MIT EL 01-003 WP

Energy Laboratory

Massachusetts Institute of Technology

Price-Cap Regulation for Transmission: Objectives and Tariffs

January 2001

Price-Cap Regulation for Transmission: Objectives and Tariffs

Yong Yoon and Marija Ilic

Energy Laboratory Publication #
MIT EL 01-003 WP

Energy Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139-4307

January 2001

Price-Cap Regulation for Transmission: Objectives and Tariffs

Yong Yoon Marija Ilić

Energy Laboratory,

Massachusetts Institute of Technology, Cambridge, MA 02139

Abstract

In this paper we construct a mathematical metric for measuring the performance of the transmission provider (TP). The heart of the problem lies in developing the systemwide social welfare function which captures the unique role of the TP in the new industry environment where the electricity is provided through the market mechanism.

Following the formulation of the benchmark performance measure we describe two possible regulation schemes to be imposed on the TP, namely the rate-of-return regulation and the price-cap-regulation (PCR). The restructuring of the electric power industry is still a relatively recent event at the time of this writing, and there is yet to be a consensus on the actual implementation scheme for regulating the TP based on the guaranteed rate-of-return. In this paper, four of the more common implementation schemes are described and examined using the corresponding systemwide social welfare functions.

The PCR is proposed as a possible alternative regulation scheme to be imposed on the TP. Starting from one of the regulation schemes described under the rate-of-return regulation we develop the systemwide social welfare function associated with the PCR and show that the main difference between these two regulation schemes is not on the functional form of the systemwide social welfare but is on the party responsible for solving the optimization problem.

I. Introduction

In this paper we construct a mathematical metric for measuring the performance of the transmission provider (TP). The heart of the problem lies in developing the systemwide social welfare function which captures the unique role of the TP in the new industry environment where the electricity is provided through the market mechanism.

The paper is organized as follows: First, the benchmark performance measure is defined while accounting for the subtlety of functional unbundling in the electric power industry. This benchmark performance measure may be compared to the systemwide social welfare function for the omnipotent social planner, whose sole objective is maximizing the consumer utilities while minimizing various costs. The maximization of the benchmark performance yields the optimal level of the investment, the control effort and the maintenance effort into transmission. It is shown that under certain conditions, optimizing the benchmark performance leads to solving the optimization problem of the omnipotent social planner.

Following the formulation of the benchmark performance measure we describe two possible regulation schemes to be imposed on the TP, namely the rate-of-return regulation and the price-cap-regulation (PCR). The TP remains a monopoly through the restructuring process due to the assumption that there exists a high degree of economies of scale and economies of scope for the network. The main function of the TP is to provide adequate transmission capacity necessary for participants to trade electricity in the electric energy market.

Then, the systemwide social welfare function is developed this time with the rate-of-return regulation imposed on the TP. The restructuring of the electric power industry is still a relatively recent event at the time of this writing, and there is yet to be a consensus on the actual implementation scheme for regulating the TP based on the guaranteed rate-of-return. In this paper, four of the more common implementation schemes are described and examined using the corresponding systemwide social welfare functions.

It is shown that even though each scheme has a few distinct peculiarities that separate it from the others, they all suffer from the shortcomings similar to that of the rate-of-return regulation imposed on the vertically integrated utility, most notably the burden put on the regulator in eliciting the social welfare optimizing behavior from the regulated firm, for the case considered in this paper, the TP.

The PCR is proposed as a possible alternative regulation scheme to be imposed on the TP. Starting from one of the regulation schemes described under the rate-of-return regulation we develop the systemwide social welfare function associated with the PCR and show that the main difference between these two regulation schemes is not on the functional form of the systemwide social welfare but is on the party responsible for solving the optimization problem.

The concluding remarks are made at the end.

II. Benchmark performance measures for the transmission provider (TP)

Through the restructuring process the electricity is provided to the load by the generators through the market mechanism, and the vertically integrated utility is divided into generation, transmission and distribution/load sectors.

With this functional unbundling we assume that the actual utility functions of the loads and the actual cost functions (related to both operation and investment) of the generators within the electric power network become highly guarded private information and are only revealed in the form of demand functions and supply functions respectively through their overall market activities regardless of the actual market implementation of a particular region for energy.[1] We denote the demand and the supply functions as $D_{d_j}(Q_{d_j}[k],k)$ and $S_{g_i}(Q_{g_i}[k],k)$. On the other hand, the cost functions associated with the transmission provider, the cost of investment into transmission, $C_l^T(K_l^T[k],I_l^T[k],k)$, the control cost, $\upsilon_{tech}(e_{tech}[k],k)$ as a function of the control effort, $e_{tech}[k]$, and the maintenance cost, $\upsilon_m(e_m[k],k)$, as a function of the maintenance effort, $e_m[k]$, which are assumed to be available so that the only uncertainty associated with these functions are associated with the stochastic nature of future values.

By combining the demand and supply functions and the cost functions related to the transmission network, the systemwide social welfare function is given, in an association with the transmission provider, as the following:

$$SW_{TP}[k] = \sum_{d_j}\int_{\tilde{Q_{d_j}}[k]=0}^{Q_{d_j}[k]} D_{d_j}(\tilde{Q_{d_j}}[k],k)d\tilde{Q_{d_j}}[k] - \sum_{g_i}\int_{\tilde{Q_{g_i}}[k]=0}^{Q_{g_i}[k]} S_{g_i}(\tilde{Q_{g_i}}[k],k)d\tilde{Q_{g_i}}[k] - \sum_l C_l^T(K_l^T[k],I_l^T[k],k) \quad (1)$$

$$-\upsilon_{tech}(e_{tech}[k],k) - \upsilon_m(e_m[k],k)$$

[1]This is not to be confused with the demand and the supply functions that are required to be bidden to so-called spot markets in some regions in the U.S. that have gone through the restructuring process. The demand and the supply functions in this paper refer to what is revealed through various market activities by the participants and is highly correlated to either the marginal utility or the marginal cost functions under the perfect market assumptions.

The benchmark performance measure is then formulated as the problem of maximizing the systemwide social welfare function given in Eq. (1) and is given as the following:

$$\left[\overline{\mathbf{I_T}}^{\star}, \overline{\mathbf{e_{tech}}}^{\star}, \overline{\mathbf{e_m}}^{\star}, \overline{\mathbf{Q_G}}^{\star}, \overline{\mathbf{Q_D}}^{\star}\right]' = \lim_{\mathcal{T}\to\infty} \arg \max_{\substack{\mathbf{I_T}[k] e_{tech}[k],\, e_m[k] \\ \mathbf{Q_G}[k],\, \mathbf{Q_D}[k]}} \sum_{k=1}^{\mathcal{T}} (1-\xi)^k \mathcal{E}\left\{SW_{TP}[k]\right\} \tag{2}$$

subject to

$$I_l[k] \geq 0 \tag{3}$$

$$e_{tech}[k] \geq 0 \tag{4}$$

$$e_m[k] \geq 0 \tag{5}$$

$$\sum_{g_i} Q_{g_i}[k] = \sum_{d_j} Q_{d_j}[k] : \qquad \lambda[k] \tag{6}$$

$$Q_{g_i}^{\min}[k] \leq Q_{g_i}[k] \leq Q_{g_i}^{\max}[k] : \qquad \eta_{g_i}[k] \tag{7}$$

$$F_l(\mathbf{Q_G}[k], \mathbf{Q_D}[k]) \leq F_l^{\max}(\mathbf{F}[k], K_l[k], e_{tech}[k], e_m[k]) : \qquad \mu_l[k] \tag{8}$$

where

$$K_l^T[k+1] = (1-\rho_T) K_l^T[l] + I_l[k] \tag{9}$$

and $Q_{g_i}^{\min}[k]$ and $Q_{g_i}^{\max}[k]$ are the minimum and the maximum output by the generator g_i and are also revealed through their overall market activities.

It is worthwhile comparing the benchmark performance measure associated with the stand-alone TP with the optimization problem of an omnipotent social planner given as the following:

$$\left[\overline{\mathbf{I_G}}^{\star}, \overline{\mathbf{I_T}}^{\star}, \overline{\mathbf{e_{tech}}}^{\star}, \overline{\mathbf{e_m}}^{\star}, \overline{\mathbf{u_G}}^{\star}, \overline{\mathbf{Q_G}}^{\star}, \overline{\mathbf{Q_D}}^{\star}\right]' = \lim_{\mathcal{T}\to\infty} \arg \max_{\substack{\mathbf{I_G}[k], \mathbf{I_T}[k] e_{tech}[k],\, e_m[k] \\ \mathbf{u_G}[k],\, \mathbf{Q_G}[k],\, \mathbf{Q_D}[k]}} \sum_{k=1}^{\mathcal{T}} (1-\xi)^k \mathcal{E}\left\{SW_{system}[k]\right\} \tag{10}$$

where

$$\begin{aligned} SW_{system}[k] = {} & \sum_{d_j} U_{d_j}(Q_{d_j}[k], k) - \sum_{g_i} c_{g_i}(x_{g_i}[k], u_{g_i}[k], Q_{g_i}[k], k) - \sum_{g_i} C_{g_i}^G(K_{g_i}^G[k], I_{g_i}^G[k], k) \\ & - \sum_{l} C_l^T(K_l^T[k], I_l^T[k], k) - \upsilon_{tech}(e_{tech}[k], k) - \upsilon_m(e_m[k], k) \end{aligned} \tag{11}$$

subject to additional constraints defined as

$$K_{g_i}^G[k+1] = (1-\rho_G) K_{g_i}^G[k] + I_{g_i}[k] \tag{12}$$

$$u_{g_i}[k+1] = u_{g_i}[k] \text{ if } t_{g_i,dn} < x[k] < t_{g_i,up} \tag{13}$$

$$Q_{g_i}^{\min}[k] \leq Q_{g_i}[k] \leq K_{g_i}^{G}[k] : \qquad \eta_{g_i}[k] \tag{14}$$

The consumer utility functions and generation costs (both operation and investment) are denoted as U_{d_j}, c_{g_i}, and $C_{g_i}^{G}$. Typically, the operating cost is given as the following [1]:

$$c_{g_i}(x_{g_i}[k], u_{g_i}[k], Q_{g_i}[k], k) = u_{g_i}[k](c_{g_i,\phi}(Q_{g_i}[k]) + I(x_{g_i}[k] < 0)S_{g_i}) + (1 - u_{g_i}[k])(c_{g_i,f} + I(x_{g_i}[k] > 0)T_{g_i}) \tag{15}$$

$$c_{g_i,\phi}(Q_{g_i}[k]) = a_{g_i}Q_{g_i}^2[k] + b_{g_i}Q_{g_i}[k] + c_{g_i} \tag{16}$$

where

- $x_{g_i}[k]$: the status of the generator indicating the number of hours that the generator has been on at hour k
- $u_{g_i}[k]$: the decision to turn on or off the generator, g_i at each hour k; $u_{g_i}[k]$ is either 0 (off) or 1 (on)
- $c_{g_i,\phi}(Q_{g_i}[k])$: the total cost of generation, excluding capacity cost but including maintenance cost, at node g_i
- $c_{G_i,f}$: the fixed costs incurred during an hour where the generator is off
- $I(\cdot)$: a conditional statement has a value of 1 if the statement is true and 0 if it is false
- S_{g_i} and T_{g_i} : a startup cost and a shutdown cost respectively

For completeness we define

$$K_l^T[k+1] = (1 - \rho_T)K_l^T[k] + I_l[k] \tag{17}$$

corresponding to the optimization problem in Eq. (10), which is identical to the expression in Eq. (9).

One of the main differences in those optimization problems lies in the treatment of variables linked to the loads and the generators.

On one hand, as evident from Eq. (10) the utility functions of the loads and the cost functions of the generators are assumed to be available when required by an omnipotent social planner. The assumption of knowing these functions in details allows for an explicit determination of the optimal control variables related to the loads and the generators, including the investment into generation, the generation scheduling and unit commitment decisions. The result is the absolute optimality of the systemwide social welfare function.

On the other hand, as apparent from Eq. (2), the utility functions of the loads and the cost functions of the generators are assumed to be unavailable but can only be inferred from their market activities in the stand-alone TP environment. The result is the loss of the absolute optimality in its solution, but at the same time, the gain of the reduction in the computational complexity. The control variables derived from solving Eq. (2) guarantee the optimality of systemwide social welfare function up to the optimality of the market activities by the loads and the generators. In order to ensure that the unique solution[2] to Eq. (2) matches the unique solution to Eq. (10), three conditions need to be met, namely (1) the demand and the supply function should equal to the marginal utility and the marginal (operating) cost of the corresponding loads and generators, respectively, (2) the minimum and the maximum generation limits under the stand-alone TP scheme should be same as those under the omnipotent social planner scheme with the generator being on, or should be zero if the generator is off under the omnipotent social planner scheme, and (3) the initial capacities for each transmission line within the network should be the same under the stand-alone TP scheme and under the omnipotent social planner scheme. This is stated more formally in the following lemma.

Lemma 1: Suppose $\overline{u_{system}}^{\star}$ and $\overline{u_{TP}}^{\dagger}$ are the unique solutions to Eqs. (10) and (2) and the systemwide social welfare functions, $SW^{\star}_{system}[k]$ and $SW^{\dagger}_{TP}[k]$, are the corresponding results to the solutions, $\overline{u_{system}}^{\star}$ and $\overline{u_{TP}}^{\dagger}$, respectively.
Let $\overline{u_{system}}^{\star} = \left[\overline{\mathbf{I_G}}^{\star}, \overline{\mathbf{I_T}}^{\star}, \overline{\mathbf{e_{tech}}}^{\star}, \overline{\mathbf{e_m}}^{\star}, \overline{\mathbf{u_G}}^{\star}, \overline{\mathbf{Q_G}}^{\star}, \overline{\mathbf{Q_D}}^{\star}\right]'$ and $\overline{u_{TP}}^{\dagger} = \left[\overline{\mathbf{I_T}}^{\dagger}, \overline{\mathbf{e_{tech}}}^{\dagger}, \overline{\mathbf{e_m}}^{\dagger}, \overline{\mathbf{Q_G}}^{\dagger}, \overline{\mathbf{Q_D}}^{\dagger}\right]$.
Then,

$$\overline{\mathbf{I_T}}^{\star} = \overline{\mathbf{I_T}}^{\dagger} \tag{18}$$

$$\overline{\mathbf{e_{tech}}}^{\star} = \overline{\mathbf{e_{tech}}}^{\dagger} \tag{19}$$

$$\overline{\mathbf{e_m}}^{\star} = \overline{\mathbf{e_m}}^{\dagger} \tag{20}$$

$$\overline{\mathbf{Q_G}}^{\star} = \overline{\mathbf{Q_G}}^{\dagger} \tag{21}$$

$$\overline{\mathbf{Q_D}}^{\star} = \overline{\mathbf{Q_D}}^{\dagger} \tag{22}$$

if and only if

$$D_{d_j}(Q_{d_j}[k], k) = \frac{dU_{d_j}}{dQ_{d_j}[k]}(Q_{d_j}[k], k) \tag{23}$$

$$S_{g_i}(Q_{g_i}[k], k) = \frac{\partial c_{g_i}}{\partial Q_{g_i}[k]}(x_{g_i}[k], u_{g_i}[k], Q_{g_i}[k], k) \tag{24}$$

[2]The convexity of functions assumed within the optimization problem ensures the uniqueness condition of the solution and is extended in this paper.

$$Q_{g_i}^{\min}[k] \text{ (in Eq. (7)) } = \begin{cases} Q_{g_i}^{\min}[k] \Big(\text{ in Eq. (14) if } u_{g_i}^\star[k] = 1 \text{ from } \overline{u_{system}}^\star \Big) \\ 0 \text{ otherwise} \end{cases} \tag{25}$$

$$Q_{g_i}^{\max}[k] \text{ (in Eq. (7)) } = \begin{cases} K_{g_i}^{G}[k] \Big(\text{ in Eq. (14) if } u_{g_i}^\star[k] = 1 \text{ from } \overline{u_{system}}^\star \Big) \\ 0 \text{ otherwise} \end{cases} \tag{26}$$

$$K_l^T[0] \text{ (in Eq. (9)) } = K_l^T[0] \text{ (in Eq. (17))} \tag{27}$$

Proof: Suppose the conditions specified in Eqs. (23) through (27) hold true. Then, the control variables of the optimization problems in Eqs. (10) and (2) are subject to the same set of constraints since the constraints on $Q_{g_i}^\star[k]$ imposed by Eqs. (12) and (13) and Ineq. (14) puts the same constraints on $Q_{g_i}^\dagger[k]$ by Eqs. (25) and (26). Plus,

$$SW_{TP}^\dagger[k] = SW_{system}^\star[k] + Const. \tag{28}$$

due to Eqs. (23) and (24), where *Const.* is the constant cost related to the systemwide operating cost of generators. For example,

$$Const. = \sum_{\substack{g_i, \\ u_{g_i}[k]=0}} c_{g_i,f} + \sum_{\substack{g_i, \\ u_{g_i}[k]=0, \ \& \\ u_{g_i}[k-1]=1}} T_{g_i} + \sum_{\substack{g_i, \\ u_{g_i}[k]=1}} c_{g_i} + \sum_{\substack{g_i, \\ u_{g_i}[k]=1, \ \& \\ u_{g_i}[k-1]=0}} S_{g_i} \tag{29}$$

where $c_{g_i,f}$, T_{g_i}, S_{g_i} and c_{g_i} are as defined in Eqs. (15) and (16). Thus, the maximizing $SW_{TP}^\dagger[k]$ and $SW_{system}^\star[k]$ lead to the same set of solutions given that the initial conditions match. This constraint on the initial condition is enforced due to Eq. (27).

Suppose, on the other hand, the conditions specified in Eqs. (18) through (22) hold true for any convex functions of $U_{d_j}(Q_{d_j}[k], k)$ and $c_{g_i}(x_{g_i}[k], u_{g_i}[k], Q_{g_i}[k], k)$. Then, the control variables in the optimization problems in Eqs. (10) and (2) must be subject to the same set of constraints while the objective functions within the optimization must match up to the constant terms. In order to match the constraints for the optimization problem, only the constraints on $Q_{g_i}[k]$ need to be examined as the constraints on other control variables are already same due to the Ineqs. (3) through (6) and (8) through (9). If $u_{g_i}[k]^\star = 0$, then $Q_{g_i}^\star[k] = 0$ while if $u_{g_i}[k]^\star = 1$, then

$$Q_{g_i}^{\min} \leq Q_{g_i}^\star[k] \leq K_{g_i}^G[k] \tag{30}$$

This implies that the constraints for $Q^{\dagger}_{g_i}[k]$ must be defined as the following:

$$Q^{\min}_{g_i}[k] \text{ (in Eq. (7)) } = \begin{cases} Q^{\min}_{g_i}[k] \left(\text{ in Eq. (14) if } u^{\star}_{g_i}[k] = 1 \text{ from } \overline{u_{system}}^{\star} \right) \\ 0 \text{ otherwise} \end{cases} \tag{31}$$

and

$$Q^{\max}_{g_i}[k] \text{ (in Eq. (7)) } = \begin{cases} K^{G}_{g_i}[k] \left(\text{ in Eq. (14) if } u^{\star}_{g_i}[k] = 1 \text{ from } \overline{u_{system}}^{\star} \right) \\ 0 \text{ otherwise} \end{cases} \tag{32}$$

In order to match the objective functions within the optimization problems in Eqs. (10) and (2),

$$U_{d_j}(Q_{d_j}[k], k) = \int D_{d_j}(Q_{d_j}[k], k) dQ_{d_j}[k] + Const1. \tag{33}$$

$$c_{g_i}(x_{g_i}[k], u_{g_i}[k], Q_{g_i}[k], k) = \int S_{g_i}(Q_{g_i}[k], k) dQ_{g_i}[k] + Const2. \tag{34}$$

where $Const1$ and $Const2$ are some constants. Finally, by imposing the restriction that the initial conditions are the same:

$$K^{T}_{l}[0] \text{ (in Eq. (9)) } = K^{T}_{l}[0] \text{ (in Eq. (17))} \tag{35}$$

the optimization problems in Eq. (10) and (2) are equivalent as specified by the conditions in Eqs. (18) through (22). Thus, the constraints defined in Eqs. (31) through (35) must be satisfied, which are identical to the constraints in Eqs. (23) through (27). ■

As the energy portion of electricity is provided through the market mechanism, the decisions for unit commitment and investment into generation are determined through the decentralized optimization by individual generators. Suppose the unit commitment and investment decisions determined by the each supplier are same as the one determined by solving the centralized optimization problem in Eq. (10) [1]. Then, the conditions given in Eqs. (25) and (26) are satisfied. Plus, the conditions defined in Eqs. (23) and (24) are satisfied based on the assumption that utility and profit maximizing entities manage their consumptions and assets based on their marginal utility and the marginal cost functions respectively, under the perfect market assumptions. Invoking Lemma 1, we conclude that the market activities modeled by the optimization problem in Eq. (2) lead to the same optimal solution with respect to the systemwide social welfare, as solving the centralized optimization problem in

Eq. (10) while the difference in the computational complexity of Eq. (2) is orders of magnitude less than that of Eq. (10). In addition, suppose there is a minimum time scale to which the investment decision into transmission can be made, i.e., T_T. Then, by eliminating the complication of temporal interactions from the unit commitment the time scale separation between the planning and the operation is valid (and is exact) within $k = (n-1)T_T + 1$ and nT_T. The time scale separation does not quite hold true for Eq. (10). This reduction in the computational complexity allows the decentralized optimization by individual entities to achieve the systemwide optimum and is the one of the main advantages of introducing competition into the electric power industry.

We introduce the regulation imposed on the *TP* as a mechanism for inducing the operation and planning of an electric power network to be close to the systemwide optimal social welfare function. This implies that the energy portion of the electricity is being provided through the market mechanism while the transmission portion of it is being provided through the regulation. In the following sections various regulation schemes that may be imposed to the TP are discussed.

III. Role of regulation in providing conditions under which the efficiency of the overall network approaches the benchmark performance measure

After the restructuring process, the operation and the planning of an electric power network consist of four entities as shown in Figure 1. The transmission provider is a monopolistic entity whose responsibility is to design the transmission network and to operate the electric power system consisting of generation and transmission by virtue of controlling the allocation of the existing transmission capacity. The energy market is a generic term used to refer to a place for trading the energy portion of electricity (rather than limiting its use to refer only to the spot market where the centralized auctioning process takes place), and is composed of loads, generators and marketers. The loads are the consumers of various electric services (generation and transmission) while the generators are the suppliers of the energy portion of electric services. The marketers participate in trading of electric services often on behalf of loads or generators and typically do not own or operate generation, transmission or distribution systems. The function of marketers is largely ignored in this paper.

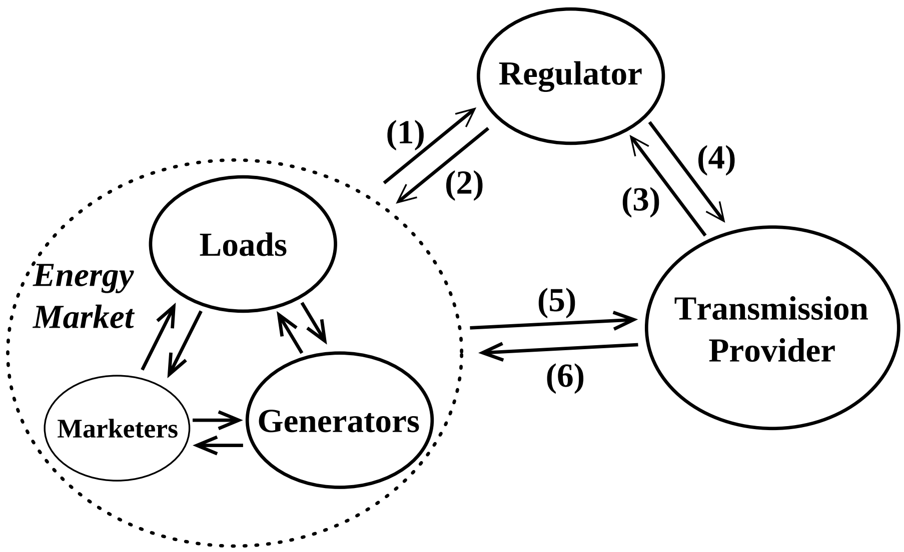

Fig. 1. Composition of the electric power network economics after the restructuring process

The regulator is typically a government agency whose responsibility is to oversee the operation and the planning of the network by the transmission provider directly and/or indirectly. The regulation by the regulator is necessary even after the restructuring process since the TP provider remains as a monopoly largely due to the economies of scale. As a monopoly the TP charges for the transmission portion of electric services above the marginal cost of network capacity so that the TP may continue to support the network as a viable business while ensuring a reasonable return on her investment. The regulation determines what the degree of reasonable return is and limits the TP from charging more than the reasonable. The rate-of-return regulation is one form of the cost-of-service regulations which guarantees the return on all of the investments that are made with an approval, up to the amount allowed by the regulator.

With the introduction of competition the function of the regulator may, at first glance, seem reduced in terms of the direct influence it imposes on the operation and the planning of a regional electric power network since the energy portion of the electricity is provided through the market mechanism. Only the transmission portion of electric service is under the direct control of the regulator through the rate approval. However, there is a significant expansion of the regulator's function in terms of the indirect control over the electric power network economics. This is due to that fact that the particular form of market mechanism governing the energy market is required to be approved by the regulator before implementation. The role of regulator is two fold, (1) designing the market mechanism for energy market (2) prescribing the rational rates for transmission capacity, so that the overall operation and planning of electric power network approaches the systemwide social welfare optimization described in Eq. (2).

In the following section we examine four different market designs under which the energy portion of electricity is provided and the *transmission capacity is allocated*. All of the four designs are based on two restrictions, namely the rate-of-return regulation and the existence of so-called spot market.[3]

A. *Effect of spot market and the rate-of-return regulation imposed on the transmission provider (TP)*

Spot market refers to the short-term market for a physical commodity, in this case electricity. In the spot market for electricity, the prices reflect the prices of power that is available to meet the near real-time demand, within a time scale of a day or just a few hours. For simplicity without the loss of generality we consider that the spot market is conducted on an hourly basis in order to match the demand and supply for electricity.

From the perspective of consumer, each load d_j chooses the optimal level of her consumption, $Q_{d_j}[k]$ at each hour k in the spot market based on the maximization function, often referred to as *competitive consumer surplus* function, given as the following:

$$Q^{\star}_{d_j}[k] = \arg\max_{Q_{d_j}[k]} \mathcal{E}\left\{\int_{\tilde{Q_{d_j}}[k]=0}^{Q_{d_j}[k]} D_{d_j}(\tilde{Q_{d_j}}[k],k)d\tilde{Q_{d_j}}[k] - \rho_{e,d_j}(\mathbf{Q_D}[k],\mathbf{Q_G}[k],k)\cdot Q_{d_j}[k] \right. \tag{36}$$

$$\left. -\hat{\rho}_{t,d_j}(\mathbf{Q_D}[k],\mathbf{Q_G}[k],k)\cdot Q_{d_j}[k]\right\}$$

where $\rho_{e,d_j}(\mathbf{Q_D}[k],\mathbf{Q_G}[k],k)$ and $\hat{\rho}_{t,d_j}(\mathbf{Q_D}[k],\mathbf{Q_G}[k],k)$ are the prices for the energy and transmission portions of electric services at load d_j, respectively.

Compared to the consumer surplus function defined under the vertically integrated utility structure, i.e.,

$$Q^{\star}_{d_j}[k] = \arg\max_{Q_{d_j}[k]} \mathcal{E}\left\{U_{d_j}(Q_{d_j}[k],k) - \rho\cdot Q_{d_j}[k]\right\} \tag{37}$$

[3]The restrictions of having the rate-of-return regulation and the existence of spot market is relaxed in the later sections.

there are three noticeable differences introduced in Eq. (36). The first is related to the separate charges for energy and transmission after the restructuring process. This is due to the functional unbundling of the vertically integrated utility. As the vertically integrated utility is unbundled from a sole electric service provider into generation and transmission, the charge associated with transmission portion of the services is separated from the energy portion so that the collected revenue from each charge goes to the respective provider. The second is related to the time dependence of each charge. The actual cost of meeting the systemwide load may vary hour-by-hour due to the changes in the demand and supply functions by the loads and the generators respectively. Under the vertically integrated utility structure, the price charged for electricity is typically an average of varying cost at each hour so that the actual fluctuation in costs is internalized. After the restructuring process, the change in cost at each hour needs to be made explicit since it is not possible to internalize this fluctuation among generators with different ownership and, at the same time, to educe the economic efficiency. The third is related to the output dependence of each charge. Because it may cost differently to provide different amounts of electricity, the price varies with respect to the production and consumption at each generator and at each load. For example, during the peak demand hours while the electricity usage is high, a number of more expensive generators may need to be utilized in order to meet the demand, thus raising the overall energy price.

The quantity dependent pricing for transmission capacity is of the particular importance[8]. On one hand, when the price for transmission capacity is set too low, some parts of the network may experience what is often referred to as transmission congestion at the peak demand hours. The electric power flow on the transmission lines are limited by the transfer capacity through the dispatch in generation and load due to the inability to direct the transfer of electricity through a particular path in the electric power network. The transmission congestion refers to the inability to dispatch additional generation from certain generators within the system due to transmission line limits. Mathematically, the transmission congestion on line l is expressed as the following:

$$F_l(\mathbf{Q_G}[k], \mathbf{Q_D}[k]) > F_l^{\max}(\mathbf{F}[k], K_l[k], e_{tech}[k], e_m[k]) \tag{38}$$

Thus, the prices for transmission capacities, $\hat{\rho}_{t,d_j}(\mathbf{Q_D}[k], \mathbf{Q_G}[k], k)$ and $\hat{\rho}_{t,g_i}(\mathbf{Q_D}[k], \mathbf{Q_G}[k], k)$

need to be chosen sufficiently to include the *penalty* for avoiding transmission congestion. On the other hand, when the price for transmission capacity is set too high, the network is under-utilized. Thus, the pricing of transmission, the congestion pricing, becomes significant in achieving economic efficiency while conforming to operational limit on power transfer through each transmission line.

Mirroring the formulation of the competitive consumer surplus function in Eq. (36), from the perspective of the supplier, each generator g_i chooses the optimal level of his production, $Q_{g_i}[k]$ at each hour k in the spot market based on the maximization function, often referred to as *competitive supplier surplus* function, given as the following[4]:

$$Q^{\star}_{g_i}[k] = \arg\max_{Q_{g_i}[k]} \mathcal{E}\left\{\rho_{e,g_i}(\mathbf{Q_D}[k],\mathbf{Q_G}[k],k)\cdot Q_{g_i}[k] - \hat{\rho}_{t,g_i}(\mathbf{Q_D}[k],\mathbf{Q_G}[k],k)\cdot Q_{g_i}[k] - \int_{\tilde{Q_{g_i}}[k]=0}^{Q_{g_i}[k]} S_{g_i}(\tilde{Q_{g_i}}[k],k)d\tilde{Q_{g_i}}[k]\right\} \tag{39}$$

where $\rho_{e,g_i}(\mathbf{Q_D}[k],\mathbf{Q_G}[k],k)$ and $\hat{\rho}_{t,g_i}(\mathbf{Q_D}[k],\mathbf{Q_G}[k],k)$ are the prices for the energy and transmission portions of electric services at generator g_i, respectively.

Since the energy portion of the electricity is provided through the market mechanism, under the perfect competition with free entry assumption, the corresponding price at each bus is identical throughout the network, i.e., $\rho_e[k] = \rho_{e,d_j}[k] = \rho_{e,g_i}[k]$. Then, the decentralized optimization by loads and generators in Eqs. (36) and (39) yield the same solution to the following optimization problem:

$$[\mathbf{Q_G}^{\star}[k],\mathbf{Q_D}^{\star}[k]]' = \arg\max_{\mathbf{Q_G}[k],\mathbf{Q_D}[k]} \mathcal{E}\left\{\sum_{d_j}\left(\int_{\tilde{Q_{d_j}}[k]=0}^{Q_{d_j}[k]} D_{d_j}(\tilde{Q_{d_j}}[k],k)d\tilde{Q_{d_j}}[k] - \rho_{t,d_j}(\mathbf{Q_D}[k],\mathbf{Q_G}[k],k)\cdot Q_{d_j}[k]\right) - \sum_{g_i}\left(\int_{\tilde{Q_{g_i}}[k]=0}^{Q_{g_i}[k]} S_{g_i}(\tilde{Q_{g_i}}[k],k)d\tilde{Q_{g_i}}[k] + \rho_{t,g_i}(\mathbf{Q_D}[k],\mathbf{Q_G}[k],k)\cdot Q_{g_i}[k]\right)\right\} \tag{40}$$

subject to

$$\sum_{g_i} Q_{g_i}[k] = \sum_{d_j} Q_{d_j}[k]: \qquad \lambda[k] \tag{41}$$

$$Q^{\min}_{g_i}[k] \le Q_{g_i}[k] \le Q^{\max}_{g_i}[k]: \qquad \eta_{g_i}[k] \tag{42}$$

$$F_l(\mathbf{Q_G}[k],\mathbf{Q_D}[k]) \le F_l^{\max}[k]: \qquad \mu_l[k] \tag{43}$$

[4]The actual competitive supplier surplus function is the decentralized unit commitment problem formulated in [1]. However, we make the assumption that the only available information regarding the supplier is his supply function at the spot market, and when the cost function of supplier is revealed in the spot market, the unit commitment decision is already internalized in his supply function.

where $\rho_{t,d_j}(\mathbf{Q_D}[k], \mathbf{Q_G}[k], k)$ and $\rho_{t,g_i}(\mathbf{Q_D}[k], \mathbf{Q_G}[k], k)$ replace $\hat{\rho}_{t,d_j}(\mathbf{Q_D}[k], \mathbf{Q_G}[k], k)$ and $\hat{\rho}_{t,g_i}(\mathbf{Q_D}[k], \mathbf{Q_G}[k], k)$ respectively so that the *penalty* associated with the transmission congestion is expressed separately through the constraint defined in Ineq. (43) under the centralized optimization. The equivalence of the decentralized decision making of Eqs. (36) and (39) and the centralized optimization in Eq. (40) is based on the result presented in [16].

The optimization problem given in Eq. (40) can be expressed as a linear programming problem. The Lagrangian function is constructed as the following:

$$\begin{aligned}\mathcal{L}(\mathbf{Q_G}[k], \mathbf{Q_D}[k], \lambda[k], \mu_\mathbf{L}[k]) = \mathcal{E}\Bigg\{ &\sum_{d_j}\left(\int_{\tilde{Q_{d_j}}[k]=0}^{Q_{d_j}[k]} D_{d_j}(\tilde{Q_{d_j}}[k], k) d\tilde{Q_{d_j}}[k] - \rho_{t,d_j}(\mathbf{Q_D}[k], \mathbf{Q_G}[k], k)\cdot Q_{d_j}[k]\right) \\ &-\sum_{g_i}\left(\int_{\tilde{Q_{g_i}}[k]=0}^{Q_{g_i}[k]} S_{g_i}(\tilde{Q_{g_i}}[k], k) d\tilde{Q_{g_i}}[k] + \rho_{t,g_i}(\mathbf{Q_D}[k], \mathbf{Q_G}[k], k)\cdot Q_{g_i}[k]\right) \\ &+\lambda[k]\left(\sum_{d_j} Q_{d_j}[k] - \sum_{g_i} Q_{g_i}[k]\right) + \sum_{g_i}\eta_{g_i}[k]\left(Q_{g_i}[k] - Q_{g_i}^{\max}[k]\right) \\ &+\sum_{l}\mu_l[k]\left(F_l(\mathbf{Q_G}[k], \mathbf{Q_D}[k]) - F_l^{\max}[k]\right)\Bigg\}\end{aligned} \quad (44)$$

Based on the Lagrangian function defined in Eq. (44), it is evident that the Lagrangian multipliers $\lambda[k]$ and $\eta_{g_i}[k]$ are the shadow prices related to the energy portion of the electric services, and the Lagrangian multiplier $\mu_l[k]$ is related to the transmission portion [4]. Then, the total transmission revenue collected at time k is given by:

$$TR[k] = \sum_{d_j} \rho_{t,d_j}(\mathbf{Q_D}[k], \mathbf{Q_G}[k], k)\cdot Q_{d_j}[k] + \sum_{g_i} \rho_{t,g_i}(\mathbf{Q_D}[k], \mathbf{Q_G}[k], k)\cdot Q_{g_i}[k] + \sum_{l} \mu_l[k]\cdot F_l^{max}[k] \quad (45)$$

If the price of transmission capacity other than the shadow cost related to the transmission congestion is set to be equal to zero, i.e., $\rho_{t,d_j}(\mathbf{Q_D}[k], \mathbf{Q_G}[k], k) = 0$ and $\rho_{t,g_i}(\mathbf{Q_D}[k], \mathbf{Q_G}[k]k) = 0$, then the Lagrangian multipliers, $\lambda[k]$, $\eta_{g_i}[k]$ and $\mu_l[k]$ in the optimization problem in Eq. (44) match the shadow prices arising from Lagrangian formulation of the optimization problem in Eq. (2) after the time scale separation. Thus, the transmission capacity pricing of $\rho_{t,d_j}(\mathbf{Q_D}[k], \mathbf{Q_G}[k], k)$, $\rho_{t,g_i}(\mathbf{Q_D}[k], \mathbf{Q_G}[k]k) \neq 0$ are the result of the economies of scale assumed for the investment into transmission. It is shown in [3] that $\rho_{t,d_j}(\mathbf{Q_D}[k], \mathbf{Q_G}[k], k)$, $\rho_{t,g_i}(\mathbf{Q_D}[k], \mathbf{Q_G}[k]k) \neq 0$ is equivalent to $\rho_{t,d_j}(\mathbf{Q_D}[k], \mathbf{Q_G}[k], k) \neq 0$, while $\rho_{t,g_i}(\mathbf{Q_D}[k], \mathbf{Q_G}[k]k) = 0$ by appropriately adjusting the supply function, $S_{g_i}(Q_{g_i}[k], k)$. For the rest of the paper, we adopt the convention of assigning the transmission charge other than the shadow cost related to the transmission congestion only to the load. In the following section we describe four commonly proposed schemes for assigning the transmission charge.

B. Transmission charge under the cost-of-service regulation

In order to ensure the TP continues to support the energy market as a viable business with a reasonable expected return on her investment, the high degree of economies of scale needs to be addressed through the transmission charge for the investment into network.

Under the rate-of-return regulation (as a particular form of the cost-of-service regulation) imposed on the TP, the regulator guarantees a reasonable rate of return on all of the approved investment into transmission made by the TP. Let $\Upsilon[n]$ be the allowed revenue of the TP for year n determined by the regulator based on the total investment cost given by:

$$\Upsilon[n] = (1 + r_{cos}) \sum_{k=(n-1)T_T+1}^{nT_T} \sum_l (1-\xi)^k C_l^T(K_l^T[k], I_l^T[k], k) \tag{46}$$

where r_{cos} is the rate of return on investment allowed by the regulator. In Eq. (46) we use the fact that typically the time scale for investment into transmission is a year, i.e., $T_T = 1$ year. From the perspective of the TP, the profit is, then, determined by:

$$\begin{aligned} \Pi_{TP}[n] &= \Upsilon[n] - \sum_{k=(n-1)T_T+1}^{nT_T} (1-\xi)^k \left(\sum_l C_l^T(K_l^T[k], I_l^T[k], k) + \upsilon_{tech}(e_{tech}[k]) + \upsilon_m(e_m[k]) \right) \\ &= \sum_{k=(n-1)T_T+1}^{nT_T} (1-\xi)^k \left(r_{cos} \sum_l C_l^T(K_l^T[k], I_l^T[k], k) - \upsilon_{tech}(e_{tech}[k]) - \upsilon_m(e_m[k]) \right) \end{aligned} \tag{47}$$

where the expression in Eq. (46) is substituted for $\Upsilon[n]$. The decision for dispensing the efforts into control and into maintenance, e_{tech} and e_m respectively, are assumed to be made only once at the beginning of each year for simplicity. Accordingly the profit maximization of the TP under the rate-of-return regulation is given as the following:

$$\left[\overline{\mathbf{I_T}}^{\star}, \overline{\mathbf{e_{tech}}}^{\star}, \overline{\mathbf{e_m}}^{\star}\right]' = \arg \max_{\substack{\mathbf{I_T}[n] e_{tech}[n], \\ e_m[n]}} \sum_{n=1}^{T_I/T_T} \mathcal{E}\left\{\Pi_{TP}[n]\right\} \tag{48}$$

where we make another simplifying assumption that the investment decision is made not over the infinite time horizon but over the time scale of T_I.

From the perspective of the regulator, the associated cost, $TC_{reg}[n]$ for year n encloses the expense arising from compensating the difference between the revenue collected from the loads and generators and the revenue guaranteed to the TP. By again employing the modeling simplification in [7] of treating the process of making up the difference in the revenue collected and allowed as an exclusive process between regulator and loads, the expression of this cost is given as the following:

$$TC_{reg} = (1 + \lambda_f)(\Upsilon[n] - TR[n]) \tag{49}$$

where λ_f is the shadow cost of public funds (the key concept in the simplification step), and $TR[n]$ is the total revenue collected over the entire year n, i.e.,

$$TR[n] = \sum_{k=(n-1)T_T+1}^{nT_T} (1-\xi)^k \mathcal{E}\left\{TR[k]\right\} \tag{50}$$

derived from Eq. (45).

Based on the cost associated with the regulator in Eq. (49) and/or in Eq. (50), it is evident that there is a significant weight placed on the transmission charge levied on the loads other than the shadow cost related to the transmission congestion. This is due to the high degree of economies of scale assumed for the investment into transmission as before, which without the transmission charge leads to a considerable difference in the revenue between the collected and the allowed.

The concept of basic importance linked to the assigning of the transmission charge is three fold, namely (1) sufficient revenue collection, (2) the distortion introduced by the charge and (3) *fairness* to the parties being levied considering their individual characteristics. The notion of the *optimal transmission pricing* lies with the scheme that allows sufficient revenue collection while minimizing the distortion introduced by the charge and appearing fair to those who pay for the charge. It turns out that the first criterion may be the easiest to comply with assuming that the investment into transmission is made with prudence, although not necessarily the optimal possible, and the relative price for the transmission portion of electricity services are much lower than that for the energy services, i.e.,

$$\sum_{k=1}^{T_T} \sum_{d_j} \rho_{e,d_j} \cdot Q_{d_j}[k] \gg \sum_{k=1}^{T_T} \sum_{d_j} \rho_{t,d_j} \cdot Q_{d_j}[k] \tag{51}$$

which is usually satisfied for many regions in US. Almost any reasonable transmission charging scheme satisfies this criterion. In comparison, the second criterion may be the hardest to comply with because the degree to which distortion is introduced in behavior of parties affected by the transmission charge depends on their respective utility functions, and thus may be quite system specific. The comparison is made with the solution to the optimization problem in Eq. (2) where no transmission charges are imposed on the loads. At the time of writing, no generalized result exists for quantifying the effect of transmission charge. The third criterion is a delicate standard by which different schemes are judged since it tends to be highly subjective. Nevertheless, we pay particular attention to the fairness criterion when we describe the following four schemes more commonly used for assigning the transmission charge, namely (1) complete *ex post* allocation scheme (2) *ex ante* access fee and *ex post* settlement scheme (3) *ex ante* injection tax on load and *ex post* settlement scheme and (4) *ex ante* flow tax on load and *ex post* settlement scheme [8] [10].

B.1 Complete *ex post* allocation scheme

Under the complete *ex post* allocation scheme, no transmission charge is initially levied on the load. At the end of the year, the difference in revenue between the allowed by the regulator and that the collected by the TP is computed. Part of this difference can be assigned to individual loads through various methods including dividing it equally among the loads, dividing it proportionally with respect to the annual peak demand for each load, and dividing it proportionally with respect to the annual demand sum of each load.

Suppose the difference is divided based on the peak demand for each load. Accounting for the profit of the TP and the cost of the regulator given in Eqs. (47) and (50) the systemwide social welfare defined under the scheme may be computed by solving the optimization problem given as the following:

$$
\begin{aligned}
\left[\overline{\mathbf{I_T}}^\star, \overline{\mathbf{e_{tech}}}^\star, \overline{\mathbf{e_m}}^\star\right]' &= \arg \max_{\substack{\mathbf{I_T}[n],\, e_{tech}[n], \\ e_m[n]}} \sum_{n=1}^{T_I/T_T} \mathcal{E}\left\{\Pi_{TP}[n] - (1+\lambda_f)(\Upsilon[n] - TR[n])\right\} \qquad (52) \\
&= \arg \max_{\substack{\mathbf{I_T}[n],\, e_{tech}[n], \\ e_m[n]}} \sum_{n=1}^{T_I/T_T} (1-\xi)^{nT_T} \mathcal{E}\Bigg\{ r_{cos} \sum_l C_l^T(K_l^T[n], I_l^T[n], n) - \upsilon_{tech}(e_{tech}[n]) \\
&\qquad -\upsilon_m(e_m[n]) - (1+\lambda_f)\Bigg(\sum_l C_l^T(K_l^T[n], I_l^T[n], n) \\
&\qquad - \sum_{k=(n-1)T_T+1}^{nT_T} (1-\xi)^k \sum_l \mu_l[k] \cdot F_l^{max}[k]\Bigg)\Bigg\}
\end{aligned}
$$

where $\mu_l[k]$ denotes the Lagrangian multiplier corresponding to solving the following optimization problem:

$$[\mathbf{Q_G}^\star[k], \mathbf{Q_D}^\star[k]]' = \arg \max_{\mathbf{Q_G}[k], \mathbf{Q_D}[k]} \mathcal{E} \left\{ \sum_{d_j} \int_{\tilde{Q_{d_j}}[k]=0}^{Q_{d_j}[k]} D_{d_j}(\tilde{Q_{d_j}}[k], k) d\tilde{Q_{d_j}}[k] \right. \tag{53}$$

$$\left. - \sum_{g_i} \int_{\tilde{Q_{g_i}}[k]=0}^{Q_{g_i}[k]} S_{g_i}(\tilde{Q_{g_i}}[k], k) d\tilde{Q_{g_i}}[k] \right\}$$

subject to

$$\sum_{g_i} Q_{g_i}[k] = \sum_{d_j} Q_{d_j}[k] : \qquad \lambda[k] \tag{54}$$

$$Q_{g_i}^{\min}[k] \leq Q_{g_i}[k] \leq Q_{g_i}^{\max}[k] : \qquad \eta_{g_i}[k] \tag{55}$$

$$F_l(\mathbf{Q_G}[k], \mathbf{Q_D}[k]) \leq F_l^{\max}[k] : \qquad \mu_l[k] \tag{56}$$

Computing the systemwide social welfare requires, as indicated before, solving the two separate optimization problems of different time scales, one dynamics evolving at the slow rate of T_T, and the other dynamics at the fast rate of an hour, as given in Eqs. (52) and (53), respectively.

At first glance, the optimization problem in Eq. (53) appears to be identical to the fast dynamics counterpart in Eq. (2) which represents the benchmark performance measure associated with the transmission provider. However, the transmission charge levied on the loads at the end of the year affects the behavior of each load in a delayed manner so that the load maximizes her consumer surplus as given in Eq. (36). For example, suppose it is possible that the transmission charge reaches a sustainable steady state after this particular scheme has been in effect for a number of years, and we denote that price as $\hat{\rho}_t$. Then, the competitive consumer surplus function is given by:

$$\mathbf{Q_{d_j}}^\star[n] = \arg \max_{Q_{d_j}[k]} \sum_{k=(n-1)T_T+1}^{nT_T} \mathcal{E} \left\{ \int_{\tilde{Q_{d_j}}[k]=0}^{Q_{d_j}[k]} D_{d_j}(\tilde{Q_{d_j}}[k], k) d\tilde{Q_{d_j}}[k] - \rho_{e,d_j}(\mathbf{Q_D}[k], \mathbf{Q_G}[k], k) \cdot Q_{d_j}[k] \right. \tag{57}$$

$$\left. - \hat{\rho}_t \cdot \max_{Q_{d_j}[k]} \left(Q_{d_j}[(n-1)T_T + 1], Q_{d_j}[(n-1)T_T + 2], \cdots, Q_{d_j}[nT_T] \right) \right\}$$

thus there exists a clear distortion to the behavior of the load d_j compared to the benchmark performance measure given in (2) although the extent to which the distortion is introduced is not clear as it depends on various factors including the actual functional form of $D_{d_j}(Q_{d_j}[k], k)$ in Eq. (1) and the relative size of $\hat{\rho}_{t,d_j}$ to $\rho_{e,d_j}(\mathbf{Q_D}[k], \mathbf{Q_G}[k], k)$. Eq. (57) further suggests that the optimization problem in Eq. (53) needs to be viewed with an implied assumption that the distortion to the behavior in the load, due to the *ex post* transmission charge, is already reflected in the demand function observed from the perspective of systemwide social welfare, i.e., $D_{d_j}(Q_{d_j}[k], k)$ in Eq. (53) is different from $D_{d_j}(Q_{d_j}[k], k)$ in Eq. (1).

In addition, the difference in revenue between the collected and the allowed is largest under the complete *ex post* allocation scheme. This is due to the fact that no transmission charge is alloted to offset this very difference required to be computed in order to assess the regulator associated cost, $TC_{ref}[n]$, defined in Eq. (49). This large difference results in considerable deviation of the solution to Eq. (52) from the slow dynamics counterpart in Eq. (1).

Similarly some variations to the complete *ex post* allocation scheme are subject to the distortion introduced to the fast dynamics of the benchmark performance measure, although the magnitude to which such distortion stands is not known in general, and are subject to, perhaps more importantly, the largest deviation permeated in the slow dynamics since no effort is made to offset the difference in revenue between the amount collected and that allowed. The variations to the scheme refers to the methods by which the actual allocation takes place, including dividing it equally among the loads, dividing it proportionally with respect to the annual peak demand for each load, and dividing it proportionally with respect to the annual demand sum of each load.

B.2 *Ex ante* access fee and *ex post* settlement scheme

Under the *ex ante* access fee and *ex post* settlement scheme, some transmission charge is levied on the load in the form of access fee at the beginning of the year, and then if there exists a difference in revenue between the amount collected through the access fee and that allowed by the regulator at the end of the year, *ex post* charges are imposed on the loads. The *ex post* charge can again take on various forms as discussed in the previous section. We make one simplifying assumption that from the sense of expected value, the adequate *ex ante* access fee can be determined so that no *ex post* charge is necessary.

Given that assumption, accounting for the profit of the TP and the cost of the regulator given in Eqs. (47) and (50), the systemwide social welfare defined under the scheme may be computed by solving the optimization problem given as the following:

$$\left[\overline{\mathbf{I_T}}^{\star}, \overline{\mathbf{e_{tech}}}^{\star}, \overline{\mathbf{e_m}}^{\star}\right]' = \arg \max_{\substack{\mathbf{I_T}[n], e_{tech}[n], \\ e_m[n]}} \sum_{n=1}^{T_I/T_T} (1-\xi)^{nT_T} \mathcal{E} \Bigg\{ r_{cos} \sum_l C_l^T(K_l^T[n], I_l^T[n], n) - \upsilon_{tech}(e_{tech}[n]) \tag{58}$$

$$-\upsilon_m(e_m[n]) - (1+\lambda_f) \Bigg(\sum_l C_l^T(K_l^T[n], I_l^T[n], n)$$

$$- \sum_{d_j} \hat{\rho}_t - \sum_{k=(n-1)T_T+1}^{nT_T} (1-\xi)^k \sum_l \mu_l[k] \cdot F_l^{max}[k] \Bigg) \Bigg\}$$

where $\hat{\rho}_t$ is the access fee charged to each load within the network, and $\mu_l[k]$ denotes the Lagrangian multiplier corresponding to solving the optimization problem given in Eq. (53) subject to the constraints given Eq. (54) and Ineqs. (55) and (56).

Once again, given that the optimization problem given in Eq. (53) requires to be solved, the solution is expected, at first glance, to be the same as the fast dynamics counterpart of the benchmark performance measure that results from solving the optimization problem given in Eq. (2). However, the access fee charged at the beginning of the year may actually distort the behavior of loads once the surplus function given in Eq. (36) is examined closely. Accordingly, the maximization of the competitive consumer surplus function by each load d_j may be represented as the following under the *ex ante* access fee scheme:

$$\mathbf{Q_{d_j}}^{\star}[n] = \arg \max_{Q_{d_j}[k]} \sum_{k=(n-1)T_T+1}^{nT_T} \mathcal{E} \left\{ \int_{\tilde{Q_{d_j}}[k]=0}^{Q_{d_j}[k]} D_{d_j}(\tilde{Q_{d_j}}[k], k) d\tilde{Q_{d_j}}[k] - \rho_{e,d_j}(\mathbf{Q_D}[k], \mathbf{Q_G}[k], k) \cdot Q_{d_j}[k] - \hat{\rho}_t \right\} \tag{59}$$

Suppose there exists a self-supplying load who requires drawing small amount of electricity from the network only at the times when her supply system fails. As the rate of failure decreases, at some point it is possible that the following condition is met:

$$\max_{Q_{d_j}[k]} \sum_{\substack{k\,=\,(n\,-\,1)T_T\,+\,1 \\ k\,\in\,\text{Failure}}}^{nT_T} \mathcal{E} \left\{ \int_{\tilde{Q_{d_j}}[k]=0}^{Q_{d_j}[k]} D_{d_j}(\tilde{Q_{d_j}}[k], k) d\tilde{Q_{d_j}}[k] - \rho_{e,d_j}(\mathbf{Q_D}[k], \mathbf{Q_G}[k], k) \cdot Q_{d_j}[k] - \hat{\rho}_t \right\} < 0 \tag{60}$$

even though $\rho_{e,d_j} \ll D_{d_j}(0,k)$ for all k's, and the assumption in Ineq. (51) holds true. In this case, the load in question is better off (in the purely economic sense) by being disconnected from the network and by not purchasing the electricity from the network even when her supply system is not operational because of the high transmission charge. Thus, there exists a clear distortion to the behavior of the load d_j compared to the benchmark performance measure given in (2) although again, the extent to which the distortion is introduced is not clear as it depends on various factors including perhaps most importantly the actual functional form of $D_{d_j}(Q_{d_j}[k], k)$ in Eq. (1). A further inference can be made from Ineq. (60)

that the optimization problem in Eq. (53) needs to be viewed with an implied assumption that the distortion to the behavior in the load is already reflected in the number of loads participating in the electricity market, i.e., d_j's in (53) are different from d_j's in Eq. (1) under the *ex ante* access fee scheme.

B.3 *Ex ante* injection tax on load and *ex post* settlement scheme

Under the *ex ante* injection tax on load and *ex post* settlement scheme, first, the tax rate for allowing injection[5], $\hat{\rho}_t[n]$, is determined. Then, the transmission charge is levied on the load in the form of injection tax proportional to the demand at each hour. If there exists a difference in revenue between the collected through *ex ante* injection tax and the allowed by the regulator at the end of the year, *ex post* charges are imposed on the loads. The *ex post* charge can again take on various forms as discussed earlier. We make one simplifying assumption that from the sense of expected value, the adequate *ex ante* injection tax rate can be determined so that no *ex post* charge becomes necessary at the end of the year.

Given that assumption, accounting for the profit of the TP and the cost of the regulator given in Eqs. (47) and (50) the systemwide social welfare defined under the scheme may be computed by solving the optimization problem given as the following:

$$\left[\overline{\mathbf{I_T}}^\star, \overline{\mathbf{e_{tech}}}^\star, \overline{\mathbf{e_m}}^\star\right]' = \arg \max_{\substack{\mathbf{I_T}[n], e_{tech}[n], \\ e_m[n]}} \sum_{n=1}^{T_I/T_T} (1-\xi)^{nT_T} \mathcal{E} \Bigg\{ r_{cos} \sum_l C_l^T(K_l^T[n], I_l^T[n], n) - \upsilon_{tech}(e_{tech}[n]) \tag{61}$$

$$-\upsilon_m(e_m[n]) - (1+\lambda_f) \left[\sum_l C_l^T(K_l^T[n], I_l^T[n], n) - \sum_{k=(n-1)T_T+1}^{nT_T} (1-\xi)^k \times \right.$$

$$\left. \left(\sum_{d_j} \hat{\rho}_t[n] \cdot Q_{d_j}[k] + \sum_l \mu_l[(n-1)T_T + k] \cdot F_l^{max}[k] \right) \right] \Bigg\}$$

where $\mu_l[k]$ denotes the Lagrangian multiplier corresponding to solving the following optimization problem:

$$[\mathbf{Q_G}^\star[k], \mathbf{Q_D}^\star[k]]' = \arg \max_{\mathbf{Q_G}[k], \mathbf{Q_D}[k]} \mathcal{E} \left\{ \sum_{d_j} \left(\int_{\tilde{Q_{d_j}}[k]=0}^{Q_{d_j}[k]} D_{d_j}(\tilde{Q_{d_j}}[k], k) d\tilde{Q_{d_j}}[k] - \hat{\rho}_t[n] \cdot Q_{d_j}[k] \right) \right. \tag{62}$$

[5]In the case of load, *withdrawal* is perhaps the more appropriate word to describe the tax scheme since the load takes electricity from the network. However, we use the word injection to mean withdrawal in order to comply with more conventional usage of the term in the electric power industry at the time of writing.

$$- \sum_{g_i} \int_{\tilde{Q_{g_i}}[k]=0}^{Q_{g_i}[k]} S_{g_i}(\tilde{Q_{g_i}}[k], k) d\tilde{Q_{g_i}}[k] \Big\}$$

subject to the constraints in Eq. (54) and Ineqs. (55) and (56).

Clearly the distortion in the behavior of each load is introduced to the market mechanism due to the transmission charge in the form of the injection tax as evident by the new term, $\hat{\rho}_t[n] \cdot Q_{d_j}[k]$ introduced in the optimization problem in Eq. (62) when compared to the fast dynamics counterpart in Eq. (2) which represents the benchmark performance measure associated with the transmission provider. Although the extent to which this distortion affects the systemwide social welfare is not known, if $\rho_{e,d_j} \ll D_{d_j}(0, k)$ for all k's, and the assumption in Ineq. (51) holds true, it may be inferred that the distortion under the *ex ante* injection tax scheme is smaller than under the *ex ante* access fee scheme since the number of loads is preserved when compared to that from Eq. (1).

B.4 *Ex ante* flow tax on load and *ex post* settlement scheme

Under the *ex ante* flow tax on load and *ex post* settlement scheme, first, the tax rate for allowing flow through the network, $\hat{\rho}_t[n]$, is determined. Then, the transmission charge is levied on the load in the form of flow tax proportional to total electric power flow throughout the network caused by the load satisfying her demand at each hour. If there exists a difference in revenue between the amount collected through *ex ante* injection tax and that allowed by the regulator at the end of the year, *ex post* charges are imposed on the loads. The *ex post* charge can again take on various forms as discussed earlier. Once again, we make the simplifying assumption that from the sense of expected value, the adequate *ex ante* flow tax rate can be determined so that no *ex post* charge becomes necessary at the end of the year.

The apparent electric power flow through transmission line l at hour k, F_l[k], is a function of the total injection into each bus in the system, i.e.,

$$F_l[k] = F_l(\mathbf{Q_G}[k], \mathbf{Q_D}[k]) \tag{63}$$

for an existing network. The vectors, $\mathbf{Q_G}[k]$ and $\mathbf{Q_D}[k]$, designate the amount of electricity injected into the network by generators and the amount of electricity withdrawn from the network by load respectively, *determined through the market clearing process* in the spot market *under the ex ante flow tax scheme*. Let f_{l,d_j} denote the flow on line l related to load

d_j derived by decomposing the apparent flow $F_l[k]$ into the flow corresponding to supplying the demand at the same load, $Q_{d_j}[k]$. Then, f_{l,d_j} can be computed using the following expression:

$$f_{l,d_j}[k] = F_l(\mathbf{Q_G}_{d_j}[k], \mathbf{Q_D}_{d_j}[k]) \tag{64}$$

where $\mathbf{Q_G}_{d_j}[k]$ and $\mathbf{Q_D}_{d_j}[k]$ are given by:

$$\mathbf{Q_G}_{d_j}[k] = \left(\frac{Q_{d_j}[k]}{\sum_{d_j} Q_{d_j}[k]} \right) \cdot \mathbf{Q_G}[k] \tag{65}$$

$$\mathbf{Q_D}_{d_j}[k] = [0, \cdots, Q_{d_j}[k], 0, \cdots, 0]' \tag{66}$$

Typically, for notational convenience, given a transmission line l connecting buses i and j, an arbitrary direction ij is defined. According to this direction the computed flow is either positive if the flow is from bus i to bus j, or negative otherwise. Let $q^+_{l,d_j}[k]$ and $q^-_{l,d_j}[k]$ denote the positive and the negative directional flow of $f_{l,d_j}[k]$, i.e.,

$$q^+_{l,d_j}[k] = \begin{cases} f_{l,d_j}[k] & \text{if } f_{l,d_j}[k] \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{67}$$

$$q^-_{l,d_j}[k] = \begin{cases} -f_{l,d_j}[k] & \text{if } f_{l,d_j}[k] \leq 0 \\ 0 & \text{otherwise} \end{cases} \tag{68}$$

For example, the apparent flow through transmission line l, $F_l[k]$, is the difference between the positive directional flow, $q^+_{l,d_j}[k]$, and the negative directional flow, $q^-_{l,d_j}[k]$, caused by supplying the individual demand Q_{d_j}, summed over all loads given by:

$$F_l[k] = \sum_{d_j} (q^+_{l,d_j}[k] - q^-_{l,d_j}[k]) \tag{69}$$

The implied reasoning for choosing this particular method of decomposing the apparent flow is that in the spot market, the demand at each load is being supplied by every generator participating in the market proportional to the total demand throughout the network. For other interesting decomposition methods, we refer to [17].

Using the decomposition method in Eq. (64) and accounting for the profit of the TP and the cost of the regulator given in Eqs. (47) and (50) the systemwide social welfare defined under the scheme may be computed by solving the optimization problem given as the following:

$$\left[\overline{\mathbf{I_T}}^\star, \overline{\mathbf{e_{tech}}}^\star, \overline{\mathbf{e_m}}^\star\right]' = \arg \max_{\substack{\mathbf{I_T}[n], e_{tech}[n], \\ e_m[n]}} \sum_{n=1}^{T_I/T_T} (1-\xi)^{nT_T} \mathcal{E} \left\{ r_{cos} \sum_l C_l^T(K_l^T[n], I_l^T[n], n) - \upsilon_{tech}(e_{tech}[n]) \right. \tag{70}$$

$$-v_m(e_m[n]) -(1+\lambda_f)\left[\sum_l C_l^T(K_l^T[n], I_l^T[n], n) - \sum_{k=(n-1)T_T+1}^{T_T}\sum_l (1-\xi)^{k-T_T}\times\right.$$

$$\left.\left.\left(\hat{\rho}_t[n]\sum_{d_j}(q^+_{l,d_j}[k]+q^-_{l,d_j}[k]) + \mu_l[k]\sum_{d_j}(q^+_{l,d_j}[k]-q^-_{l,d_j}[k])\right)\right]\right\}$$

where $\mu_l[k]$ denotes the Lagrangian multiplier corresponding to solving the following optimization problem:

$$[\mathbf{Q_G}^\star[k], \mathbf{Q_D}^\star[k]]' = \arg\max_{\mathbf{Q_G}[k],\mathbf{Q_D}[k]} \mathcal{E}\left\{\sum_{d_j}\left(\int_{\tilde{Q_{d_j}}[k]=0}^{Q_{d_j}[k]} D_{d_j}(\tilde{Q_{d_j}}[k],k)d\tilde{Q_{d_j}}[k] - \sum_l \hat{\rho}_t[n](q^+_{l,d_j}[k]+q^-_{l,d_j}[k])\right)\right. \quad (71)$$

$$\left.-\sum_{g_i}\int_{\tilde{Q_{g_i}}[k]=0}^{Q_{g_i}[k]} S_{g_i}(\tilde{Q_{g_i}}[k],k)d\tilde{Q_{g_i}}[k]\right\}$$

subject to the constraints in Eq. (54) and Ineqs. (55) and (56).

Similar to the optimization problem in Eq. (62), there is a clear distortion in the behavior of each load according to the optimization problem in Eq. (71) under the *ex ante* flow tax scheme when compared to the fast dynamics counterpart in Eq. (1) which represents the benchmark performance measure associated with the transmission provider. It may be inferred that the distortion under the *ex ante* flow tax scheme is smaller than that under the *ex ante* access fee scheme assuming $\rho_{e,d_j} \ll D_{d_j}(0,k)$ for all k's, and that Ineq. (51) holds true although it is not known the exact extent to which this distortion affects the systemwide social welfare.

B.5 Comparison of transmission charging methods

We compare the four charging methods described above with respect to the three criteria for assigning the transmission charge introduced earlier, namely (1) sufficient revenue collection, (2) the distortion introduced by the charge and (3) *fairness* to the parties being levied considering their individual characteristics.

First, with respect to the sufficient revenue collection criterion, the methods described above fare well except for one method. For *ex ante* access fee, *ex ante* injection tax, and *ex ante* flow tax schemes, it is assumed that, from the sense of expected value, the adequate *ex ante* access fee, injection tax rate and flow tax rate can be determined so that no *ex post* charge becomes necessary at the end of the year. If the condition in Ineq. (51) holds true, as is the case most of the time, this assumption is valid even with the distortion insinuated in the behavior of loads. Thus, under any one of these schemes sufficient revenue is collected to match the allowed revenue given the assumption. In the case of the complete *ex post* allocation scheme, however, no additional revenue other than that from congestion pricing is raised in order to reduce the difference in revenue between the amount collected and that allowed. By requiring the settlement after the fact, a significant regulatory oversight becomes apparent under this scheme. With the simplification in modeling of the regulatory process adopted from [7] throughout development, the higher need for regulatory oversight surfaces as a loss of systemwide social welfare proportional to the shadow cost of public fund, λ_f, in Eq. (49).

With respect to the minimal distortion criterion it is not clear which scheme may be more enabling since the extent to which the transmission charge affects the behavior of loads is highly system specific as it depends on various factors including the actual functional form of $D_{d_j}(Q_{d_j}[k], k)$ in Eq. (1) and the relative size of ρ_{t,d_j} to ρ_{e,d_j}. Nevertheless, in the case of the *ex ante* access fee scheme, the loads with a small and occasional demand are entirely discouraged from participating in the market. As the number of load centers exploring unconventional means of satisfying their energy need increases, this type of distortion seems to be much more unfavorable than under the other schemes.

With respect to the fairness criterion the *ex ante* flow tax scheme seems to be preferred over the other schemes. Suppose that there are three different loads, d_i, d_j and d_k with the following characteristics. The peak demand for each load is the same, i.e., $Q^{peak} = Q^{peak}_{d_i} = Q^{peak}_{d_j} = Q^{peak}_{d_k}$. At each hour throughout the year, the demand at load d_i and d_j is at its peak while the demand at load d_k is zero except for hour t when the demand at load d_k is at its peak. Mathematically, $\forall \tau \in [(n-1)T_T + 1, nT_T]$ and $\tau \neq t$, $Q^{peak} = Q_{d_i}[\tau] = Q_{d_j}[\tau]$ and $Q_{d_k}[\tau] = 0$. For t, $Q^{peak} = Q_{d_i}[t] = Q_{d_j}[t] = Q_{d_k}[t]$. Plus, the load d_i is located near cheaper generation sources while d_j and d_k are located near expensive sources. The load d_i is located quite far away from the other loads. Then, under the complete *ex post* scheme and under the *ex ante* access fee scheme described above, each load in the network would pay the same amount of transmission charge. Under the *ex ante* injection tax scheme, the loads d_i and d_j would pay the same amount of transmission charge while the load d_k would pay much less. Under the *ex ante* flow tax scheme, the load d_j pays the most and the load d_k pays the least of the transmission charge. Considering that the value of the network is different to each load based on the corresponding characteristics, the *ex ante* flow tax scheme seems more equitable than the other methods although this matter is somewhat subjective and is debatable. Nevertheless, the complete *ex post* scheme favors the load with consistent usage pattern, the *ex ante* access fee scheme favors the captive load with large demand, and the *ex ante* injection tax scheme favors the load situated far from the generation sources.

Although the four schemes described above are admittedly very different from one another, they, along with various other methods under the rate-of-return regulation, all suffer similarly from several defects, admitting inefficiency. The optimization problem for the regulator as expressed in Eq. (52), (58), (61), or (70) becomes a tremendous burden because various functions necessary for solving the problem are highly uncertain from the perspective of the regulator.

First, there are the transmission related costs to be assessed. Even if it is assumed that a reasonable estimate for the cost associated with investment in transmission may be possible, it is unlikely that the regulator is able to accurately evaluate the costs associated with the control effort and the maintenance effort as these costs tend to be highly dependent on the constantly evolving network operating conditions [6]. Suppose the regulator makes the estimate with the help of the TP. From the optimization problem for the TP given in Eq. (48), it is clear that the incentive structure is such that the TP favors expanding the investment into transmission over increasing either the control effort or the maintenance effort. Given this incentive structure, the cost estimates for control effort and for maintenance effort may be much higher than the actual values with the resulting consequences being again the infamous Averch-Johnson effect as described in the context of the rate-of-return regulation on the vertically integrated utility in [2].

Then, there are the transmission related benefits to be evaluated. This requires forecasting the demand and the supply functions of loads and generators within the network as precisely as possible. The forecasting is not an easy task because only after actually participating in the market and acquiring substantial amount of knowledge about the loads and the generators, is it possible to make an accurate forecast for the demand and the supply functions. Unfortunately for the regulator this knowledge is lacking because she seldom actually participates in the market process, and this task may be better left to the TP since the TP is actually in the market dealing with the loads and the generators all the time in order to provide with their transmission portion of the electric services. However, because, as discussed earlier, each scheme under the rate-of-return regulation carries a safety net of the *ex post* settlement in case the forecast is not precise, so that the revenue requirement for the TP is always fulfilled, the TP lacks the motivation to make the maximum effort for an accurate forecast. This is again likely to lead to inefficiency.

To remedy the situation for an improved efficiency, the rate-of-return regulation needs to be replaced with a more appropriate regulatory structure. The PCR is one form of the performance-based regulation (PBR) commonly suggested as an alternative to the rate-of-return regulation to be imposed on the TP. When applied correctly, the PCR bestows the responsibility of cost-benefit analysis similar to the optimization problem in Eq. (52), (58), (61), or (52) to the regulated firm, in this case the TP, and the result is an increase in the systemwide social welfare function [9]. In the following section we examine the possible application of PBR in the electric power industry after the restructuring process.

IV. Performance-based-regulation (PBR)

Under the cost-of-service regulation a close link is made between the cost of providing the service and the price charged for the service by the regulated firm. In the context of the electric power industry after the restructuring process, this means the price charged for providing transmission capacity by the TP is strictly based on the cost of investment into the transmission network. As it is pointed out in earlier discussions, in this environment there is little or no incentive for the TP to reduce costs by improving productivity.

The PBR is a regulatory structure where this linkage between the cost and the price of the service is broken by offering instead financial incentives to the regulated firm, the TP, to lower the cost. Under PBR, an improvement in efficiency by the TP is rewarded with higher profit while a loss in productivity is penalized with dwindled profit. Thus, the PBR is viewed as an alternative to the traditional cost-of-service regulation to be placed on the TP as there exist many success stories related to applying the PBR in the telecommunications and railroad industries [9]. One of the main advantages of adopting the PBR is its capability to encourage the implementation of new and more advanced technology in providing the services because the PBR provides incentives that are similar to those provided by competition.

It is possible to devise different approaches to applying PBR so that specific objectives are met. For example, to provide customers with lower prices, the regulator can first set a baseline revenue requirement for the firm. A set of incentives is then proposed to encourage the firm to lower its costs relative to the baseline revenue requirement, and in the case of a realized cost saving, the firm and the customers share the benefit. For the most part, these different approaches can be grouped into three principal categories: price caps, revenue caps and sliding scale mechanisms.

Under the price cap approach, first the regulator determines an appropriate price for providing the service and sets the initial ceiling price. This first step of setting the initial price is *similar to that under the cost-of-service regulation.* Once the initial price is set, then the regulator decides on various indices to be used to compute the ceiling prices for the specified period into the future. These indices include the changes in productivity and unanticipated changes in costs not under the control of the regulated firm. The change in productivity is often referred to as the X factor and prescribes the targeted improvement in efficiency to be achieved by the firm. The unanticipated changes are called the exogenous factor or the Z factor and include such elements as low-income program expenditures and sometimes research and development (R&D) costs [9] [3].

The firm's incentives to reduce costs comes from the higher profit expected under this approach. Any reduction in costs increases the profit of the firm given the price ceilings for the specified period into the future. It is interesting to note that the period over which the price ceilings (typically 5 years) are determined is usually much longer than the price review by the regulator under the rate-of-return regulation (1 year). Such stability in regulation also adds to induce higher efficiency since the firm is assured by keeping the additional profits realized from cost reduction without causing regulatory interference.

Under the revenue cap approach, the regulator sets the ceiling on the firm's allowed revenues instead of prices. Since the revenue is composed of the price and the quantity of the service, the adjustment on the revenue cap is subject to the factors pertaining to the price as well as to the quantity. The factors pertaining to the price are same as the factors described earlier under the price cap approach, including the X factor and the Z factor. The factors pertaining to the quantity are mainly related to the customer growth.

A noted feature of revenue cap approach is the flexibility on the service determining the overall output level endowed to the firm compared to the price cap approach. This flexibility is related to the fact that since only the overall revenue is constrained, the firm can adjust both price and quantity accordingly to achieve the higher profit.

The sliding scale PBR may be considered as a refinement to either the revenue cap or price cap approach. Under this approach, in conjunction with the revenue cap, for example, the regulator, first defines the band of acceptable level of revenues by assigning the minimum and the maximum desired, and determines the sharing mechanism. The price of the service is allowed to fluctuate to enable the firm to attain the acceptable level of revenue. the regulator, then, tracks the revenue of the firm on a yearly basis and invoke the sharing mechanism on the difference between the collected revenue and the desired minimum or maximum (whichever is closer) if the actual revenue falls outside of the band. The advantage of the sliding scale PBR is on its sharing mechanism especially if the economics related to the service being provided by the regulated firm is highly uncertain. Both the firm and the customers are protected by sharing the risks above and below certain threshold.

In the following section we introduce a possible PCR to be imposed on the TP based on the *ex ante* flow tax scheme and examine the merit of the newly proposed mechanism.

A. Price-cap regulation

Consider the *ex ante* flow tax scheme discussed earlier. From Eq. (70) it is evident that the TP's revenue for the year n is given by:

$$TR[n] = \mathcal{E}\left\{ \sum_{k=(n-1)T_T+1}^{nT_T} \sum_{l} (1-\xi)^k \left(\hat{\rho}_t[n] \sum_{d_j} (q^+_{l,d_j}[k] + q^-_{l,d_j}[k]) + \mu_l[k] \sum_{d_j} (q^+_{l,d_j}[k] - q^-_{l,d_j}[k]) \right) \right\} \tag{72}$$

where $\mu_l[k]$ denotes the Lagrangian multiplier corresponding to solving the following optimization problem:

$$[\mathbf{Q_G}^\star[k], \mathbf{Q_D}^\star[k]]' = \arg \max_{\mathbf{Q_G}[k], \mathbf{Q_D}[k]} \mathcal{E}\left\{ \sum_{d_j} \left(\int_{\tilde{Q_{d_j}}[k]=0}^{Q_{d_j}[k]} D_{d_j}(\tilde{Q_{d_j}}[k], k) d\tilde{Q_{d_j}}[k] - \sum_{l} \hat{\rho}_t[n] (q^+_{l,d_j}[k] + q^-_{l,d_j}[k]) \right) \right. \tag{73}$$

$$\left. - \sum_{g_i} \int_{\tilde{Q_{g_i}}[k]=0}^{Q_{g_i}[k]} S_{g_i}(\tilde{Q_{g_i}}[k], k) d\tilde{Q_{g_i}}[k] \right\}$$

subject to the constraints in Eq. (54) and Ineqs. (55) and (56). Suppose the rate of the flow tax, $\hat{\rho}_t[n]$, is allowed to vary hour-by-hour denoted as $\hat{\rho}_t[k]$. By rearranging the expression inside $(\cdot)$ on the right-hand-side (RHS) of Eq. (72) and substituting $\hat{\rho}_t[k]$ for $\hat{\rho}_t[n]$ we have

$$TR[k] = \mathcal{E}\left\{\sum_{l}\sum_{d_j}\left[\left(\hat{\rho}_t[k] + \mu_l[k]\right) q^{+}_{l,d_j}[k] + \left(\hat{\rho}_t[k] - \mu_l[k]\right) q^{-}_{l,d_j}[k]\right]\right\} \tag{74}$$

where

$$TR[n] = \sum_{k=(n-1)T_T+1}^{nT_T} (1-\xi)^k TR[k] \tag{75}$$

From Eq. (74) it is clear what service the TP provides and what price is charged for the service, namely the transmission capacity in the positive direction and in the negative direction, $q^{+}_{l,d_j}[k]$ and $q^{+}_{l,d_j}[k]$, and the transmission rent, $\hat{\rho}_t[k] + \mu_l[k]$ and $\hat{\rho}_t[k] - \mu_l[k]$, respectively.

Before introducing the PCR to be applied to providing the transmission capacity service, it is noted here that the structure of transmission rent as defined in Eq. (74) is *not* in the sense of usual multi-part tariff [13] [12] [11]. The usual sense multi-part tariff refers to the pricing of a service with several prices corresponding to several mutually exclusive cost elements. For example, suppose providing a particular service requires incurring some fixed costs and some operating costs. Then, it is possible to apply the multi-part tariff for this service by charging a fixed price plus a variable price which depends on the overall quantity of service provided. It is true that the pricing appears to have two components, $\hat{\rho}_t[k]$ and $\mu_l[k]$. Furthermore, it is also true that $\hat{\rho}_t[k]$ is placed so that the high fixed cost is recovered for the TP related to the economies of scope, and $\mu_l[k]$ is dependent on the quantity of transmission capacity being demanded. However, $\mu_l[k]$ is zero unless the transmission line l is congested and reflects the marginal *value* of the scarcity in transmission capacity rather than any increase in actual cost incurred to meet the demand growth. In this aspect the analogy is closer to the *peak load pricing* than to the *multi-part tariff* [15] [7].

The newly proposed PCR mechanism consists of regulating the price elements, $\hat{\rho}_t[k]$ and $\mu_l[k]$, for providing the transmission capacity service with the ceiling prices determined by the regulator, $\hat{\rho}_t[n]$ and $\mu_l[n]$, respectively.

First, the regulator defines the initial ceiling prices, $\hat{\rho}_t[1]$ and $\mu_l[1]$. Following the initial prices, the regulator sets the appropriate indices for price adjustment including the inflation i factor and the X factor. Suppose the period of the price review by the regulator is set to be 5 years. Then, the ceiling prices for the subsequent years up to the year 5 are determined by:

$$\hat{\rho}_t[n+1] = \hat{\rho}_t[n]\,(1 + i_\rho - X_\rho) + Z_\rho \tag{76}$$

$$\mu_l[n+1] = \mu_l[n]\,(1 + i_\mu - X_\mu) + Z_\mu \tag{77}$$

for $n = 1, 2, \cdot, 4$. In case there is a significant effect from exogenous factor, which requires an adjustment to the price before the end of the review period, the Z factor is defined for each price element.

Having defined the price cap for each year until the end of the review period the conventional application of PCR means *transferring* the operation and planning authority completely from the regulator to the regulated firm, in this case the TP, so long as the following constraints are met:

$$\hat{\rho}_t[k] \leq \hat{\rho}_t[n] \tag{78}$$

$$\mu_l[k] \leq \mu_l[n] \tag{79}$$

where $k = (n-1)T_T + 1, (n-1)T_T + 1, \cdots, nT_T$. However, $\mu_l[k]$ reflects the value of scarcity in transmission capacity and is determined exogenously through solving the optimization problem in Eq. (73). Thus, some modifications are necessary in enforcing the PCR on the TP. In the following section the necessary modifications are described, and thus, the complete PCR structure of the newly proposed scheme for regulating the TP is presented.

B. Complete formulation of the newly proposed price-cap-regulation (PCR) scheme

It is recognized that the each of the two price elements in Eq. (74) has a different impact on the operation and planning of the electric power network because the rate for flow tax, $\hat{\rho}_t[k]$ is closely related to the recovery of investment cost while the congestion price, $\mu_l[k]$, is intimately associated with the allocation of transmission capacity at the time of scarcity. On one hand, the rate for flow tax, $\hat{\rho}_t[k]$ is necessary for assuring the recovery of the investment cost into the transmission for the optimal systemwide social welfare even when there exists a high degree of the economies of scale in the network. On the other hand, the congestion price, $\mu_l[k]$, sets the marginal value for the transmission capacity so that the allocation of the capacity leads to the optimal systemwide social welfare while abiding by the network constraints, notably the transfer limits on each line in the near real time operation, and the effect of accumulated congestion price over time, $\sum_{n=1}^{T_I/T_T} \sum_{k=(n-1)T_T}^{T_T} (1-\xi)^k \mu_l[k]$ for each line l determines the optimal investment strategy in the planning.

Suppose we make an assumption that there exists a set of prices, $\hat{\rho}_t^\dagger[k] = 0$ and $\mu_l^\dagger[k] \neq 0$ if and only if $F_l(\mathbf{Q_G}[k], \mathbf{Q_D}[k]) = F_l^{\max}(\mathbf{F}[k], K_l[k], e_{tech}[k], e_m[k])$ such that there is a unique vector, $[\mathbf{Q_G}[k], \mathbf{Q_D}[k]]'$ which solves the optimization problem in Eq. (73) for given $K_l[k]$, $e_{tech}[k]$, and $e_m[k]$. The required conditions for this assumption may be roughly described as that there exists an adequate supply of energy portion of electric service [14] [4]. Fortunately this condition is always satisfied under the perfect competition with free entry assumption made earlier for the energy market. The vector, $[\mathbf{Q_G}^\dagger[k], \mathbf{Q_D}^\dagger[k]]'$, is often referred to as the *optimal power flow* solution [5]. Given this assumption consider the application of the PCR and the resultant allocation of transmission capacity by the regulated TP. As before, the ceiling prices, $\hat{\rho}_t[n]$ and $\mu_l[n]$, are determined by initial ceiling prices and the indices for price adjustment imposed by the regulator. Then, in the near real time operation, the TP may choose a set of prices, $\hat{\rho}_t^\ddagger[k] \leq \hat{\rho}_t[n]$ and $\mu_l^\ddagger[k]$ so that there exists a solution, $[\mathbf{Q_G}^\ddagger[k], \mathbf{Q_D}^\ddagger[k]]'$ to the optimization problem in Eq. (73) for given $K_l[k]$, $e_{tech}[k]$, and $e_m[k]$. This is always true since by the assumption above there is at least one such solution obtained by setting $\hat{\rho}_t^\ddagger[k] = \hat{\rho}_t^\dagger[k]$ and $\mu_l^\ddagger[k] = \mu_l^\dagger[k]$.

Under the PCR scheme, the desired result is that the following condition:

$$\mu_l^\ddagger[k] \leq \mu_l[n] \tag{80}$$

is also satisfied. However, this is not assured even under the perfect competition with free entry assumption for the energy market because $\mu_l[k]$ depends not only on the demand and supply functions of the loads and the generators but also on the transmission network conditions, $K_l[k]$, $e_{tech}[k]$, and $e_m[k]$. The perfect market assumption cannot be applied for the transmission network since the TP is a monopoly. If the condition in Ineq. (80) is required to be satisfied absolutely, then there may be no vector, $[\mathbf{Q_G}[k], \mathbf{Q_D}[k]]'$ that solves the optimization problem in Eq. (73) while satisfying the constraints in Eq. (54) and Ineq. (55) and (56). Thus, some modifications are required to enforce the PCR scheme on the TP.

The following modifications are proposed to the conventional PCR scheme to be imposed on the TP. On one hand, for the rate on flow tax, $\hat{\rho}_t[k]$, a strict ceiling price, $\hat{\rho}_t[n]$ applies so that $\hat{\rho}_t[k] \leq \hat{\rho}_t[n]$ for all $k = (n-1)T_T + 1, (n-1)T_T + 2, \cdots, nT_T$. On the other hand, instead of enforcing a rigid ceiling price, the congestion price, $\mu_l[k]$, is free to vary if the transmission line l is congested and is set to zero, otherwise. It is assured by the perfect competition assumption on the energy market that there is a set of prices, $\hat{\rho}_t^{\ddagger}[k] \leq \hat{\rho}_t[n]$ and $\mu_l^{\ddagger}[k]$ such that there exists a solution, $[\mathbf{Q_G}^{\ddagger}[k], \mathbf{Q_D}^{\ddagger}[k]]'$ to the optimization problem in Eq. (73) for given $K_l[k]$, $e_{tech}[k]$, and $e_m[k]$. The reward and penalty scheme under the proposed PCR scheme is such that the performance of the TP is, compensated through the transmission revenue collected at hour k in Eq. (74) amended as the following:

$$TR^{\ddagger}[k] = \begin{cases} \sum_l \sum_{d_j} \left[\left(\hat{\rho}_t^{\ddagger}[k] + \mu_l^{\ddagger}[k]\right) q^+_{l,d_j}[k] + \left(\hat{\rho}_t^{\ddagger}[k] - \mu_l^{\ddagger}[k]\right) q^-_{l,d_j}[k] \right] & \text{if } \hat{\rho}_t^{\ddagger}[k] \leq \hat{\rho}_t[n] \text{ and } \mu_l^{\ddagger}[k] \leq \mu_l[n] \\ (1 - r_{penalty}) \sum_l \sum_{d_j} \mu_l[n](q^+_{l,d_j}[k] - q^-_{l,d_j}[k]) & \text{otherwise, i.e. } \hat{\rho}_t^{\ddagger}[k] = 0 \text{ and} \\ & \mu_l^{\ddagger}[k] > \mu_l[n] \text{ for any } l \end{cases} \tag{81}$$

where $r_{penalty}$ is the penalty rate imposed on the TP for the poor performance. The poor performance of the TP refers to the ill conceived decisions on the amount of investment, control effort and the maintenance effort into the transmission network so that it becomes necessary to invoke congestion prices higher than what is allowed under the PCR scheme. The difference in the revenue between the amount collected and that retained by the TP at hour k when $\hat{\rho}_t[k] = 0$ and $\mu_l[k] > \mu_l[n]$ for any l is given by

$$\sum_l \sum_{d_j} \left[\left(\hat{\rho}_t^{\ddagger}[k] + \mu_l^{\ddagger}[k]\right) q^+_{l,d_j}[k] + \left(\hat{\rho}_t^{\ddagger}[k] - \mu_l^{\ddagger}[k]\right) q^-_{l,d_j}[k] - \mu_l[n](1 - r_{penalty})(q^+_{l,d_j}[k] - q^-_{l,d_j}[k]) \right] \tag{82}$$

and is assumed to be returned to the loads in the spot market[6] indirectly through the regulator, based on the modeling simplification in [7] of treating the process of making up the difference in the revenue collected and allowed as an exclusive process between regulator and loads.

According to Eq. (81) the penalty imposed on the TP for violating the ceiling price on congestion charge is two fold. For hour k when $\mu_l[k] > \mu_l[n]$ for any l, it is required that the rate of flow tax is set to zero as indicated by $\hat{\rho}_t^{\ddagger}[k] = 0$. Plus, the poor performance penalty factor, $r_{penalty}$, reduces the revenue retained by the TP. Due to the imposed poor-performance penalty, the incentive structure for the TP is such that the TP is encouraged to reduce the level of congestion throughout the network below the allowed level. Considering the level of congestion is inversely correlated to the reliability of the network operation, maintaining the congestion price below the desired level is equivalent to maintaining the reliability above the advisable level.

[6]We emphasize here that the refund is made only to the spot market participants without further explanation.

The optimization problem for the slow dynamics associated with the TP under the newly proposed PCR scheme is given as the following:

$$\left[\overline{\mathbf{I_T}}^\star, \overline{\mathbf{e_{tech}}}^\star, \overline{\mathbf{e_m}}^\star\right]' = \arg \max_{\substack{\mathbf{I_T}[n],\, e_{tech}[n], \\ e_m[n]}} \mathcal{E}\left\{\sum_{n=1}^{T_I/T_T} (1-\xi)^{nT_T} TR^\star[n] - \upsilon_{tech}(e_{tech}[n]) - \upsilon_m(e_m[n]) - \sum_l C_l^T(K_l^T[n], I_l^T[n], n)\right\} \tag{83}$$

where

$$TR^\star[n] = \sum_{k=(n-1)T_T+1}^{nT_T} (1-\xi)^k TR(\hat{\rho}_t^\star[k], k) \tag{84}$$

The complementing optimization problem for the fast dynamics is given by

$$\hat{\rho}_t^\star[k] = \arg\max_{\hat{\rho}_t[k]} \sum_l \sum_{d_j} \left[\left(\hat{\rho}_t[k] + \mu_l^\star[k]\right) q_{l,d_j}^+[k] + \left(\hat{\rho}_t[k] - \mu_l^\star[k]\right) q_{l,d_j}^-[k]\right] \tag{85}$$

if $\hat{\rho}_t[k] \leq \hat{\rho}_t[n]$ and $\mu_l^\star[k] \leq \mu_l[n]$, or

$$\hat{\rho}_t^\star[k] = 0 \tag{86}$$

otherwise, i.e. $\mu_l^\star[k] > \mu_l[n]$ for any l. The Lagrangian multiplier, $\mu_l^\star[k]$, are the result of solving the following optimization problem:

$$[\mathbf{Q_G}^\star[k], \mathbf{Q_D}^\star[k]]' = \arg \max_{\mathbf{Q_G}[k], \mathbf{Q_D}[k]} \mathcal{E}\left\{\sum_{d_j}\left(\int_{\tilde{Q}_{d_j}[k]=0}^{Q_{d_j}[k]} D_{d_j}(\tilde{Q}_{d_j}[k], k)\, d\tilde{Q}_{d_j}[k] - \sum_l \hat{\rho}_t^\star[k](q_{l,d_j}^+[k] + q_{l,d_j}^-[k])\right) - \sum_{g_i} \int_{\tilde{Q}_{g_i}[k]=0}^{Q_{g_i}[k]} S_{g_i}(\tilde{Q}_{g_i}[k], k)\, d\tilde{Q}_{g_i}[k]\right\} \tag{87}$$

subject to the constraints in Eq. (54) and Ineqs. (55) and (56). The revenue of the TP at each hour, $TR(\hat{\rho}_t^\star[k], k)$, is given by:

$$TR(\hat{\rho}_t^\star[k], k) = \begin{cases} \sum_l \sum_{d_j} \left[\left(\hat{\rho}_t^\star[k] + \mu_l^\star[k]\right) q_{l,d_j}^+[k] + \left(\hat{\rho}_t^\star[k] - \mu_l^\star[k]\right) q_{l,d_j}^-[k]\right] & \text{if } \hat{\rho}_t^\star[k] \leq \hat{\rho}_t[n] \text{ and } \mu_l^\star[k] \leq \mu_l[n] \\ (1 - r_{penalty}) \sum_l \sum_{d_j} \mu_l[n](q_{l,d_j}^+[k] - q_{l,d_j}^-[k]) & \text{otherwise, i.e. } \hat{\rho}_t^\star[k] = 0 \text{ and } \mu_l^\star[k] > \mu_l[n] \text{ for any } l \end{cases} \tag{88}$$

where $r_{penalty}$ is, as before, the penalty rate imposed on the TP for the poor performance.

According to the optimization problem in Eqs. (85) and (86) the incentive structure of the proposed PCR mechanism is such that the TP increases the transmission revenue by reducing the congestion within the network up to the desirable level determined by the regulator. This incentive structure allows for placing the responsibility *on the TP*, and not on the regulator, of making decisions for the amount of investment, control effort and the maintenance effort into the transmission network as identified in the optimization problem in Eq. (83), whereas the role of the regulator is limited to defining the desirable level of congestion. Given that the level of congestion is inversely correlated to the reliability of the network operation, this means that the regulator has the ultimate responsibility of determining the *minimum* level of reliability desirable in the network. This resembles very closely the role of regulator defining the quality of the service being provided by the regulated firm under the conventional PCR scheme.

V. Illustrative examples

We illustrate some of the ideas presented in this paper through a numerical example using the 5-bus electric power network shown in Figure 2. Table I summarizes the initial capacity

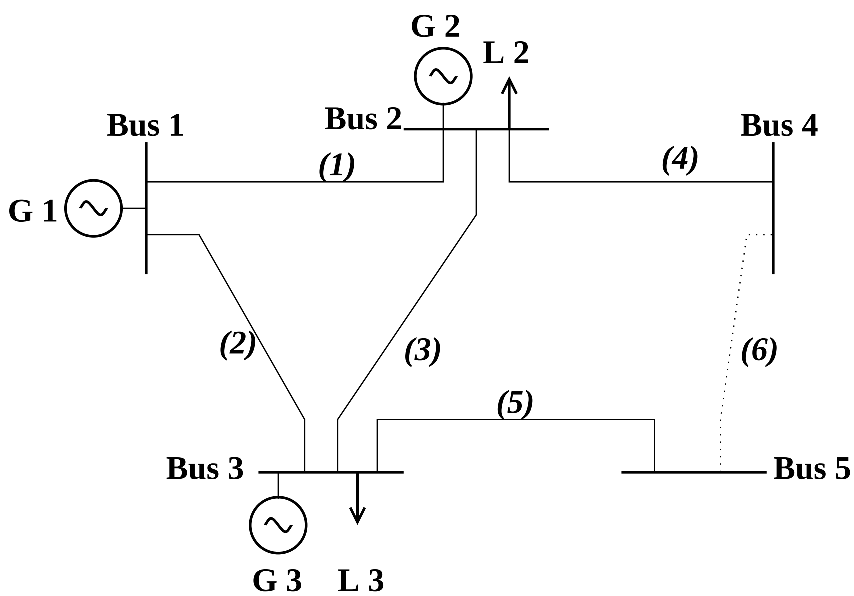

Fig. 2. One-line diagram of the 5-bus electric power network

of each transmission line in the network. The network capacity of zero between bus 4 and bus 5 indicates that currently no line exists between those buses. It is assumed that the small network capacity on the transmission line between bus 2 and bus 3 relative to the other lines

Line #	(1)	(2)	(3)	(4)	(5)	(6)
Initial capacity (MW)	300	300	120	300	300	0

TABLE I

INITIAL CAPACITY OF EACH TRANSMISSION LINE IN THE 5-BUS ELECTRIC POWER NETWORK EXAMPLE

is a result of physical restriction for network expansion in the area, such as zoning limits. Thus, no additional transmission capacity is allowed on that line. No other restriction exists for expanding other lines including between bus 4 and bus 5. As before, there are 4 thermal units at bus 1, 7 hydro units at bus 2, and 1 gas-turbine unit and 1 nuclear unit at bus 3, for the total of 13 units. This time, however, the capacity of each generating unit in the network is assumed to be infinite. Table II summarizes the characteristics of these units for the first year ($n = 1$), including their marginal costs of the form, $S_{g_i}(Q_{g_i}[k]) = 2a_{g_i}Q_{g_i}[k]+b_{g_i}$. These marginal cost functions are useful in computing the systemwide generation cost in this example according to the perfect market assumptions. In this example, a year is composed of 2 seasons, each having 3 days. Depending on the demand of the loads, the seasons and the days are differentiated as peak, shoulder and off-peak. Figure 3 shows the load characteristics at each bus for the first year. For simplicity we assume that the demand is

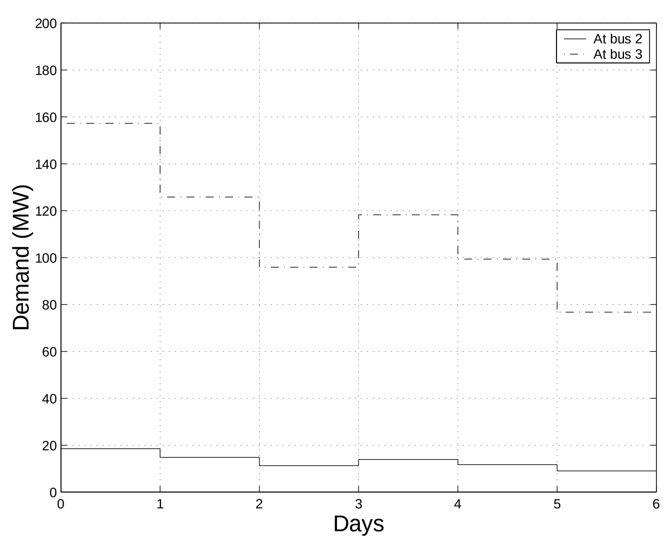

Fig. 3. Load characteristics in years $n = 1, 2$

inelastic throughout the year. The expected system conditions for the next few years are,

Unit #	Type	Bus # (g_i)	a_{g_i}	b_{g_i}
1	thermal	1	60	0
2	thermal	1	60	0
3	thermal	1	250	0
4	thermal	1	122.5	0
5	hydro	2	25	0
6	hydro	2	25	0
7	hydro	2	2.5	0
8	hydro	2	70	0
9	hydro	2	70	0
10	hydro	2	80	0
11	hydro	2	80	0
12	gas-turbine	3	1000	100
13	nuclear	3	3	0

TABLE II

Characteristics of generating units in the 5-bus network

then given as follows: At the beginning of the second year ($n = 2$) the nuclear unit at bus 3 is taken out of service for maintenance and is not expected to come on line until the beginning of year 4. The expected demand of loads in this year is same as shown in Figure 3. In year 3 ($n = 3$), the projected demand of loads increases by 5% from the previous year throughout the network while no change is expected to take place in the generation. Figure 4 shows the load characteristics at each bus for year 3. At the beginning of year 4, the nuclear plant is expected to come back on line while the projected demand stays the same from the previous year as shown in Figure 4.

At the beginning of each year the TP decides the amount of investment into transmission, $I_l^T[k]$, and determines the size of expenses for the control effort, $e_{tech}[k]$, and the maintenance effort, $\upsilon_m(e_m[k], k)$. Figures 5 and 6 show the actual and marginal costs of investment into transmission. For simplicity we assume that the marginal cost of investment into transmission is piece-wise constant: \$30.4616 for 0 to 20MW, \$60.9233 for 20 to 40MW,

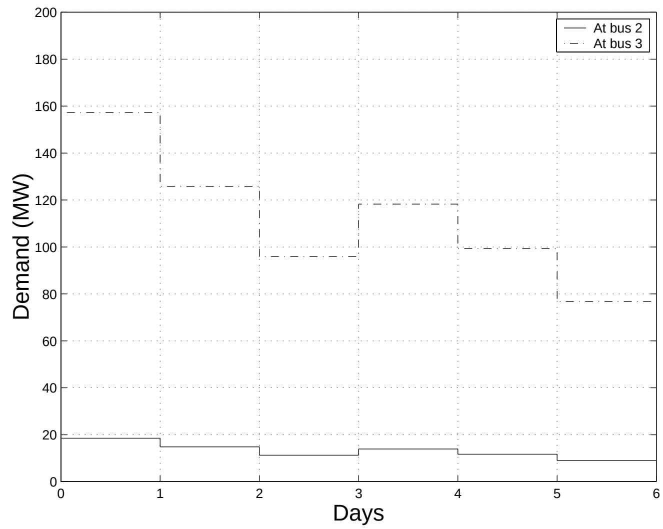

Fig. 4. Load characteristics in year $n = 3, 4$

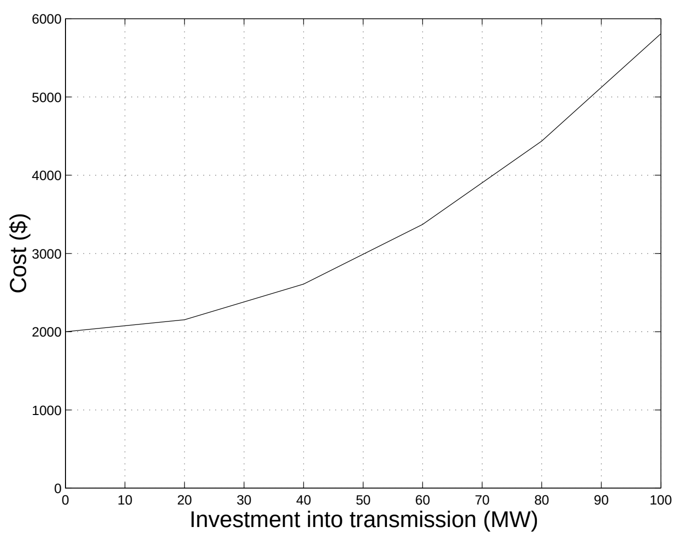

Fig. 5. The cost of investment into transmission

\$91.3849 for 40 to 60MW, \$121.8465 for 60 to 80MW and \$152.3082 for 80 to 100MW. Consequently the actual cost of investment is piece-wise linear. As discussed earlier, it is evident from the figures that the marginal cost of the investment into transmission is much smaller than the average cost for the ranges of investment being considered. In addition, it is assumed that the cost function associated with the maintenance effort is \$0, while the cost function associated with the control effort into transmission is given by \$180 if the control

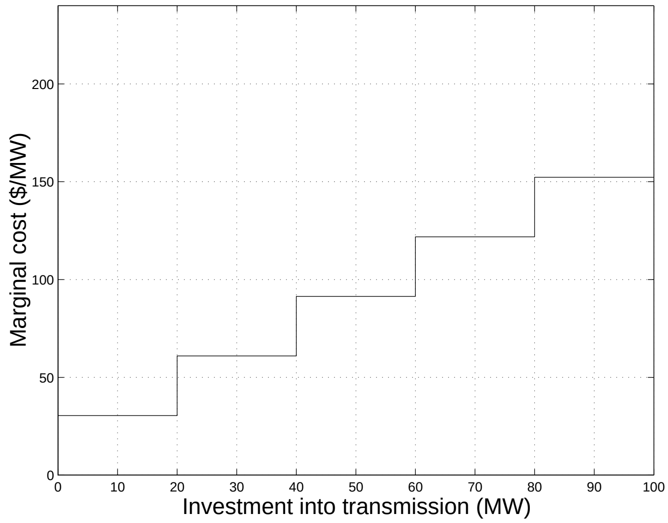

Fig. 6. The marginal cost of investment into transmission

effort is made, or $0 otherwise, i.e,

$$\upsilon_m(e_m[k], k) = 0 \tag{89}$$

$$\upsilon_{tech}(e_{tech}[k], k) = \begin{cases} 180 & \text{if } e_{tech}[k] = 1 \\ 0 & \text{otherwise} \end{cases} \tag{90}$$

Suppose the operational limit on power transfer through line l is given as either a half of the line capacity if no control effort is made or additional 5MW otherwise, i.e,

$$F_l^{\max}(\mathbf{F}[k], K_l[k], e_{tech}[k], e_m[k]) = \begin{cases} 0.5K_l[k] & \text{if } e_{tech}[k] = 0 \\ 0.5K_l[k] + 5 & \text{otherwise} \end{cases} \tag{91}$$

Further, suppose that we apply the so-called DC load flow assumption. Then, the expression for the flow on transmission line l, $F_l(\mathbf{Q_G}[k], \mathbf{Q_D}[k])$ is given by:

$$F_l(\mathbf{Q_G}[k], \mathbf{Q_D}[k]) = \sum_{g_i} H_{lg_i} Q_{g_i}[k] - \sum_{d_j} H_{ld_j} Q_{d_j}[k] \tag{92}$$

where H_{li} denotes the power transfer distribution factor (PTDF) of line l with respect to bus i.

Then, the benchmark performance measure for the TP can be established by substituting various cost functions described above into Eq. (1) and solving the optimization problem given in Eq. (2) as:

$$\overline{\mathbf{I_T}}^{\star} = \begin{bmatrix} \left[I^T_{(1)}[n=1], I^T_{(1)}[n=2], I^T_{(1)}[n=3], I^T_{(1)}[n=4]\right] \\ \left[I^T_{(2)}[n=1], I^T_{(2)}[n=2], I^T_{(2)}[n=3], I^T_{(2)}[n=4]\right] \\ \left[I^T_{(3)}[n=1], I^T_{(3)}[n=2], I^T_{(3)}[n=3], I^T_{(3)}[n=4]\right] \\ \left[I^T_{(4)}[n=1], I^T_{(4)}[n=2], I^T_{(4)}[n=3], I^T_{(4)}[n=4]\right] \\ \left[I^T_{(5)}[n=1], I^T_{(5)}[n=2], I^T_{(5)}[n=3], I^T_{(5)}[n=4]\right] \\ \left[I^T_{(6)}[n=1], I^T_{(6)}[n=2], I^T_{(6)}[n=3], I^T_{(6)}[n=4]\right] \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 20 & 0 & 0 \end{bmatrix} \tag{93}$$

$$\overline{\mathbf{e_{tech}}}^{\star} = \begin{bmatrix} e_{tech}[n=1] \\ e_{tech}[n=2] \\ e_{tech}[n=3] \\ e_{tech}[n=4] \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} \tag{94}$$

$$\overline{\mathbf{e_m}}^{\star} = \begin{bmatrix} e_m[n=1] \\ e_m[n=2] \\ e_m[n=3] \\ e_m[n=4] \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \tag{95}$$

In addition, Table III gives the complete solution. The significance of the investment into transmission and the expense in control effort is the savings in overall cost for meeting the demand in years 2 and 3. For example, without the network reinforcement, the total generation cost is \$7,527.86. With the reinforcement, whose cost amounts to \$2969.23, the total generation cost is reduced to \$2,491.72. This is a savings of \$2,066.90.

In comparison, either 20MW or 30MW of the investment into transmission alone amounts to \$2,609.23 or \$3,370.77 while the total generation cost corresponding to the investment reduces to \$2,945.08 or \$2,164.62. This results in savings of \$1,973.54 and \$1,992.46, respectively. The combined 20MW investment and the control effort in only year 2 or year 3 amounts to the network cost of \$2,789.23, and the corresponding total cost of generation is given by \$2754.72 or \$2,682.08. This is savings of \$1,983.91 and \$2,056.54, respectively.

Under the rate-of-return regulation scheme, it is likely that the TP would prefer 30MW of the investment into generation alone since while the systemwide savings is still at a reasonable level, in this case the rate base is higher than when combined with the expense in control

	Peak season			Off-peak season		
	Peak day	Shoulder day	Off-peak day	Peak day	Shoulder day	Off-peak day
$n = 1$						
Q_{g_1}	8.76	7.01	5.35	6.59	5.54	4.28
Q_{g_2}	102.78	82.22	62.69	77.29	64.96	50.16
Q_{g_3}	64.21	51.37	39.17	48.28	40.58	31.33
$n = 2$						
Q_{g_1}	59.50	11.05	8.42	10.38	8.73	6.74
Q_{g_2}	116.25	129.55	98.78	121.78	102.35	79.03
Q_{g_3}	0.00	0.00	0.00	0.00	0.00	0.00
$n = 3$						
Q_{g_1}	75.23	11.60	8.84	10.90	9.16	7.08
Q_{g_2}	109.31	136.03	103.72	127.87	107.46	82.98
Q_{g_3}	0.00	0.00	0.00	0.00	0.00	0.00
$n = 4$						
Q_{g_1}	9.20	7.36	5.61	6.92	5.82	4.49
Q_{g_2}	107.92	86.33	65.83	81.15	68.20	52.66
Q_{g_3}	67.42	53.93	41.13	50.70	42.61	32.90

TABLE III

Generation dispatch of the benchmark performance

effort. Therefore, unless a regulator is aware of this particular advantage of the control effort, the network is likely to be operated at a suboptimal level, i.e. Averch-Johnson effect.

By contrast, suppose the ceiling prices on congestion charge are set to be \$45 for years 2 and 3 on the transmission line between bus 4 and bus 5 with no additional penalty for exceeding this limit (i.e., $r_{penalty} = 0$) under the proposed PCR scheme. Then, with 40MW of investment into transmission alone, the transmission congestion occurs 4 times in years 2 and 3 with the shadow prices of the line being \$43.43, \$0.40 (year 2), \$54.19 and \$9.00 (year 3). Since the TP is only entitled up to \$45 for congestion, the total congestion charge collected by the TP is given as \$1,957. With 40MW of investment and the control effort made only in year 2, the corresponding shadow prices are given as \$32.70 (year 2), \$54.19 and \$9.00 (year 3), resulting in the total congestion charge of \$1,898. With 40MW of investment and the control effort made only in year 3 this time, the corresponding shadow prices change to \$43.43, \$0.40 (year 2) and \$43.46 (year 3) yielding the total congestion charge of \$1,963. If the control effort made in both years 2 and 3 in addition to 40MW of investment, the resulting congestion charge is reduced to \$1,904. Finally, with 60MW investment, the shadow prices for the transmission line between bus 4 and bus 5 are \$21.96 (year 2) and \$32.72 (year 3) producing the total of \$1,367 congestion charge. Thus, clearly the decision for the 40MW investment followed by the control effort made only in year 3 is favorable to the TP compared to other decisions under the proposed PCR scheme. This is because the maximum profit is obtained while staying within the ceiling prices set by the regulator. It is interesting to note that if the ceiling prices are set as \$33 for year 2 and \$45 for year 3 instead, then the optimal decision by the TP results in the benchmark performance.

Finally, it is noted that the cost of the investment into transmission is not recovered solely based on the congestion charge. This is because of the high fixed cost element of \$2,000 as shown in Figure 5. For example, for the optimal decision by TP for the ceiling price of \$45 for years 2 and 3, the total investment cost is over \$2,600 yet the congestion charge collected is only \$1,963. Therefore, some form of supplemental charge is required in order to induce the optimal decision by the TP. In the case of the proposed PCR scheme, this results in 6.92 (\$/MW) to be imposed as one possible *ex ante* flow tax, which provides an added incentive for the TP to implement various control methods so that the congestion charge does not exceed the ceiling price, as described in this chapter.

VI. CONCLUSION

The importance of a properly functioning forward market for energy has been well understood including from the perspective of solving the unit commitment problem [1]. It is also well recognized that it is practically impossible to have a liquid forward market for energy without well-thought through delivery (transmission) provision.

In order to create a long term transmission market, the ability of the TP to take on the financial risks is also very important when implementing the longer term transmission rights under uncertainty. In this paper we have shown that the propose price-cap regulation provides a possible framework for performance based regulation necessary for such undertaking of financial risks.

References

[1] Allen, E., "Stochastic Unit Commitment in a Deregulated Electric Utility Industry." Ph.D. thesis, Massachusetts Institute of Technology, 1998

[2] Harvey Averch, Leland L. Johnson, "Behavior of the Firm Under Regulatory Constraint", The American Economic Review, Vol. 52, No. 5., December 1962, pp. 1052-1069.

[3] Awerbuch, S., Hyman, L. Vesey, A. (Edited), *Independent Transmission Companies: Unlocking the Benefits of Electricity Restructuring*, 1999.

[4] Bertsimas, Dimitris, Tsitsiklis, John, *Introduction to linear optimization*, Athena Scientific, Belmont, MA, 1997.

[5] William W. Hogan, "Contract Networks for Electric Power Transmission; Technical Reference", John F. Kennedy School of Government, Harvard University, September 1990.

[6] Ilić, M., Galiana, F., Fink, L., "Power systems Restructuring. Engineering and Economics", Kluwer Academic Publishers. 1998

[7] Laffont, Jean-Jacques, Tirole, Jean, "The Regulation of Multiproduct Firms: Part I: Theory", Journal of Public Economics vol. 43, no. 1, October 1990: pp. 1-36.

[8] Leotard, Jean-Pierre, *Transmission pricing and incentives for investments under uncertainty in the deregulated power industry*, MIT Master's Thesis, 1999.

[9] Bridger M. Mitchell, Ingo Vogelsang, *Telecommunications pricing : theory and practice*, Cambridge University Press, New York, NY, 1991.

[10] Nasser, Thomas-Olivier, *Imperfect markets for power: competition and residual regulation in the electricity industry*, MIT Ph.D. Thesis, 1997.

[11] Michael C. Naughton, "The Efficiency and Equity Consequences of Two-Part Tariffs in Electricity Pricing", Review of Economics and Statistics, Vol. 68, No. 3., August 1986, pp. 406-414.

[12] Roger Sherman, Michael Visscher, "Rate-Of-Return Regulation and Two-Part Tariffs", Quarterly Journal of Economics, Vol. 97, No. 1., February 1982, pp. 27-42.

[13] Padmanabhan Srinagesh, "Nonlinear Prices and the Regulated Firm", Quarterly Journal of Economics, Vol. 101, No. 1., February 1986, pp. 51-68.

[14] Schweppe, F., Caramanis, M., Tabors, R., Bohn, R., "Spot Pricing of Electricity", Kluwer Academics, 1988.

[15] Steiner, Peter, "Peak Loads and Efficient Pricing", Quarterly Journal of Economics, Vol. 71, November 1957: pp. 585-610.

[16] Yong T. Yoon, *Electric Power Network Economics; Designing Priciples for For-profit Independent Transmission Company and Underlying Architectures for Reliability*, Ph.D. Thesis, Department of Electrical Engineering and Computer Science, January 2001.

[17] Zobian, A., Ilić, M.D., "Unbundling of Transmission and Ancillary Services", IEEE Power Winter Meeting Baltimore, MD. Jan, 1996.