from the energy exchange or from the transmission company (e.g., switching of capacitor banks). The paper [2] shows one way of allocating system reactive power costs to individual contracts. The dispatch of reactive power must be coordinated with the participants, since reactive power changes affect losses and flows in the power system, and hence the contractual transactions.

2. Real time operating stage In this stage, the ISO ensures that the reactive power demands are instantaneously satisfied so that voltages stay within acceptable levels, instantaneous operating reserves come on line if they are required, etc.

3. Ex-post stage There will be a need for participants to settle contracts ex-post for reactive power dispatch. The ISO also computes the costs for any use of operating reserves, and keeps track of which contract(s), if any, were responsible. At a later stage, this would affect each contract’s insurance premiums for operating reserves (the costs of only those operating reserves must be computed for which the ISO has contracted; contract settlements for operating reserves not involving the ISO will of course be left to the contract participants).

5. Pricing issues and the role of the transmission company

In the previous sections, we have described a possible transmission access model. While most of the pricing issues can be settled by the invisible hand of the marketplace with minimal but important input from the ISO, we discuss two issues that need closer attention: congestion costs, and transmission pricing.

Consider the example in Figure 2a that illustrates a lossless power system that is constrained by a flow limit on the transmission line between east and west, and by the capacity limit of the generators in the east. G west, G east represent generators in the west and east respectively. Both G west and G east are assumed to be price takers, i.e., perfect competition exists in the regions where G west and G east operate. Therefore, when G east is not at a capacity limits, its price, $7/MW, is the marginal cost of producing power; at capacity limits, its price will be determined by the demand.

Figure 2b shows the marginal price curve for power delivery. The cost curve reflects the effect of these constraints on price. For simplicity, we ignore the prices related to ancillary services.

Assume that L east contracts with G west and G east to have power delivered to them at least cost. It can be shown that:
- If the system operates at a level less than 100 MW (point “a”), L east incurs no congestion costs.
- At “b”, L east incurs congestion costs of $200
- At “c”, L east incurs congestion costs of $400
- At “d”, L east incurs congestion costs of $800

Who Collects the Congestion Revenues? For reasons of efficiency and fairness, the ISO should collect and transfer the congestion revenues to the transmission company. It may not be intuitively obvious that power supplier G west should not collect congestion costs-- G west in fact is paid simply the price of $3/MW -- at “b”, “c” or “d,” while G east profits (by $80=20x4) as a result of scarce supply in the east. The reasoning is that suppliers should not benefit from the constraints of a regulated transmission system (the 100 MW line in the example above at “b”, “c” and “d”). Otherwise, suppliers get perverse incentives to locate in places that increases congestion, to the overall detriment of society. For example, letting G west collect congestion costs at “b”, “c” or “d” gives G west no incentive to relocate on the east to relieve the 100 MW constraint. Indeed, it sends the wrong signals to suppliers to locate on the west to lay a stake on the congestion revenues.

How does the transmission company use the collected congestion revenues? To answer this question, we digress to consider first how the transmission company will recover its investment and operating costs of the transmission facilities. We sketch an outline for analyzing costs and benefits for each transmission facility.
generation and transmission) will have an obligation to serve the best interests of society. This statement implies that the regulated company bears a responsibility of reliable and efficient operation of transmission services, including regular operation and maintenance.

Regulators will allow recovery of prudently incurred investment and operating costs of the transmission facilities based on many factors including the performance of the transmission company, rate of return promised to investors, intrinsic value of a transmission facility to the users, stranded investments on obsolete transmission technologies, etc. The criteria for determining the recovery of costs for transmission facilities is beyond the scope of this paper. For the purposes of this discussion, however, we simply assume that the regulators have identified the transmission costs that can be recovered for each facility.

To the extent that part of the transmission costs of a facility can be identified as infrastructure and access costs that affect a region/area as a whole, these costs will be borne by every user within the region/area as access charges. We call the remaining portion of the transmission costs to be recovered as usage transmission costs. The transmission congestion revenues corresponding to a particular facility should be applied as a premium tax towards meeting usage transmission costs for that facility. If the congestion revenues under-recover the costs, the difference should be allocated as a non-premium tax among the users in proportion to usage (usage during off-peak periods could be valued the same as usage during peak periods). Note that: (a) the premium tax represents a surcharge for using the facility during periods when the facility is constrained or in

---

7 It may be impractical to do cost recovery analysis for each transmission facility in the system. Rather, most transmission facilities within a region can be lumped together and considered as one facility. However, there may be a few facilities that are substantially more expensive than other facilities, and that benefit only a few users. Regulators permitting, the cost recovery analysis for these facilities could be done separately.

8 It may not always be possible to identify with any one particular facility, the congestion revenues received from certain security constraints, such as voltage collapse limits. One way of allocating these revenues among the facilities could be according to sensitivities of the various parameters of the facility to the constraint. For example, if transmission lines are parameterized by their reactance, the sensitivity of these parameters on the security constraint identifies each facility’s contribution to the security constraint, and hence its share of the corresponding congestion revenues.

9 Various techniques have been published for allocating “transmission usage” among the various users, including Postage Stamp, Contract Path, various versions of the MW-Mile method, etc. These methods do not take full account of the nature of power system flows. Recently [14,20,21] have proposed methods to identify transmission usage that take full account of the AC load flows.

are sent to users because only those participants who use a facility contribute towards the costs of the facilities.

If the congestion revenues for the facility over-recover costs for a facility, then the excess congestion revenues should be used towards transmission grid expansion (thus reducing overall expansion costs) if the benefits of expansion outweigh its costs (see the next paragraph for a discussion on cost-benefit analysis of grid expansions). If the transmission system cannot be expanded, the excess revenue must be somehow returned to the participants. Note that there is no point in returning to the users their share of congestion costs because it defeats the purpose of having a market clearing process for a constraint in the first place! Instead, the over-collected revenues must be used to reward participants if they improve congestion in that facility by changing their behavior, e.g., offering credits for use of the system during periods when the facility is less congested. The simultaneous combination of rewards and market clearing for constraining the use of a facility is a carrot-and-stick approach. This would result in a more uniform and efficient utilization of the facility.

The transmission company must consider the costs and benefits of grid expansion at all times (especially during periods of over-recovery of costs). When calculating the costs and benefits, many other factors must be considered. These include the possibility of new generators planning to enter the market, possible improvement or deterioration in other aspects of system operation, etc. For instance, in the example of Figure 2, if a generator whose marginal cost is $3/MW locates in the east, the congestion costs at “b”, “c” and “d” become zero; consequently the transmission company will not over-recover its costs on the 100 MW transmission line. In such a case, the transmission company may not upgrade the transmission network, unless there are other benefits in doing so. It is also possible that upgrading a facility at one location may create other problems, leaving other users worse off. For example, adding new transmission lines to a system may actually increase congestion in other lines. As another example, in Figure 2, if the capacity of the 100 MW transmission line is extended significantly, the generators in the east may be driven out of business. Ultimately, the decision to expand the transmission grid must be made in cooperation with all the participants and with the approval of the regulators.

6. Conclusions

This paper presents a vision of an efficient minimal role for an independent system operator (ISO) in competitive power markets. The ISO framework could be achieved during the final stages of the transition to the deregulated
system operations at all times. The ISO would not participate directly in the markets for bulk power. The ISO must be unaffiliated with any supply or consumer entity. Furthermore, the ISO must be directly regulated or subject to a strict operations protocol imposed by a consortium of network interests. To achieve operational reliability, the ISO coordinates the various contract transactions by following technically sound rules of the road. The rules may be classified according to time stages: the real time operating time scale ("the present"), the ex-ante planning stage ("the future") in various possible time scales, and the ex-post stage ("the past"). The paper has discussed the ISO's role in great detail for each of the three stages. Of these, the most important is the role of the ISO to determine congestion charges to be imposed ex-ante on transactions in order to help the marketplace attain a feasible operating point.

The transmission grid will be managed by a regulated transmission company. This company will have an obligation to serve the best interests of society. For optimal societal benefits, the transmission company should collect the congestion revenues. These revenues should be applied towards the costs (including any expansion costs) of the transmission network. If the congestion revenues under-recover the costs, the difference will be allocated to the users based on their usage of the facilities. If the congestion revenues over-recover the costs, the transmission company could use the over-collected revenues to either expand the transmission grid to reduce congestion (if benefits exceed costs of expansion), or to reward participants if they improve system congestion. The transmission company will explore transmission expansion options by performing cost-benefits analysis.

Acknowledgment
The paper has greatly benefited from discussions with Laurence Kirsch of Laurits R. Christensen Associates. This paper does not necessarily represent the views of Laurits R. Christensen Associates. The authors remain solely responsible for any errors and omissions in this paper.

References