Electric Power Network Economics: Designing Principles for a For-profit Independent Transmission Company and Underlying Architectures for Reliability

by

Yong Tae Yoon

S.B., Mathematics
Massachusetts Institute of Technology (1995)

S.B., Electrical Engineering and Computer Science
Massachusetts Institute of Technology (1997)

M.Eng., Electrical Engineering and Computer Science
Massachusetts Institute of Technology (1997)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Computer Science and Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2001

© Massachusetts Institute of Technology 2001. All rights reserved.

Author..................................................
Yong Tae Yoon
Department of Electrical Engineering and Computer Science
April 11, 2001

Certified by...........................................
Marija D. Ilić
Senior Research Scientist
Thesis Supervisor

Accepted by...........................................
Arthur C. Smith
Chairman, Department Committee on Graduate Students
Electric Power Network Economics: Designing Principles for a For-profit Independent Transmission Company and Underlying Architectures for Reliability

by

Yong Tae Yoon

Submitted to the Department of Electrical Engineering and Computer Science on April 11, 2001, in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Computer Science and Engineering

Abstract

In this thesis we examine the problem of designing architecture of an electric power network with the emphasis on transmission provision that supports various electricity market structures while ensuring the systemwide reliability, following the electric power industry restructuring process. Specifically, the thesis proposes a possible regulatory incentive-based framework critical for creating a for-profit independent transmission company (ITC), and defines the role of ITC under the newly introduced regulatory framework, such that an adequate level of transmission capacity is provided for trading energy services between generators and loads over long-term.

Before the restructuring process all of the electric services are provided by a so-called vertically integrated utility. A vertically integrated utility is a monopolistic entity who is responsible for meeting the region-wide demand for electricity at some acceptable quality by designing and operating entire generation, transmission and distribution assets within a regional electric power network. A vertically integrated utility fulfills its responsibilities by first projecting the demand of its customers over some period of time (typically for the next 5 to 10 years) often assuming inelasticity. Then, based on the load projection, the utility plans for additional generation necessary for meeting the demand over the same period. Due to the lack of practical means of storing electricity and the uncertainty in equipment (both generation and transmission) availability, the adequate generation reinforcement includes not only the increase in projected demand but also the sufficient amount of reserves.

Under the restructuring process, a completely new environment is created for producing, delivering and consuming electric power. A market is implemented for efficient trading of energy, and a vertically integrated utility is functionally unbundled into generation, transmission and distribution sectors by divesting the generation and distribution assets to many market participants including its affiliates. In the market the newly formed generation companies compete with other generation companies to provide energy to individual distribution companies who serve the ultimate consumer loads. The transmission asset owner and the network operator then together become a transmission provider (TP) who exists as a natural monopoly and provides network capacity to generation and distribution companies under the strict oversight of a regulator.

With the introduction of a market mechanism there are a couple of new categories of uncertainties that are not encountered before the restructuring process, namely regulatory uncertainties and the market designs. It is asserted that these uncertainties play a critical role along with the well recognized uncertainties in equipment status/functionality in determining systemwide efficiency as well as reliability in the operation and planning of an electric power network in the new environment.

In dealing with the regulatory uncertainties, the price-cap regulation (PCR) scheme is suggested to replace the traditional cost-of-service regulation. Due to the particular characteristics of the industry including the lack of practical means of storing electricity and the lack of directly controlling the transmission path, some modifications to the conventional form of PCR scheme are necessary before actual application. In this thesis a practicable PCR scheme is proposed for regulating the operation and planning of a TP, which establishes the framework necessary for creating a for-profit ITC.
In dealing with the uncertainties in market designs, the thesis introduces a usable business model for an ITC and the transmission products to be provided by this ITC. It is asserted that much of uncertainties in network usage, due to constantly changing supply and demand of the network users, can be eliminated through offering longer term transmission contracts by the ITC. A liquid market for these transmission contracts is essential for information revelation on the supply and demand of network users. The thesis proposes a workable mechanism for designing the market for transmission.

In dealing with the uncertainties in equipment status/functionality, the thesis only suggests that there is a strong need for analytical tools in accurately computing the operational limit on power transfer through transmission lines within the network. Much work is needed for re-visiting the (short-term) reliability standards created under the old industry regime since various contractual agreements among the market participants now have very different interpretation on how the uncertainties in equipment status/functionality need to be handled. An active future research is urged for defining the market mechanisms essential for unbundling reliability in parallel to the functional unbundling.

As attempted in this thesis, we believe that any proposed designs for electricity market structures should be examined with a clear understanding of the implications on the overall industry performance, as well as with an understanding of the implications on the individual industry participants, such as power suppliers, provider of wires, and consumers. Particular emphasis should be on understanding the long-term (in contrast to only short-term) effects of various changes on the adequacy of supply and evolution of the grid necessary to support the long-term needs of the energy markets. It will take some deep thinking and patience to get the entire electric power industry to function properly following the restructuring process.

Thesis Supervisor: Marija D. Ilić
Title: Senior Research Scientist
To My Family
Acknowledgments

It is difficult to find a good way to express my appreciation to all those who, in one way or another, helped me in the course of my studies at M.I.T. I wish however to thank my advisor, Dr. Marija Ilić, for her support throughout the period of writing of this thesis. Marija transcended the boundaries of her duties by patiently poring over every chapter, commenting at length and improving various earlier drafts. Her encouragement and insights were essential to the completion of this work. I am also indebted to two other advisors, Professor Paul Joskow and Professor John Tsitsiklis, who performed their advisory role with dedication and brilliance.

The dissertation is the last hurdle in the graduate program. I wish to thank Professor George Verghese and Professor Alan Willsky without whose help I would never have crossed the first hurdle. A special thanks to those who have contributed valuable information on different aspects of electric power network economics: Professor Yu-Chi Ho, Professor Paul Kleindorfer, Professor Ingo Vogelsang, and Stephen Connors.

My colleagues in LEES, E-Lab and LIDS have been most helpful, especially Kenneth Collison, who provided some valuable last-minute help that will not be forgotten.

My studies at M.I.T. were financed generously by US DOE’s Energy Information Administration, Korean Electric Power Research Institute, and the industry and government sponsors of a consortium entitled, “New Concepts and Software for Competitive Power Systems: Operations and Management,” including ABB, Constellation Power Source, Electricité de France, and TransÉnergie. However, the conclusion, opinions, and other statements in this dissertation are those of the author and not necessarily those of the sponsors.

To my family who have supported me during my stay at M.I.T., I humbly offer my sincere gratitude. My parents are the ones who made it possible from the first time I came to the Institute as a confused undergraduate student to now as a bewildered PhD. To my brother, Peter, who is about to enter law school and will walk down the Ph.D. path soon after, I promise that I will do my best to be as supportive as he has been for me at such a young age. Finally, Meine Freundin, Yong-mi Kim, is the one who ultimately got me to the end by continually sacrificing of herself in order to give me time to work.

Finally, I give thanks to You, Lord, for all Your blessings.
Contents

1 Introduction .......................................................... 13
   1.1 Problem statement ........................................ 15
   1.2 Thesis organization ...................................... 16
   1.3 Contribution of the thesis ................................ 18

2 Dynamics of the electric power network: reliability, pricing, and regulation .......... 21
   2.1 Full problem formulation - benign social planner .................. 22
       2.1.1 Identifying uncertainties and constraints in the systemwide social welfare maximization problem ........................................ 25
       2.1.2 Applying dynamic programming technique to the systemwide social welfare maximization problem ........................................ 30
       2.1.3 Solving the systemwide social welfare maximization problem and the need for regulation in the actual operation and planning of an electric power network ........................................ 33
   2.2 Operations and planning by a vertically integrated utility: problem formulation .......... 37
       2.2.1 Defining the basic structure of the vertically integrated utility ........................................ 38
       2.2.2 Cost-of-service regulation as an optimization problem ........................................ 42
   2.3 Restructuring process and functional unbundling ........................................ 50
       2.3.1 Defining the objectives of the transmission provider (TP) ........................................ 51
       2.3.2 Defining the objectives of network users ........................................ 52

3 Measuring the performance of the transmission provider (TP) using a systemwide social welfare function under uncertainties ................. 53
   3.1 Benchmark performance measures for the transmission provider (TP) ................. 54
   3.2 Role of regulation in providing conditions under which the efficiency of the overall network approaches the benchmark performance measure ........................................ 59
       3.2.1 Effect of the spot market and the cost-of-service regulation imposed on a transmission provider (TP) ........................................ 60
3.2.2 Transmission charge under cost-of-service regulation ........................................ 63
3.3 Performance-based-regulation (PBR) ................................................................. 73
   3.3.1 Price-cap regulation ...................................................................................... 75
   3.3.2 Complete formulation of the newly proposed price-cap-regulation (PCR) scheme 77

4 Current development of the restructuring process in some regions in the US with an
   emphasis on prudent transmission provision ......................................................... 81
   4.1 Progress in the restructuring process in some parts of the US ......................... 82
       4.1.1 Restructuring process in the California region ............................................. 82
       4.1.2 Restructuring process in the Pennsylvania-New Jersey-Maryland (PJM) region 84
       4.1.3 Restructuring process in the Midwest region ................................................ 87
   4.2 Simplifying models for representing the markets in the California, Pennsylvania-New Jersey-
       Maryland (PJM) and Midwest regions ................................................................. 88
       4.2.1 Voluntary system operator model; California region ..................................... 88
       4.2.2 Mandatory system operator model; Pennsylvania-New Jersey-Maryland (PJM) region 89
       4.2.3 Multilateral Transaction model; Midwest ISO .............................................. 90
   4.3 Some critical observations .................................................................................. 91
       4.3.1 Observation 1: lack of practical means for storing electricity and its implication on the
           need for forward markets for energy .................................................................... 92
       4.3.2 Observation 2: lack of practical means for storing electricity and its implication on the
           need for forward markets for transmission ....................................................... 92
       4.3.3 Observation 3: hidden reliability risks in current market development under the old
           regulatory paradigms ......................................................................................... 93
       4.3.4 Observation 4: need for new reliability standards ........................................... 94
   4.4 Proposed market design in this thesis .................................................................. 95
       4.4.1 Single transmission network ownership within a single energy market ............. 96
       4.4.2 Multiple transmission owners, multiple energy markets: inter-regional transmission
           organization (IRTO) ...................................................................................... 98

5 Independent transmission company (ITC) for profit and markets for transmission .... 99
   5.1 Functions and services ....................................................................................... 100
       5.1.1 Long term network services ......................................................................... 103
       5.1.2 Intermediate term network services ............................................................ 106
       5.1.3 Short term network services ......................................................................... 108
   5.2 Services and pricing ........................................................................................... 110
       5.2.1 Pricing of short term network services ......................................................... 110
       5.2.2 Pricing of intermediate term network services ............................................. 113
### List of Figures

2-1 Typical electric power network consisting of generating substations, load centers and transmission lines .................................................. 22

2-2 Illustrative example of operational and thermal limits on the power transfer through transmission lines .................................................. 28

2-3 Composition of the electric power network under the traditional regulation mechanism .................................................. 38

2-4 The functions of the vertically integrated utility in designing the electric power network .................................................. 38

2-5 The functions of the vertically integrated utility in operating the electric power network .................................................. 40

3-1 Composition of the electric power network economics after the restructuring process .................................................. 59

4-1 Voluntary System Operator Model .................................................. 88

4-2 Mandatory System Operator Model .................................................. 89

4-3 Multilateral Transaction Model .................................................. 91

5-1 Overall market composition under the newly proposed structure .................................................. 100

5-2 Changing Role of Transmission Provider .................................................. 102

5-3 The information exchange between the market participants and the ITC for long term transmission contracts .................................................. 104

5-4 The information exchange among the market participants, the ITC and the SO for intermediate term transmission contracts .................................................. 107

5-5 The information exchange between the market participants and the SO for the allocation of residual transmission capacity in the spot market for energy .................................................. 109

5-6 Discretized load duration curve .................................................. 116

5-7 Membership functions for the individual load pattern .................................................. 116

5-8 One-line diagram of 3-bus electric power network .................................................. 119

5-9 Load characteristics in years $n = 1, 2$ .................................................. 120

5-10 Load characteristics in years $n = 3, 4$ .................................................. 120
6-1 The information exchange among the market participants, the SO and the ITC for intermediate term transmission contracts involving the secondary market ........................................ 125
6-2 The financial exchange between the market participants and the secondary markets for transmission rights ................................................................. 127
6-3 One-line diagram of 3-bus electric power network .................................................. 131
6-4 The physical exchange among participants for each transaction through forward contracts ........ 133
6-5 (Marginal) Supply bid of generator $g_i$ .................................................................. 152
6-6 Markovian Chain Modeling of Equipment Status ................................................... 155
6-7 Aggregation of Marginal Supply Bids in Zone $k$ ...................................................... 158
6-8 Membership Functions for Individual Load Pattern ................................................ 165
6-9 Discretized Load Duration Curve for the Range of System Load .............................. 166
6-10 One line diagram of 118 bus (power flow test case) system ..................................... 167
6-11 Demand pattern for load $i$ in the system .............................................................. 168
6-12 First cut cluster design for line between buses 30 and 38 ...................................... 169

7-1 North American Transmission Systems and NERC Reliability Council Regions .......... 174
7-2 One-line diagram of 5-bus electric power network ................................................... 176
7-3 The change in the area-wide frequency deviation in region I measured at bus 2 following an unanticipated load increase if region I is isolated from region II ........................................ 177
7-4 The change in the network-wide frequency deviation in region I measured at bus 2 following an unanticipated load increase when region I is interconnected to region II ......................... 177
7-5 The supply functions at buses 1, 2, 3 and 4 based on the individual marginal costs of the generation units ........................................................................ 179
7-6 Electric power flows through network lines caused by 50MW transactions between the suppliers at bus 2 and the loads at bus 5 ........................................................................ 181
7-7 Electric power flows based on the supply and demand determined through the market mechanisms separately between regions I and II .................................................... 181
7-8 Actual electric power flows through the network determined based on the supply and demand of regions I and II ........................................................................ 182
7-9 The change in the network-wide frequency deviation in region II measured at bus 4 following an unanticipated load increase when region I is interconnected to region II ........................................ 183
7-10 Cost associated with the exchange through tie-line 4 from the perspective of TP's in terms of reliability ......................................................................................... 186
7-11 Cost associated with the exchange through tie-line 5 from the perspective of TP's in terms of reliability ......................................................................................... 187
7-12 Benefit associated with the transaction between the suppliers at bus 2 and the loads at bus 5 in terms of cost savings ........................................... 188
7-13 Transmission cost to be levied on the network users involved in inter-regional transactions using tie-line 4 ................................................................. 188
7-14 Transmission cost to be levied on the network users involved in inter-regional transactions using tie-line 5 ................................................................. 189
7-15 Actual tie-line flow throughout day n with the use of tertiary level control .................. 191
8-1 One-line diagram of 2-bus electric power network ................................................. 201
8-2 Decision tree for the social welfare optimization problem in the 2-bus network example .... 208
8-3 The sum of the expected generation cost and the expected transmission cost for years 2, 3, and 4 ..................................................................................... 209
8-4 Computation of the optimal decision for the systemwide social welfare maximization problem 210
8-5 Expected prices at bus 1 and at bus 2 for different investment decisions ....................... 214
8-6 Expected revenue of a new gas-turbine generator at bus 1 ........................................ 215
8-7 Expected revenue from the marginal pricing of new capacity subtracted by the transmission cost ...................................................................................... 216
8-8 Expected prices at which the electricity is traded in years 2 and 3 through long term contracts 218
B-1 $\pi$-equivalent model of a transmission line ...................................................... 227
List of Tables

5.1 Marginal cost of generating units in the 3-bus network ........................................... 119
5.2 Transfer capacity of each transmission line in years 1, 2, 3, and 4 in the 3-bus electric power
    network example ........................................................................................................... 121
5.3 Expected prices for network services in year 1 ................................................................ 121
5.4 Expected prices for network services in year 2 ................................................................ 122
5.5 Expected prices for network services in year 3 ................................................................ 122
5.6 Expected prices for network services in year 4 ................................................................ 122

6.1 Exchanges among the market participants and the ISO under the TCC scheme for $k = 1, 2$ . 137
6.2 Exchanges among the market participants and the ISO under the TCC scheme for $k = 3, 4$ . 138
6.3 Exchanges among the market participants and the ISO under the flowgate scheme for $k =
    1, 2, 3, 4 ..................................................................................................................... 143
6.4 Exchanges among the market participants and the SO under the proposed scheme for $k =
    1, 2, 3, 4 ..................................................................................................................... 147
6.5 Prices of intermediate term transmission contracts ............................................................ 149
6.6 Updated prices of intermediate term transmission contracts ............................................. 150
6.7 Estimated sample performance for $\theta_t \in S$ .................................................................. 170
6.8 Individual bus cluster affiliation for $\theta_9$ ...................................................................... 170
6.9 Individual bus cluster affiliation for $\theta_{13}$ .................................................................... 171
6.10 Individual bus cluster affiliation for $\theta_{32}$ ................................................................... 171

7.1 Nominal dispatch schedule for the 5-bus electric power network at hour $k$ .................... 175
7.2 Generation dispatch schedule if regions I and II are isolated ........................................... 179
7.3 Generation dispatch schedule if the regions I and II belong to the same energy market of a
    single control area ......................................................................................................... 180
7.4 Inelastic portion of demands in regions I and II ............................................................... 185
7.5 Dispatch schedule for regions I and II at hour 1 on day $n$ ............................................... 190
7.6 Dispatch schedule for regions I and II at hour 2 on day $n$ ............................................... 190
Chapter 1

Introduction

In this thesis we examine the problem of designing an architecture for an electric power network with an emphasis on transmission provision that supports various electricity market structures while ensuring systemwide reliability, following the electric power industry restructuring process. Specifically, the thesis proposes a possible regulatory framework critical for creating a for-profit independent transmission company (ITC), and defines the role of the ITC under the newly introduced regulatory (incentive-based) framework, such that an adequate level of transmission capacity is provided for trading energy services between the generators and loads over long-term.

Nasser’s work in [57] marks one of the first studies performed on the regulatory framework necessary for overseeing the naturally monopolistic entity providing transmission services. One of the keys to the work in [57] is realizing that transmission lines are isomorphic to horizontally differentiated goods because, once this is realized, the notion of optimal regulation derived in [51] becomes directly applicable. In realizing this isomorphism, the spatial aspect of the transmission provision problem is well recognized. However, because the problem of transmission provision requires studies covering not only a wide geographical area but also a long period of time, the temporal as well as the spatial aspects need to be considered carefully.

In [54] and [55] Leotard et al expand the result in [57] to incorporate the temporal aspects of the problem and present many additional notions of optimal regulation. Opportunity, in [73], Yu identifies a number of uncertainties related to the temporal aspects of the transmission provision problem and asserts that these uncertainties need to be studied carefully in order to maintain an acceptable level of reliability in the operation and planning of an electric power network.

In this thesis we attempt to hybridize the results in [54] and [55] and in [73] to describe the reliability related notion of the optimal regulation of an ITC. As with the results in [57] and in [54] and [55] the difficulties lie in creating the right set of incentives for the ITC to expand its socially optimal transmission capacity. What is interesting are the number of important transmission products which need to be offered

\[1\text{Some descriptions of the restructuring process taking place in the US are presented in Appendix A.}\]
by the ITC in order to deal with these uncertainties.

Before the restructuring process, all of the electric services were provided by a so-called vertically integrated utility. A vertically integrated utility is a monopolistic entity which is responsible for meeting the region-wide demand for electricity at some acceptable quality by designing and operating all the generation, transmission and distribution assets within a regional electric power network. Here the acceptable quality of electricity can actually be defined in two ways, namely (1) planning an adequate amount of economic energy and transmission resources to meet the demand and (2) operating within a satisfactory level of frequency and voltage deviations around a nominal condition. The deviation in frequency (or in voltage) occurs when the supply of real power (or reactive power) does not equal the demand. For a (vertically integrated) utility taking on these responsibilities, a regulator defines and allows a reasonable rate of return so that the utility can collect on its total cost.

A vertically integrated utility fulfills its responsibilities by first projecting the demand of its customers over some period of time (typically for the next 5 to 10 years), often assuming inelasticity. Then, based on the load projection, the utility plans for the additional generation necessary to meet the demand over the same period. Due to the lack of a practical means of storing electricity and the uncertainty in equipment (both generation and transmission) availability, adequate generation reinforcement includes not only the increase in projected demand but also a sufficient amount for ancillary services. Ancillary services refer to the generation and transmission services required to balance the constantly fluctuating demand in the presence of random equipment outages. Once the additional generation is planned, the utility plans the transmission reinforcement necessary for meeting the projected demand. Again, a sufficient amount of transmission capacity needs to be determined while taking various uncertainties into consideration. So as to avoid the planned generation not being feasible due to the inability to design the necessary addition of transmission capacity or vice versa, the planning of generation and the planning of transmission need to be closely coordinated. Finally, the utility manages the existing generation and transmission resources so that the demand can be met reliably.

After the restructuring process a completely new environment is created for producing, delivering and consuming electric power. A market is implemented for the efficient trading of energy, and a vertically integrated utility is functionally unbundled into generation, transmission and distribution sectors by divesting the generation and distribution assets to many market participants, including its affiliates. In the market the newly formed generation companies compete with other generation companies to provide energy to individual distribution companies who serve the ultimate consumer loads. The transmission asset owner and the network operator then together become a transmission provider (TP), which exists as a natural monopoly and provides network capacity to generation and distribution companies under the strict oversight of a regulator.
1.1 Problem statement

With the introduction of the market mechanism in trading energy while the TP remains a monopoly, there is an enormous number of issues to be addressed in order to replace the operational and planning practices of the vertically integrated utility with the functions of various market participants so that the desired increase in efficiency is achieved while an acceptable level of reliability is ensured. For example, in the new environment, who secures enough generation and transmission resources to meet the projected demand for real and reactive power? how are the ancillary services provided along with energy trading? who is responsible for dynamic stability, etc.? Although it is not possible even to list all of the issues, some attempts are made in this thesis to define the role of the TP in fostering competition for increased efficiency and in circumscribing market abuse in order to maintain the reliability related to these issues. The thesis develops the overall scope of the studies by defining a few simplifying assumptions.

We assume that there exist well defined markets for providing sufficient ancillary services from the generation resources. These include voltage control services for matching the reactive power supply and demand, regulation services for balancing the minute-to-minute deviations in loads from the hourly mean, reserve services for offsetting the effect of random outages, and dynamic control services for suppressing any fast dynamics problems in the operation of the electric power network. Further, it is assumed that there is a well defined mechanism for defining the maximum transfer over each individual network line for the given resources available for ancillary services so that the supply and constantly fluctuating demand around the hourly mean can be balanced even in the presence of random equipment outages. With these assumptions, a study of the operation and planning of an electric power network can be performed using the minimum time scale of an hour. In addition, we assume that there exists a spot market conducted on the hourly basis where the market participants can trade energy and purchase the available transmission capacity from a TP as a bundled service. Besides the spot market, it is assumed that there are active energy trades taking place in the form of bilateral contracts among the market participants, and this creates a need for market mechanisms for trading transmission capacity between the market participants and the TP as a separate commodity from energy. Finally, no market power is assumed to be present in the energy market.

Given these assumptions, we examine the design architecture of the electric power system that supports the various electricity market structures while ensuring short-term and long-term reliability, which allows for the optimal evolution of the underlying physical transmission systems. Specifically, in this thesis we address the following issues:

1. what an optimal (and adequate) transmission capacity is for stimulating long term efficiency while ensuring long term reliability

2. what the role of a TP is under the various regulation schemes
   - how to create proper incentives for investing in optimal transmission capacity
3. what the possible structure of a TP is, which promotes active involvement by the TP in supporting energy trades among market participants

4. what possible set of transmission products is necessary for creating a close coordination between generation planning and transmission planning

Other than the minimum time scale of an hour, the problem formulation given in the thesis is fully general.

1.2 Thesis organization

The thesis is organized as follows:

Chapters 2 and 3: Review of electric power system operation and planning In these chapters we review the electric power system operation and planning from the perspective of system dynamics modeling. The emphasis is on defining the notions of optimal transmission configuration and of optimal generation management. The problem of setting a benchmark is closely related to defining the notion of optimal transmission configuration. The problem often referred to as unit commitment is important in defining the notion of optimal generation management. It is asserted that the computational complexity of the decision making process is almost too big to solve both the benchmark problem and the unit commitment under uncertainty.

Chapter 4: Survey of current development in the restructuring process In this chapter we survey the current development in the restructuring process taking place in the California, Pennsylvania-New Jersey-Maryland (PJM) and Midwest regions in the US. There are several characteristics unique to each region, and three simplified models developed earlier in [54] are reviewed for describing these regions. Despite the differences, it is explained that many of the common problems associated with the restructuring process are the result of focusing only on the short term provision of transmission. The explanation given in this chapter is the basis for developing the new market structure proposed in the thesis.

Chapter 5: Postulation of the basic framework for a transmission provider in the new environment In this chapter we postulate the basic framework for the transmission provider (TP), which ensures the reliable provision of electricity in the new environment. The new environment refers to the market induced operation and planning of the electric power system after the restructuring process takes place. The TP functions in a dependent mode under the vertically integrated utility structure and in a passive mode through the current restructuring process, meaning there is little participation by the TP in the market activities. It is asserted that continuing to operate under either one of these modes is not sustainable without degrading the reliability of the entire system. The emphasis is given to creating a for-profit Independent
Transmission Company with two separate functions: market maker and service provider, in order to support the market activities properly.

**Chapters 3 and 5: Investigation of possible architectures for regulating the market maker function of the transmission provider (TP)** In these chapters we investigate possible regulatory architectures necessary for overseeing the market maker function of the transmission provider (TP). The market maker function refers to the day-to-day administration of congestion pricing in the residual market in order to balance supply and demand without being involved in issuing the rights for longer term transmission access to the users. The regulatory architectures to be examined are cost-of-service regulation and performance-based regulation. Under each architecture, several tariff designs, including postage stamp, two-part tariff and peak-load pricing, are analyzed for their performance. It is shown that, in administrating performance-based regulation, the well-defined performances are critical for devising a revenue cap, price cap or sliding scale in conjunction with a tariff design. The emphasis is given to the architecture that allows the TP the flexibility of acting as a service provider while yielding proper incentives for investment compared to the optimal grid defined in Chapters 2 and 3.

**Chapter 6: Examination of the role of the transmission provider (TP) as a service provider in auctioning transmission rights** In this chapter we examine the role of the TP as a financial entity in auctioning transmission rights. Transmission rights are typically classified as physical rights and financial rights depending on their relationship to the underlying transactions for which the rights are traded. Currently the majority of existing proposals, whether they are for financial rights such as transmission congestion contracts or physical rights such as flowgates, segregates the TP from actively participating in the auctioning process for fear of unfair market practices given the TP’s monopolistic status. It is asserted in Chapter 3 that a well-designed regulatory architecture would allow the TP to have profit/risk interests in the auction, in the form of optional pricing without such fear. The emphasis is given for to an option-like scheme of transmission rights.

**Chapter 7: Examination of the role of the transmission provider (TP) as a service provider in implementing inter-regional transactions** In this chapter we examine the role of the TP as a service provider in implementing inter-regional transactions. Inter-regional transactions refer to the wheeling of electricity across several regional boundaries including (but not limited to) the boundaries of control areas, vertically integrated utilities, and markets. A possible framework needed to implement such transactions is proposed first. In this framework many functions of the security coordinator (SC) at each Regional Council level are transferred to new economic agents called inter-regional transmission organizations (IRTO) and to the inter-regional markets, created to control tie-line flows between regions. Then, the effect on the revenue stream of allowing such inter-regional transactions is studied. It is asserted and shown through simplified
simulations that the outcome of the proposed framework depends on the effective interaction between the IRTO and the TP of each region based on the tariff designed in Chapter 3. An ill-designed tariff is likely to distort the value of using the grid in each region, thus preventing economically attractive transactions. The emphasis is given to the framework that allows the co-existence of the various developing market structures in different regions.

Chapter 8: The simplest 2 bus topology example, recapturing assumptions, results and open questions In this chapter we re-iterate the assumptions under which this thesis is written, describe conceptual ways for their relaxations and implications in needed new research, and our results, and make concluding remarks. Next, we suggest the development of the software required to process real-time transaction management. The congestion pricing defined in Chapter 3 and the services defined in Chapter 5 need to be administered in (near) real-time as the requests are made for various transmission products. It is shown that adequate software tools can be implemented if and only if the studies in Chapters 3 and 5 are carried out while incorporating the critical uncertainties defined in Chapter 2. The emphasis is given to introducing advanced approximate methods, including the theory of ordinal optimization, into the software.

1.3 Contribution of the thesis

The thesis establishes a necessary framework, using dynamic programming formulation, for examining the impact of various regulatory structures on a TP’s investment decisions by expanding on the notion of optimal regulation found in [57] and further developed in [54] and [55] to include a number of uncertainties identified in [73] pertinent to the temporal aspect of the transmission provision problem as the electric power industry goes through the restructuring process. The sources of uncertainties can be put into three qualitatively different categories, namely regulatory uncertainties, market designs, and equipment status/functionality.

Regulatory uncertainties are related to the influence regulators have over the decision making processes of the TP and of the network users. Market designs are associated with the ability of the TP to capture the uncertainties in network usage based on the constantly changing supply and demand of the network users. Equipment status/functionality is related to the random outages of generators and of transmission lines. For the most part, much of the attention in power system studies has been given only to dealing with the uncertainties in equipment status/functionality while ignoring the regulatory uncertainties and the market designs. In this thesis it is asserted that the uncertainties in all three categories play equally critical roles in determining systemwide efficiency, the reliability of the operation and the planning of an electric power network in the new environment.

There are many ways of dealing with the above uncertainties. One possibility analyzed in the thesis is as follows. In dealing with regulatory uncertainties, the price-cap regulation (PCR) scheme is suggested to replace the traditional cost-of-service regulation. The PCR scheme has one main advantage over the cost-
of-service regulation, namely the transfer of the decision making process from the regulator to the regulated firm.

Under cost-of-service regulation the regulated firm first proposes to the regulator the necessary investment to be made. Upon receiving the proposal, the regulator decides on the prudence of the proposed plan of investment and either approves or rejects the plan. When the investment is approved, the price of the product by the regulated firm is adjusted as necessary while the regulator guarantees the regulated firm a reasonable rate of return on the investment. Therefore, under cost-of-service regulation, it is evident that the ultimate authority for investment lies with the regulator and not with the regulated firm. Furthermore, any investment decision is subject to the uncertainty of the regulation process.

In contrast, under the PCR scheme the regulator initially sets a reasonable price cap to be imposed on the services being provided by the regulated firm. This price cap is then allowed to evolve over time, based on some formula agreed upon between the regulator and the regulated firm ahead of time. As long as the services are provided below the agreed upon price cap, the regulated firm is autonomous in making any necessary operational and planning decisions. Thus, much of the uncertainties of the regulatory process are eliminated under the PCR scheme.

The regulated firm is a TP in the case of the electric power industry in the new environment. Due to the particular characteristics of the industry, including the lack of a practical means of storing electricity and the inability to directly control the transmission path, some modifications to the conventional form of PCR scheme are necessary before actual application. In this thesis a practicable PCR scheme is proposed for regulating the operation and planning of a TP, which establishes the framework necessary for creating a for-profit ITC.

In addition to reducing the regulatory uncertainties, it is asserted that the proposed PCR scheme also has the advantage over traditional cost-of-service regulation of creating proper incentives for expansion as initially suggested in [57], and thus deterring imprudent investments (and may have an advantage in fostering new technology).

In dealing with the uncertainties in market designs, the thesis introduces a usable business model for an ITC and the transmission products to be provided by this ITC. A dynamic programming formulation for examining the investment decisions by the ITC is needed in order to capture the uncertainties in network usage because the supply and demand of the network users are constantly changing and evolving with time. However, many of these uncertainties can be eliminated through the ITC offering various transmission products. These products are long term transmission contracts and intermediate term transmission contracts.

The long term transmission contracts refer to any point-to-point network capacity offered at an increment of a year starting from the year following the current one. The intermediate term transmission contracts refer to the link-based network capacity design and are offered by the ITC and the system operator (SO), who may or may not be a part of the ITC, but together they make up the entire TP. By offering these products the ITC can enjoy the effect of information revelation on the supply and demand of network users. A liquid
market for these transmission contracts is essential for such information revelation. The thesis proposes a workable mechanism for designing a market for transmission.

Once the supply and demand can be known with a certain level of accuracy, there is a reduced need for projecting the market uncertainties, thus simplifying the mathematical formulation necessary for examining the planning by the ITC. Furthermore, it turns out that when the ITC makes the investment decisions based on the transmission capacity trading over the longer term and solidifies much of its future revenue stream, the volatility associated with the ITC’s profit is reduced as a side benefit. Finally, it is shown that the transmission products introduced in this thesis admit a proper placement of financial and operation risks to the ITC and to the market participants. The proposed PCR scheme is instrumental in order for the ITC to exist as a for-profit entity which has the incentives and the ability to offer these transmission products.

In dealing with the uncertainties in equipment status/functionality, the thesis only suggests a strong need for analytical tools in accurately computing the operational limits on power transfer through transmission lines within the network. The operational limits on transmission lines should be defined such that the existing electric power network can be operated without the loss of synchronism even under some plausible contingencies. Much work is needed for re-visiting the (short-term) reliability standards created under the old industry regime since various contractual agreements among the market participants now have a very different interpretation of how the uncertainties in equipment status/functionality need to be handled. Active future research is urged in order to define the market mechanisms essential for unbundling reliability in parallel to functional unbundling.
Chapter 2

Dynamics of the electric power network: reliability, pricing, and regulation

In this chapter we review a mathematical formulation of the operations and planning problem in an electric power industry composed of generation, transmission and load sectors. This formulation captures the various characteristics unique to the industry including the lack of practical means of storing electricity and the inability to transfer electricity directly through a particular path in the electric power network. These characteristics place very difficult constraints on the operations and planning of the network, and the effect of such constraints is discussed in detail in the chapter. Related to this, the concept of reliability is briefly reviewed and is shown to play a significant role in determining suitable operation and planning methods by defining the social welfare function based on the mathematical formulation developed here.

Following the formulation we make an assumption that there exist a high sunk cost, economies of scale, and economies of scope for the network and explain the intricate relationship between the pricing of electricity and the regulation of the monopolistic entity called the vertically integrated utility whose responsibility is to plan and operate the regional electric power network before the restructuring process. Cost-of-service regulation is commonly imposed to regulate the vertically integrated utility. Employing the same mathematical formulation as earlier, we assert that there is a significant long-term efficiency loss under this form of regulation if there is a significant asymmetry of information on technology when meeting the demand for electricity between the regulator and the utility. This assertion is already well recognized in regulatory economics [51].

Then, the changes in the industry brought on by the restructuring process are described, most notably the functional unbundling of the generation, transmission and load sectors. Due to the assumption made
earlier, the transmission sector remains a natural monopoly following the functional unbundling. It is emphasized that the preeminent purpose of the restructuring is to reduce the long-term efficiency loss. With the introduction of competition, long term efficiency gain is expected on the condition that a proper incentive structure is placed on the transmission sector. We defer to Chapter 3 for a detailed discussion of the proper incentive structure.

2.1 Full problem formulation - benign social planner

Consider the simplified model of an electric power network shown in Figure 2-1. A network is composed of

Figure 2-1: Typical electric power network consisting of generating substations, load centers and transmission lines

three basic elements representing three sectors in the industry: the generator or the generating substations (generation) that produce electricity; the network or transmission lines (transmission) that deliver power from the generators to the distribution system/loads\(^1\) covering a wide geographical region; and the distribution system or the load centers (load) that consume electricity. A bus refers to the place where either a generating substation or a load center is connected to the network through transmission lines.

When a consumer of the load center at bus \(d_j\) consumes electricity of the amount \(Q_{d_j}\) at hour \(k\), the consumer satisfies a certain level of utility, \(U_{d_j}\). When a different amount of electricity, \(Q'_{d_j}\), at hour \(k\) is consumed, we expect that a certain other level of utility is achieved by the consumer or when the same amount of electricity \(Q_{d_j}\) is consumed but at hour \(h\) different from hour \(k\), then again we expect that a certain other level of utility is achieved by the consumer. Thus, the behavior of the consumer at bus \(d_j\) can

\(^1\)When only the bulk transmission network is considered, we make no distinctions between the load centers, which consume electricity, and the distribution systems, which deliver electricity to the actual consumers.
be described using its utility function as the following:

\[ U_{d_j} = U_{d_j}(Q_{d_j}[k], k) \] (2.1)

Typical of electric power industry practice, an assumption is made about the utility function in Eq. (2.1) that, although the utility function of a consumer may change hour-by-hour, it remains constant within the same hour.

Suppose a generator of capacity, \( K_{g_i} \), at bus \( g_i \) produces electricity of the amount \( Q_{g_i} \) at hour \( k \). Then, the generator incurs an operating cost of \( c_{g_i} \). Typically, the operating cost is given as the following [1]:

\[
c_{g_i}(x_{g_i}[k], u_{g_i}[k], Q_{g_i}[k], k) = u_{g_i}[k](c_{g_i, \phi}(Q_{g_i}[k]) + I(x_{g_i}[k] < 0)S_{g_i}) + (1 - u_{g_i}[k])(c_{g_i,f} + I(x_{g_i}[k] > 0)T_{g_i})
\] (2.2)

where

\( x_{g_i}[k] \): the status of the generator indicating the number of consecutive hours that the generator has been on (or off) 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 when the generator is off

\( I(\cdot) \): a conditional statement that 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

The status of the generator, \( x_{g_i}[k] \), is then determined by the following [1]:

\[
x_{g_i}[k + 1] = \begin{cases} 
\max(1, x_{g_i}[k] + 1) & : u_{g_i}[k] = 1 \\
\min(-1, x_{g_i}[k] - 1) & : u_{g_i}[k] = 0 
\end{cases}
\] (2.3)

Because for most generators, once the unit is turned on (or off), it must remain on (or off) for a number of hours before it can be turned off (or on), we impose the following constraints:

\[
u_{g_i}[k + 1] = u_{g_i}[k] \text{ if } t_{g_i, dn} < x[k] < t_{g_i, up}
\] (2.4)

where \( t_{g_i, dn} \) and \( t_{g_i, up} \) are the minimum down time and the minimum up time respectively. Eq. (2.2) models closely the typical operating cost of a generating unit along with the constraints in Eqs. (2.3) and (2.4) while ignoring the ramping up and ramping down constraints.

At any hour, \( k \), the economics of an electric power network can then be described using Eqs. (2.1) and (2.2) along with the existing capacities in generation, \( K_{g_i}[k] \), at bus \( g_i \) and in transmission, \( K^T_{l}[k] \), at line \( l \),
and various operational constraints to be defined later, in the sense of static equilibrium at which the quantity of electricity demanded equals the quantity supplied within the capacity. However, in order to describe the economics of the network in the sense of an equilibrium at which the evolving quantity of electricity demanded equals the quantity supplied, we also need to consider the unit commitment, maintenance and investment.

As the system-wide demand for electricity increases, the supply of generation as well as transmission needs to increase commensurably in order to meet the demand efficiently. Suppose a generation capacity of \( G_{g_i}^C[k] \) at bus \( g_i \) is added at hour \( k \). Then, the current capacity of generation, \( K_{g_i}^C[k] \), changes according to the following:

\[
K_{g_i}^C[k + 1] = (1 - \rho_G)K_{g_i}^C[k] + I_{g_i}[k]
\]  

(2.5)

where \( \rho_G \) is the depreciation rate of the generation capacity [54] [55]. This increase in generation capacity comes at the expense of the investment cost, \( C_G^G(K_{g_i}^C[k], I_{g_i}^G[k], k) \). Similarly, an investment of transmission capacity, \( I_l^T[k] \), changes the thermal limit/rating of transmission line \( l \) at hour \( k \) to the following:

\[
K_l^T[k + 1] = (1 - \rho_T)K_l^T[k] + I_l[k]
\]  

(2.6)

at the cost of \( C_l^T(K_l^T[k], I_l^T[k], k) \) where \( \rho_T \) is the depreciation rate of the transmission capacity [54] [55]. The thermal rating of a transmission line refers to the maximum loading of electricity on the line specified by the manufacturer without compromising the integrity of the line. We emphasize the words, thermal limit, in order to differentiate this from the actual limit set by the system operator, usually much lower.

Whereas the major part of the operating cost of the generator is the production cost as described in Eq. (2.2), the operating cost of the transmission for the entire network is limited to the control cost, \( \nu_{tech}(e_{tech}[k], k) \) as a function of the control effort, \( e_{tech}[k] \), and the maintenance cost, \( \nu_m(e_m[k], k) \), as a function of the maintenance effort, \( e_m[k] \). We defer the discussion on the significance of \( \nu_{tech} \) and \( \nu_m \) to the later sections.

Having defined the utility function for loads and the cost functions for generators and transmission lines, we can define the systemwide social welfare function, \( SW_{system}[k] \), at hour \( k \) as

\[
SW_{system}[k] = \sum_{d_j} U_{d_j}(Q_{d_j}[k], k) - \sum_{g_i} c_{g_i}(x_{g_i}[k], u_{g_i}[k], Q_{g_i}[k], k) - \sum_{g_i} C_G^G(K_{g_i}^C[k], I_{g_i}^G[k], k) \]
\[
- \sum_l C_l^T(K_l^T[k], I_l^T[k], k) - \nu_{tech}(e_{tech}[k], k) - \nu_m(e_m[k], k)
\]  

(2.7)

The idealistic modeling of electric power network economics as solved by a benign social planner, then, may be defined as the maximization of the systemwide social welfare function given in Eq. (2.7) with respect to the generation dispatch, the rate of investments in generation and in transmission and the control and

24
maintenance cost in the network for the anticipated demand, given as the following:

$$\lim_{T \to \infty} \arg \max_{(\mathbf{I}_{G[k]}, \mathbf{I}_{T[k]}, \mathbf{e}_{\text{tech}[k]}, \mathbf{e}_{m[k]}, \mathbf{u}_{G[k]}, \mathbf{Q}_{G[k]}, \mathbf{Q}_{D[k]})} \sum_{k=1}^{T} (1 - \xi)^k \mathcal{E} \{ SW_{\text{system}[k]} \}$$

where

- $\mathcal{E}\{\cdot\}$ : expectation operator
- $\xi$ : discount rate of risk-free investment
- $[\cdot]'$ : transpose matrix/vector of $[\cdot]$

In this formulation the optimization is performed over a very long time, i.e., $T \to \infty$, due to the intertemporal interactions between variables. Due to these interactions, the discount rate of risk-free investment is included in order to account for the present value of income as required in an intertemporal utility maximization problem [23]. The problem formulation given here is a generalization of the problem formulation given in [54], [55], and [75].

The optimization problem in Eq. (2.8) is not complete since the various uncertainties and constraints related to the problem have not been defined. For example, there are many operational constraints to be considered, as mentioned earlier, including the one given in Eq. (2.4). We describe those uncertainties and constraints next.

### 2.1.1 Identifying uncertainties and constraints in the systemwide social welfare maximization problem

The operation and planning of electric power systems is subject to various uncertainties evolving at different rates.

#### Uncertainties

We begin by identifying the two definitive uncertainties to be considered in solving the optimization problem given in Eq. (2.8).

The first type of uncertainty is not necessarily limited to the electric power industry but may also be intrinsic to other industries as well; it is related to the stochastic nature of the future value of utility and cost functions in Eq. (2.7). Nevertheless, it is noted that the utility function for electricity in Eq. (2.7) is much more indeterminant than for many other products due to the fact that the function is strongly correlated to highly uncertain weather and long-term macroeconomic trends. Although no specific form of functions describing these uncertainties is proposed in this thesis, it is recognized that there needs to be an inherent intertemporal effect in modeling these uncertainties [1], i.e.,

$$U_{d_j}(Q_{d_j[k], k}) = f_{U_{d_j}}(U_{d_j}(Q_{d_j[k-1], k-1}), \omega_{d_j[k]})$$

(2.9)
where \( \omega_{d_j}[k] \) is the uncertainty in the utility function of the load at bus \( d_j \) for hour \( k \) given the utility function of the previous hour, \( U_{d_j}(Q_{d_j}[k-1], k-1) \). Incidentally, since the uncertainties in utility functions are much greater than in cost functions, it is often customary to neglect the modeling of uncertainties in cost functions.

The second type of uncertainty is related to the equipment outages particular to the industry. The frequent outages of generators and the rare outages of transmission lines must be weighed carefully in the operations and planning of the electric power network. The modeling of these uncertainties again involves an understanding of the intertemporal effect but can be much more cumbersome than the modeling of the first type of uncertainty [61]. Fortunately, by limiting the overall network to operate only within certain system constraints, the second type of uncertainty may be ignored in considering the operation and planning of the electric power network which evolve in the time scales slower than an hour. This is described further in the following section.

**Reliability-related uncertainties and constraints in the social welfare optimization**

Usually, well designed power systems do not experience significant constraints under the anticipated load demand patterns and when all equipment (generation plants and/or transmission lines) is in service. The system constraints become much more of an issue when the equipment is out of service for planned maintenance or due to an unexpected technical problem.

The overall objective of operating and planning electric power systems can be thought of as optimizing that social welfare function defined in Eq. (2.8) subject to the technical constraints resulting from the uncertainties. The problem of reliable operation and planning then becomes related to the decision making which ensures that supply meets demand under uncertainties.

The overall efficiency (measured by Eq. (2.8)) is generally strongly affected by how the uncertainties are managed. Therefore, for the purposes of this thesis, it is important to recognize that reliability management is closely related to uncertainty management. This provides an important link between the market instruments evolving in the electric power industry under restructuring, reliable operations and/or planning and the financial instruments in place to assess and value reliability-related risks. The hidden differences between various market designs regarding rules, rights and responsibilities can often be attributed to how the reliability-related risks are managed. We argue later in this thesis that subtle distinctions between proposed financial instruments for delivery services, proposed throughout the literature, can be accounted for and understood only in the light of uncertainty management and the related risks to the system providers and/or system users. At the carefully defined equilibrium, void of uncertainties, the proposed financial mechanisms are often equivalent, as we describe in Chapter 6 of this thesis.

It is at this point in the thesis that one needs a careful understanding of the interdependence between technical assumptions made and their relation to the market mechanisms. For the first time in this thesis, this issue is stressed and clarified. Unfortunately, it turns out, as we describe later, that solving the social
welfare optimization under uncertainties requires solving a very major dynamic programming problem. Even establishing actual operating limits on individual pieces of equipment, as operating conditions and the status of equipment changes, is a major problem for which no general solution is readily available. For this reason, the simpler problem of optimizing social welfare for given operating limits on individual pieces of equipment, transmission lines in particular, is attempted in this thesis. The risks and opportunities related to defining the operating limits are, therefore, not studied in this thesis.

**Characteristics unique to the electric power systems and their effects on the optimization problem in Eq. (2.8)**

Beside the uncertainties there are several characteristics unique to the industry which need to be considered carefully in optimizing the social welfare function given in Eq. (2.8), most notably the lack of a practical means of storing electricity and the inability to direct the transfer of electricity along a particular path in the electric power network. Combined with the uncertainties described above, these characteristics yield the following constraints to the optimization problem. Because the electricity cannot be stored, the amount of generation and the load need to be matched at each hour $k$, i.e.,

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

while

$$
Q_{g_i}^{\text{min}}[k] \leq Q_{g_i}[k] \leq K^{\text{Q}}_{g_i}[k] : \eta_{g_i}[k]
$$

where $\lambda[k]$ and $\eta_{g_i}[k]$ are the Lagrangian multipliers accompanying the above two constraints often referred to as the load flow constraints. In addition, the electric power flows through the lines are limited in transfer capacity, i.e.,

$$
F_l(Q_{G}[k], Q_{D}[k]) \leq F_l^{\text{max}}(F[k], K_l[k], e_{\text{tech}}[k], e_{m}[k]) : \mu_l[k]
$$

where

$$
F_l(Q_{G}[k], Q_{D}[k]) : \text{electric power flow through line } l \text{ as a function of the dispatch in generation and load at hour } k
$$

$$
F_l^{\text{max}}(F[k], K_l[k], e_{\text{tech}}[k], e_{m}[k]) : \text{operational limit on power transfer through line } l \text{ as a function of operating conditions (generation and demand), } F[k], \text{ the thermal rating on the line, } K_l[k], \text{ the control effort, } e_{\text{tech}}[k], \text{ and the maintenance effort, } e_{m}[k]
$$

$$
\mu_l[k] : \text{the Lagrangian multiplier accompanying the above constraint}
$$

The power flow on each line $l$, $F_l$, in Ineq. (2.12) is expressed as a function of generation dispatch and load demand rather than as an independent variable due to the inability to direct the transfer of electricity along a particular path in the electric power network. Because the transfer capacity needs to be respected at
each hour $k$ even in the event of equipment outages, the flow through each transmission line is constrained much more conservatively than the thermal limit. For example, consider the two identical transmission lines connecting buses 11 and 17 as shown in Figure 2-2. Suppose the thermal rating on the lines is specified to

\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{figure2-2}
\caption{Illustrative example of operational and thermal limits on the power transfer through transmission lines}
\end{figure}

be 100MW by the manufacturer, and sometimes, though rarely, one of the lines is not functional for the duration of less than an hour at a time. Then, the operational limits on these lines are 50MW assuming no corrective actions are available since, as soon as one line becomes unavailable, the other line needs to carry its own load as well as the load from the line that is out. Thus, $F_{i}^{\text{max}}(F[k], K_{1}[k], \epsilon_{\text{tech}}[k], \epsilon_{\text{m}}[k]) \leq K_{1}[k]$ and sometimes, $F_{i}^{\text{max}}(F[k], K_{1}[k], \epsilon_{\text{tech}}[k], \epsilon_{\text{m}}[k]) \ll K_{1}[k]$. By internalizing the effect of the equipment outages into the loading limits on transmission lines, we can disregard the uncertainties related to equipment outages (or the issues related to short term reliability) for the rest of the thesis.
By incorporating the uncertainties and constraints into the operation and planning of the electric power network we can present a complete formulation of the optimization problem to be solved by a benign social planner as the following:

\[
\left[ I_G^+, I_T^+, e_{tech}^+, e_m^+, uG^+, Q_G^+, Q_D^+ \right]^T = \lim_{T \to \infty} \arg \max_{I_G[k], I_T[k], e_{tech}[k], e_m[k]} \sum_{k=1}^{T} (1 - \xi)^k \mathcal{E} \{ SW_{system}[k] \}
\]

where

\[
SW_{system}[k] = \sum_{d_j} U_{d_j}(Q_{d_j}[k], k) - \sum_{g_i} c_{g_i}(x_{g_i}[k], u_{g_i}[k], Q_{g_i}[k], k) - \sum_{g_i} G_{g_i}^G(K_{g_i}^G[k], I_{g_i}^G[k], k)
\]

\[
- \sum_{l} C_l^T(K_l^T[k], I_l^T[k], k) - v_{tech}(e_{tech}[k], k) - v_m(e_m[k], k)
\]

\[
x_{g_i}[k + 1] = \begin{cases} 
\max(1, x_{g_i}[k] + 1) & : u_{g_i}[k] = 1 \\
\min(-1, x_{g_i}[k] - 1) & : u_{g_i}[k] = 0
\end{cases}
\]

\[
K_{g_i}^G[k + 1] = (1 - \rho)K_{g_i}^G[k] + I_{g_i}[k]
\]

\[
K_l^T[k + 1] = (1 - \rho)K_l^T[l] + I_l[k]
\]

\[
U_{d_j}(Q_{d_j}[k], k) = f_{U_{d_j}}(U_{d_j}(Q_{d_j}[k - 1], k - 1), \omega_{d_j}[k])
\]

subject to

\[
I_{g_i}[k] \geq 0
\]

\[
I_l[k] \geq 0
\]

\[
e_{tech}[k] \geq 0
\]

\[
e_m[k] \geq 0
\]

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

\[
\sum_{g_i} Q_{g_i}[k] = \sum_{d_j} Q_{d_j}[k] : \lambda[k]
\]

\[
Q_{g_i}^{min}[k] \leq Q_{g_i}[k] \leq K_{g_i}^G[k] : \eta_{g_i}[k]^2
\]

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

It is implied that \( Q_{g_i}[k] = 0 \) if \( u_{g_i}[k] = 0 \) without defining it explicitly as a constraint. The first four

\[3\] There are also maximum and minimum limits associated with \( Q_{d_j}[k] \), but for the ranges of operating conditions considered in this thesis, typically those limits play no role and are, thus, ignored here.
constraints in this formulation state that any investment to be made is non-negative. In the following section we recognize that this optimization problem is indeed a dynamic programming (DP) [11] problem.

### 2.1.2 Applying dynamic programming technique to the systemwide social welfare maximization problem

The systemwide social welfare maximization problem for an electric power network can be expressed as a dynamic programming (DP) problem which includes state variables, control variables and uncertain (random) variables [11]. For a given network the number of generators, transmission lines and loads are denoted by $N_G$, $N_T$ and $N_D$ respectively.

We, first, define the state variables, $x_k$ as the following:

$$x_k = [x_{1_k}, x_{2_k}, x_{3_k}, x_{4_k}]'$$  \hspace{1cm} (2.27)

where

$$x_{1_k} = K^G[k]$$  \hspace{1cm} (2.28)

$$= [K^G_{g_1}[k], K^G_{g_2}[k], \cdots K^G_{g_{N_G}}[k]]'$$

$$x_{2_k} = K^T[k]$$  \hspace{1cm} (2.29)

$$= [K^T_1[k], K^T_2[k], \cdots K^T_N_T[k]]'$$

$$x_{3_k} = x_G[k]$$  \hspace{1cm} (2.30)

$$= [x_{g_1}[k], x_{g_2}[k], \cdots x_{g_{N_G}}[k]]'$$

$$x_{4_k} = U_D[k]$$  \hspace{1cm} (2.31)

$$= [U_{d_1}[k], U_{d_2}[k], \cdots U_{N_D}[k]]'$$

We, then, define the control variables, $u_k$ as the following:

$$u_k = [u_{1_k}, u_{2_k}, u_{3_k}, u_{4_k}, u_{5_k}, u_{6_k}, u_{7_k}]'$$  \hspace{1cm} (2.32)

where

$$u_{1_k} = I_G[k]$$  \hspace{1cm} (2.33)

$$= [I^G_{g_1}[k], I^G_{g_2}[k], \cdots I^G_{g_{N_G}}[k]]'$$

$$u_{2_k} = I_T[k]$$  \hspace{1cm} (2.34)

$$= [I^T_1[k], I^T_2[k], \cdots I^T_{N_T}[k]]'$$

$$u_{3_k} = e_{tech}[k]$$  \hspace{1cm} (2.35)

$$u_{4_k} = e_m[k]$$  \hspace{1cm} (2.36)
\begin{align*}
\text{u5}_k &= \text{u}_G[k] \\
&= [\text{u}_{g_1}[k], \text{u}_{g_2}[k], \cdots \text{u}_{g_{NG}}[k]]' \\
\text{u6}_k &= \text{Q}_G[k] \\
&= [\text{Q}_{g_1}[k], \text{Q}_{g_2}[k], \cdots \text{Q}_{g_{NG}}[k]]' \\
\text{u7}_k &= \text{Q}_D[k] \\
&= [\text{Q}_{d_1}[k], \text{Q}_{d_2}[k], \cdots \text{Q}_{d_{ND}}[k]]'
\end{align*}

Finally, we define the random variables, \( w_k \) as the following:

\begin{align*}
\text{w}_k &= [\text{w}_1[k] \\
&= [\omega_{d_1}[k], \omega_{d_2}[k], \cdots \omega_{d_{ND}}[k]]'
\end{align*}

where \( \text{w}_1[k] \) is the modeling of uncertainties related to the utility functions of individual loads. Using the newly defined variables, \( x, u, w \) we can rewrite the social welfare function in Eq. (2.14) as

\[ SW_{\text{system}}[k] = g_k(x_k, u_k, w_k) \]

Similarly, using the newly defined state variables, \( x_k \) we can reduce Eqs. (2.15) through (2.18) to

\[ x_{k+1} = f(x_k, u_k, w_k) \]

where

\begin{align*}
x_{1k+1} &= (1 - \rho_G)x_{1k} + u_{1k} \\
x_{2k+1} &= (1 - \rho_T)x_{2k} + u_{2k} \\
x_{3k+1}(i) &= \begin{cases} 
\max(1, x_{3k}(i) + 1) : & u_{5k}(i) = 1 \\
\min(-1, x_{3k}(i) - 1) : & u_{5k}(i) = 0
\end{cases} \\
x_{4k+1} &= f_U(x_{4k}, w_{1k})
\end{align*}

where \( i = g_1, g_2, \cdots g_{NG} \) and \( x_{3k}(i) \) (and \( u_{5k}(i) \)) denotes the \( i \)th element of \( x_{3k} \) (and \( u_{5k} \), respectively), and

Finally, we can rewrite the optimization problem defined in Eq. (2.13) using the newly defined variables, as the following:

\[ J^*_\pi(x_0) = \lim_{T \to \infty} \max_{\pi \in \Pi} \mathcal{E} \left\{ \sum_{k=1}^{T} (1 - \xi)^k g_k(x_k, \nu_k(x_k), w_k) \right\} \]

\[ k = 1, \cdots, T \]
where \( \nu_k \) defines the control law which maps states, \( x^k \), into controls, \( u_k \), i.e.,

\[
 u_k = \nu_k(x^k) \tag{2.48}
\]

and is subject to the admissible control laws denoted as \( U_k(x_k) \). The admissibility is defined by

\[
 u_{1_k} \geq 0 \tag{2.49}
\]

\[
 u_{2_k} \geq 0 \tag{2.50}
\]

\[
 u_{3_k} \geq 0 \tag{2.51}
\]

\[
 u_{4_k} \geq 0 \tag{2.52}
\]

\[
 u_{5_{k+1}}(i) = u_{5_k}(i) \text{ if } t_{g_i, dn} < x_{3_k}(i) < t_{g_i, up} \tag{2.53}
\]

\[
 1'u_{6_k} + 1'u_{7_k} = 0 : \quad \lambda_k \tag{2.54}
\]

\[
 Q_{G_{\min}} \leq u_{6_k} \leq x_{1k} : \quad \eta_k \tag{2.55}
\]

\[
 F_l(u_{6_k}, u_{7_k}) \leq F_l^{\max}(u_{3_k}, u_{4_k}, u_{6_k}, u_{7_k}, x_{2_k}) : \quad \mu_k \tag{2.56}
\]

as defined in Eqs. (2.19) through (2.26). \( \pi \) (and \( \Pi \)) denotes the sequence of functions of the control laws, i.e.,

\[
 \pi = \{ \nu_1, \nu_2, \cdots \} \tag{2.57}
\]

(and the sequence of admissible control laws).

With the above transformations, Eqs. (2.41), (2.42) and (2.47), the systemwide social welfare maximization problem is now ready to put into standard DP formulation. It is recognized that in [19] the public utility problem is formulated as a stochastic DP problem. This is further carried out in [54] and [75] to include the constraints in the state variables. The formulation here also includes unit commitment constraints and the value of \( e_{tech}[k] \) and \( e_m[k] \). With these constraints, there is a stronger intertemporal interplay among variables, and the problem now must be solved over a very long time. The problem, however, is not solvable in the current form because as \( T \to \infty \) the number of time steps and the number of states blows up to infinity. We employ the heuristic justification from the economics standpoint and reduce the problem in a finite stage (and thus, a finite state) DP problem. This formulation is given in the following section.
2.1.3 Solving the systemwide social welfare maximization problem and the need for regulation in the actual operation and planning of an electric power network

In the typical operation and planning of an electric power network, the control variables connected with the planning, namely \( u_{1k}, u_{2k}, u_{3k} \) and \( u_{4k} \) in Eq. (2.32), are subject to additional restrictions due to a few practical considerations. The first practical consideration is related to the fact that once the investment decisions are made, it takes several months to a few years before the newly installed capacities become functional. Thus, there is a minimum time scale on which the decisions can be made. For simplicity without the loss of generality, we define the typical time scales for an investment in generation, \( T_G \), and in transmission, \( T_T \), as a season (typically 3 months) and a year, respectively.\(^3\)

The second practical consideration is related to the fact that the valuation of investment is performed over a finite number of years. In the case of the electric power industry, the valuation of investment is made usually over 20 years since the recovery of investment cost in either generation or transmission is expected to be made over this period of time. Thus, we define a couple of new time scales for the recovery of investment, \( T_R \) (typically \( T_R = 20T_T \)), and for the investment evaluation, \( T_I \) \((T_I = 3T_R)\); \( T_I \) functions as an artificial termination stage in applying the DP technique.

Having minimum time scale restrictions imposes an additional constraints on the control variables. This is that the investment in generation (and in transmission) is made only once every season (and year, respectively) up to the second recovery of investment period since only the investment made up to that period influences the investment in the first period.

The last major practical consideration is related to the fact that the investments in generation and especially in transmission are lumpy in nature and have upper limits. Typically the investment decision cannot be made in an increment less than 0.5 to 5MW up to 500MW for generation and 5 to 10MW up to 1000MW for transmission. For example, we may assume that the investment in generation can be made in increments of 10MW up to 50MW and in transmission in increments of 20MW up to 100MW, for simplicity. The control and maintenance efforts are decided as either made \((e(c) = 1)\) or not made \((e(c) = 0)\).

Taking these considerations into account we have the systemwide social welfare maximization problem as a standard finite stage (and finite state) DP problem as the following:

\[
J^*_n(x_0) = \max_{\pi \in \Pi} \mathcal{E}
\left\{ g_{T_I}(x_{T_I}) + \sum_{k=1}^{T_I-1} (1 - \xi)^k g_k(x_k, \nu_k(x_k), w_k) \right\}
\]

\[k = 1, \ldots, T_I - 1\]

\(^3\)Although there are no actual minimum time scale restrictions on \( u_{3k} \) and \( u_{4k} \), we assume that the decisions about control effort and maintenance effort are made on the same time scale as the investment in transmission. The reason for this assumption is related to the budgetary consideration of the transmission provider (TP) as described in Chapter 3.
where the terminal cost is now given by

\[ g_{T_f}(x_{T_f}) = 0 \]  

(2.59)

the state transition is according to

\[ x_{k+1} = f(x, u, w) \]  

(2.60)

t.e.,

\[ x_{1k+1} = (1 - \rho_G)x_{1k} + u_{1k} \]  

(2.61)

\[ x_{2k+1} = (1 - \rho_T)x_{2k} + u_{2k} \]  

(2.62)

\[ x_{3k+1}(i) = \begin{cases} 
\max(1, x_{3k}(i) + 1) & : \ u_{5k}(i) = 1 \\
\min(-1, x_{3k}(i) - 1) & : \ u_{5k}(i) = 0 
\end{cases} \]  

(2.63)

\[ x_{4k+1} = f_U(x_{4k}, w_{1k}) \]  

(2.64)

and the control laws are subject to

\[ u_{1k} \geq 0 \]  

(2.65)

\[ u_{2k} \geq 0 \]  

(2.66)

\[ u_{3k} \geq 0 \]  

(2.67)

\[ u_{4k} \geq 0 \]  

(2.68)

\[ u_{5k+1}(i) = u_{5k}(i) \text{ if } t_{g_{i, da}} < x_{3k}(i) < t_{g_{i, up}} \]  

(2.69)

\[ 1' u_{6k} + 1' u_{7k} = 0 : \quad \lambda_k \]  

(2.70)

\[ Q_G^{\min} \leq u_{6k} \leq x_{1k} : \quad \eta_k \]  

(2.71)

\[ F_1(u_{6k}, u_{7k}) \leq F_1^{\max}(u_{3k}, u_{4k}, u_{6k}, u_{7k}, x_{2k}) : \quad \mu_k \]  

(2.72)

\[ \pi = \{ \nu_1, \nu_2, \cdots \} \]  

(2.73)

\[ u_{1k} = \begin{cases} 
0, 10, \cdots , 50 & \text{if and only if } k = mT_G, \ m = 1, 2, \cdots , \frac{2T_G}{T_G} \\
0 & \text{otherwise}
\end{cases} \]  

(2.74)

\[ u_{2k} = \begin{cases} 
0, 20, \cdots , 100 & \text{if and only if } k = nT_T, \ n = 1, 2, \cdots , \frac{2T_T}{T_T} \\
0 & \text{otherwise}
\end{cases} \]  

(2.75)

\[ u_{3k} = \begin{cases} 
0, 1 & \text{if and only if } k = nT_T, \ m = 1, 2, \cdots , \frac{2T_T}{T_T} \\
0 & \text{otherwise}
\end{cases} \]  

(2.76)
\[ u_{4k} = \begin{cases} 
0, 1 & \text{if and only if } k = nT_T, m = 1, 2, \cdots, \frac{2T_T}{T_T} \\
0 & \text{otherwise} 
\end{cases} \] (2.77)

The more familiar expression of the DP algorithm for the problem in Eq. (2.58) is written as the following [11]:

\[ J_{T_1} = g_{T_1}(x_{T_1}) \] (2.78)

\[ J_k(x_k) = \max_{u_k \in U_k(x_k)} \mathcal{E} \{ g_k(x, u, w) + J_{k+1}(f_k(x, u, w)) \} \] (2.79)

such that

\[ J_0(x_0) = J^*_0(x_0) \text{ from Eq. (2.58)} \] (2.80)

However, there are a couple of difficulties in applying this problem formulation directly to the real life operation and planning of an electric power network. The first is rather straightforward and is related to the computational complexity of the problem. The second is not readily understood and has to do with the high sunk cost, and the economies of scale assumed to be inherent in an electric power network [57]. We discuss these difficulties in more detail.

It is generally well recognized that the DP algorithm is applicable to many stochastic optimization problems. However, it also has the disadvantage of a non-polynomial (NP) growth of operation count with respect to problem size [11], as in the case of the optimization problem here. Even before considering the uncertainties, \( w_{1k} \), and without accounting for the control variables related to the investment decisions in generation and transmission, \( u_{1k}, u_{2k}, u_{3k}, u_{4k} \), the computational complexity of solving the optimization problem in Eq. (2.58) is expressed, in terms of the total number of states, as the following [1]:

\[ \prod_{g_i}(t_{gi, up} + t_{gi, dn}) \]

Suppose the uncertainties are considered in discretized steps of sizes, \( \Delta_w 1, \Delta_w 2, \) and \( \Delta_w 3 \) encompassing 5 standard deviations above and below the initial expected values of each variable through the normal distribution assumption mentioned earlier. Then this adds to the total number of states at stage \( k \) by

\[ (10k\sigma_{w1}(\Delta_w)^{-1} + 1)^{N_G} \cdot (10k\sigma_{w2}(\Delta_w)^{-1} + 1)^{N_T} \cdot (10k\sigma_{w3}(\Delta_w)^{-1} + 1)^{N_G} \times \]

\[ (10k\sigma_{w4}(\Delta_w)^{-1} + 1)^{N_D} \cdot \prod_{g_i}(t_{gi, up} + t_{gi, dn}) \]

Finally, when the investment decisions about generation and transmission are accounted for, the total number of states is given by

\[ 2 \cdot \left( \frac{N_G}{N_T} \right) \cdot (6 \left[ \frac{N_G}{N_T} \right] + 1)^{N_G} \cdot (6 \left[ \frac{N_G}{N_T} \right] + 1)^{N_T} \cdot (10k\sigma_{w1}(\Delta_w)^{-1} + 1)^{N_G} \cdot (10k\sigma_{w2}(\Delta_w)^{-1} + 1)^{N_T} \times \]

\[ (10k\sigma_{w3}(\Delta_w)^{-1} + 1)^{N_G} \cdot (10k\sigma_{w3}(\Delta_w)^{-1} + 1)^{N_D} \cdot \prod_{g_i}(t_{gi, up} + t_{gi, dn}) \]
where \([\lfloor \cdot \rfloor]\) denotes the nearest integer less than \((\cdot)\). The simplifying assumptions in Eqs. (2.74) through (2.77) are used to obtain the coefficients of 2, 6 and 6 for the second, the third and the fourth terms in the expression. Given that typically \(T_f \equiv \mathcal{O}(10^6)\), \(N_G = \mathcal{O}(10^2)\), \(N_T = \mathcal{O}(10^3)\) and \(N_D = \mathcal{O}(10^2)\), the total number of states is not manageable even with the most potent computers available at the moment. A brief discussion of the reduction of computational complexity based on the function bundling of the electric power industry is deferred to later in the chapter and more in Chapter 3.

Even if the problem related to the computational complexity can be overcome somehow, the solution to the optimization problem in Eq. (2.58) is not likely to be realized in an actual electric power network, especially considering the investment in transmission under the high sunk cost and the economies of scale assumptions [57]. This is stated more formally in the following proposition.

**Proposition 1**: Suppose at the stage \(k = nT_T\) where \(n < 2(T_f/T_T)\), the optimal level of investment in transmission, \(u^*_nT_T\), is determined. Then, by invoking the *Principle of Optimality* in applying the DP technique [11], from Eq. (2.79), we have the following:

\[
 u^*_nT_T = \arg \max_{u^2_nT_T u^*_nT_T} E \left\{ g_{nT_T}(x^*_{nT_T}, \bar{u}_{nT_T}, w_{nT_T}) + J_{nT_T+1}(f_{nT_T+1}(x^*_{nT_T}, \bar{u}_{nT_T}, w_{nT_T})) \right\} \tag{2.81}
\]

where

\[
\begin{align*}
 u^1_{nT_T} &= u^1_{nT_T} \\
 u^3_{nT_T} &= u^3_{nT_T} \\
 u^4_{nT_T} &= u^4_{nT_T}
\end{align*}
\tag{2.82}
\]

Here \((\cdot)^*\) denotes the variables associated with the optimal solution of the complete problem in Eq. (2.79).

By expanding the right hand side of Eq. (2.81) we have

\[
 u^*_nT_T = \arg \max_{u^2_nT_T u^*_nT_T} E \left\{ g_{nT_T}(x^*_{nT_T}, \bar{u}_{nT_T}, w_{nT_T}) + \sum_{k=nT_T+1}^{T_f} g_k(x^*_k, u^*_k, w_k) \right\} \tag{2.83}
\]

Given that the optimal solution, \(u^*_k\), at the optimal state \(x^*_k\) except for \(u^*_nT_T\) is given for the problem in Eq. (2.83) for \(k = nT_T, nT_T + 1, \ldots, T_f\), \(u^*_nT_T\) can be obtained by applying the linear programming technique [13]. After constructing the Lagrangian equation and taking the first derivative of the equation with the convexity assumption discussed earlier, we have

\[
 \frac{dg_{nT_T}}{du^*_{nT_T}} \bigg|_{u^*_{nT_T}} = \sum_{k=T_T+1}^{T_f} \mu_k \tag{2.84}
\]

where \(\mu_k\) is the Lagrangian multiplier associated with the transmission line flow constraint in Ineq. (2.72)
By substituting $dC_l^T / dI_l^T$ for $dg/du2$ we can equate the individual elements in $u2$ as the following:

$$
\frac{dC_l^T}{dI_l^T} \mid_{(I_l^T)^*[k]} = \sum_{k=T_r+1}^{T_f} \mu_t[k]
$$

(2.85)

According to the result in Eq. (2.85) the pricing theory under perfect competition with free entry suggests that the pricing of the capacity of the transmission lines is equal to the marginal cost value suggested by $\sum_{k=T_r+1}^{T_f} \mu_k$ [60]. However, the effect of the high degree of economies of scale can be expressed as [54] [55]

$$
C_l^T((I_l^T)^*[k]) \geq \sum_{k=T_r+1}^{T_f} \mu_t[k]
$$

(2.86)

Thus, the solution to the optimization problem in Eq. (2.58) is not likely to be realized in an actual electric power network when there is an inherent assumption of a high degree of economies of scale concerning the investment in transmission.

Based on Proposition 1 it is suggested that when the high sunk cost and the economies of scale assumption concerning the investment in transmission holds true, the operation and planning of an electric power network cannot exist as a viable business under marginal cost pricing with perfect competition and free entry. Although in the future it may be possible to create higher (financial option-like) values for transmission with an increased uncertainty in energy, by considering the substitutability of transmission for generation in certain cases, for now some artificial mechanisms of marginal cost pricing are needed to achieve the systemwide optimum. Strict regulation currently provides such mechanisms. The benefit of the strict regulation on a single service provider is that it allows the provider to exist as a monopoly so that the prices for its services are not limited to the marginal cost pricing principle. In the following section, we examine the traditional regulation mechanism in the electric power industry, and an entity called the “vertically integrated utility” which operates as a service provider under this traditional regulation mechanism.

### 2.2 Operations and planning by a vertically integrated utility: problem formulation

Under the traditional regulation mechanism, the operation and planning of an electric power network is carried out by the three major decision makers shown in Figure 2-3. A vertically integrated utility is a monopolistic entity whose responsibility is to design and operate an electric power network. This includes generation and transmission investment. The regulator is typically a government agency whose responsibility is to approve the design and oversee the operation of the network by the vertically integrated utility. The loads are the consumers of the various electric services. In the following section we describe the relationship among these three players.
2.2.1 Defining the basic structure of the vertically integrated utility

As an entity which is responsible for designing and operating an electric power network, the utility carries out numerous functions. Figure 2-4 describes the planning aspect of the utility's function in the form of a flow chart. We look at each of the functions listed in Figure 2-4 in more detail.

(a) Project the demand for the next 5, 10 years The loads being served by the utility are often referred to as the native customers since they are bound to that particular utility by geographical location. Because the utility is required to provide electric services to all of its loads in a non-discriminatory way by the regulator, it needs to accurately project the demand of the loads over 5 to 10 years in advance. In
projecting the load, the demand is often assumed to be inelastic. Even if an accurate assessment can project the demand at each load bus by hour for the next 5 to 10 years, this is neither practical nor necessary. Usually, the demand is projected for the systemwide sum of individual loads corresponding to only several loading conditions: seasons (spring, summer, fall and winter), within the week (weekdays and weekends), within the day (peak hours, off-peak hours, shoulder-hours), etc. The uncertainty is then reflected directly in the uncertainty on the demand, in the case of inelastic loads.

(b) Plan for additional generation to meet the projected demand Based on the load projection, the utility plans for the additional generation necessary to meet the demands for the next 5 to 10 years. Several generation investment schemes, $I_G$, are drawn so that

$$P \left\{ \sum_{g_i} K^C_{g_i}[k] \leq \sum_{g_i} \dot{Q}_{d_i}[k] \right\} \leq \hat{P} \tag{2.87}$$

where $P \{\cdot\}$ denotes the probability of the event $\{\cdot\}$. The probability defined in Ineq. (2.87) is often referred to as the loss of load probability (LOLP)$^4$ [36]. The LOLP is required to be less than some level, $\hat{P}$, predetermined by the regulator and imposed as one of the planning standards that must be followed by the utility.

(c) Conduct off-line studies to ensure an acceptable level of reliability Related to the LOLP defined in Ineq. (2.87) the utility conducts numerous off-line reliability studies so that the probability is accurately computed by carefully accounting for several uncertainties in the operation and planning of an electric power network, as discussed earlier in the chapter. There are several indices associated with such off-line studies so that the reliability seen by the actual consumers is at an acceptable level. These are the customer average interruption index (CAIDI), the customer average interruption frequency index (CAIFI), and the expected value of energy not served (EENS). These indices serve as a strict planning standard for the utility [36].

(d) Plan network enhancement in order to accommodate the anticipated generation and demand Using the off-line reliability studies, the utility estimates the network enhancement necessary to accommodate the anticipated generation and demand for the next 5 to 10 years while meeting the planning standards set by the regulator. Several transmission network enhancement schemes, $I_T$, are devised as a result.

(e) Apply for rate adjustment for planned additions Once the essential generation additions, $I_G$, and the necessary transmission enhancement, $I_T$, are identified, the utility needs to apply for the rate

---

$^4$The probability given in Ineq. (2.87) is a highly simplified definition of LOLP since the actual computation of LOLP also takes various uncertainties such as outages into consideration.
adjustment at which its native customers are charged for the electric services in order to ensure the recovery of investment cost. The arrow numbered (3) leaving the planning box of the utility in Figure 2-4 is connected to the regulator as in Figure 2-3 because the regulator holds the authority to approve the proposed rate adjustment (thus the proposed generation and transmission investments). The approved rate sets the allowed revenue for the utility.

The regulator, in return, verifies the proposed investment in generation and in transmission by the utility in conformity with the predetermined planning standards. If the planned addition meets the standards with an economic efficiency judged by the regulator, the proposed rate is approved. The arrow numbered (4) in Figures 2-3 and 2-4 indicates this process.

Once the proposed rate adjustment is approved by the regulator, the utility goes ahead with the investment, and the modification/enhancement is made to the network accordingly as described by Eqs. (2.5) and (2.6).

Figure 2-5 describes the operational aspect of the utility’s function, after the investment made, in the form of a flow chart. In a similar fashion as before with the planning by the utility, we look at each of the functions listed in Figure 2-5 in more detail. Notice that an additional step needs to be made with respect to approving the right of ways for building. This step has become increasingly difficult, particularly for new transmission.

(a) Maintain generation and transmission Having the sole ownership of both the generation and transmission assets after the investment, the responsibility of maintaining the generation and transmission equipment falls fully on the shoulders of the utility. It is important to recognize that when the reliability
indices are computed, including LOLP, EENS, CAIDI and CAIFI, there is an implied assumption about the status of the various equipment in the network. Thus, in order to continue to meet the reliability standards, the utility must incur some expenses for maintenance, $v_m(c_m)$ (using the notation defined earlier in the chapter), where we assume for simplicity, $c_m = 1$ as a minimum. These expenses are a decision to be made at the beginning of each year.\(^5\)

(b) **Schedule and make generation unit commitment decisions**  The scheduling and the unit commitment are the process of deciding in advance whether to turn on or off each generator, $u_G$, on the given electric power network at a given hour. The utility needs to determine in a given hour which generator should be operating in order to meet the anticipated system load while conforming with the operating standards, including the Control Performance Standard (CPS1 and CPS2), and the Disturbance Control Standard (DCP) [81]. The CPS1 and the CPS2 specify that the average control area error (ACE), which measures the systemwide imbalance between the generation and load, needs to be less than a specific limit determined in advance. The DCP limits the ACE after a disturbance such as a generator outage or transmission line outage; the ACE must return to pre-disturbance level within 10 to 15 minutes following the disturbance [81]. Since generators cannot be turned on instantly and start producing power, the units need to be scheduled in advance, typically a week ahead, so that they are available to handle the system demand. This additional standby generation capacity is referred to as the reserve.

(c) **Adjust the network controllers in order to meet the reliability standards**  There are a number of network controllers [45] installed in an electric power network in order to improve the reliability of operations. These controllers range from more static devices such as on-load tap changers to more dynamic devices such as flexible AC transmission systems (FACTS) [26] [77]. Because these devices function properly only in a limited scope, they must be tuned for some anticipated loading conditions when generation dispatch scheduling is done. This step is critical along with the unit commitment in order to ensure the stability of the network operation. It is important to point out here that there are many innovative ways of ensuring reliability by adequately tuning the network related controllers at the expense of the control effort, $v(e_{tech})$, in lieu of incurring the cost related to the investment in transmission, $I_T$; the decision to utilize the network controllers, $e_{tech}$ is assumed to be made ahead of time at the beginning of each year [45].\(^6\)

(d) **Dispatch generation in order to meet the hourly demand for electric services by the load**

Once the unit commitment decisions are made, and all the network controllers are tuned so that the stability of the network is ensured, the committed generators are dispatched hourly in order to meet the demand for

\(^5\)The utility actually varies its effort for maintenance throughout the year, depending on the near real time operating conditions, in order to maximize its profit. However, this level of detail is beyond the scope of this thesis.

\(^6\)Analogous to $c_m$ the expense of control effort varies, depending on the near real time operating conditions, even if the decision to use advanced control is assumed to be made only once at the beginning of the year. However, this level of detail is beyond the scope of this thesis.
electricity by the loads [45]. As described earlier in the chapter, the amount of the generation dispatched, \( Q_G \), needs to match the amount of the demand by the loads, \( Q_D \), in order to keep the integrity of the system, without an adequate means of storing the electricity. The actual provision of electric services is performed under this step as indicated by the arrow numbered (6) in Figure 2-5 connecting to the loads as shown in Figure 2-3. The regulator keeps its watchful eyes on this function of the utility so that compliance with the operating standards is verified. This process is indicated with the arrow numbered (4) in Figure 2-5 connecting to the regulator as shown in Figure 2-3.

(e) Collect revenue and resolve any inadequacy or surplus: Upon the provision of electric services, the utility is expected to collect the revenue from the load according to the rate set by the regulator in the planning stage. If the demand for the revenue period (typically over a year) is realized as projected at the beginning of the period, then there should be no inadequacy or surplus of the revenue collected by the utility. However, if the actual load is different from the projected, or if the expected cost changes, then the difference in the revenue collected by the utility and the revenue allowed by the regulator needs to be made up on behalf of the utility. This is because unlike a competitive business entity, the monopoly, the vertically integrated utility in this case, is providing the service, electricity, on behalf of the regulator and in return is granted a guaranteed rate of return on its investment unless otherwise specified.\(^7\) If the revenue collected by the utility is greater than that allowed, the difference is rebated to the loads in a reduced rate for the electric services in the following revenue period. On the other hand, if the revenue collected is smaller than that allowed, the difference is made up in the following year through an increase in the rate at which the service is provided even when no investment is made to the network. For simplicity, however, this process can be modeled such that the difference is made up by the regulator on behalf of the customer as indicated by the arrow numbered (5) in Figure 2-5 connecting to the regulator as shown in Figure 2-3. The process of the actual rebating or taxing of the loads can be modeled as strictly between the regulator and the loads as shown through the arrows numbered (1) and (2) in Figure 2-3. In order to account for the reduced complication achieved through this simplified resulting model for any inadequacy or surplus in revenue, some weighing factor can be assigned to the process involving the regulator and the loads, as is done in [51].

2.2.2 Cost-of-service regulation as an optimization problem

The regulation scheme described in the previous section is referred to as cost-of-service regulation and is typically imposed on the vertically integrated utility. By modeling each process under this regulation with the mathematical notation defined earlier, we can construct the optimization problem relevant to the traditional mode of the operation and planning of the electric power network and compare it to the systemwide social

\(^7\)One form of regulation called price-cap regulation (PCR) under the performance-based regulation (PBR) does not guarantee the certain rate of return described in Chapter 3.
welfare maximization problem defined in Eq. (2.13). Here the financial interplay among the three players in Figure 2-3 is of particular interest.

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

$$Q_{d_j}^*[k] = \arg \max_{Q_{d_j}[k]} \mathcal{E}\{U_{d_j}(Q_{d_j}[k], k) - \rho \cdot Q_{d_j}[k]\} \quad (2.88)$$

where $\rho$ is the rate approved by the regulator and is the price paid by the load in order to consume the amount of electricity, $Q_{d_j}[k]$ at hour $k$. As before we make a convexity assumption about $U_{d_j}(Q_{d_j}[k], k)$. Then, there exists a unique solution to Eq. (2.88), and the vector form of the individual load, $\mathbf{Q_D}^*[k]$, i.e.,

$$\mathbf{Q_D}^*[k] = [Q_{d_1}^*[k], Q_{d_2}^*[k], \ldots Q_{d_N}^*[k]]^\prime \quad (2.89)$$

is a unique solution to

$$\mathbf{Q_D}[k]^* = \arg \max_{\mathbf{Q_D}[k]} \mathcal{E}\left\{\sum_{d_j} U_{d_j}(Q_{d_j}[k], k) - \rho \sum_{d_j} Q_{d_j}[k]\right\} \quad (2.90)$$

This is a result of the following lemma.

**Lemma 1:** Suppose that $\mathbf{x}^* = [x_1^*, x_2^*, \ldots, x_N^*]^\prime$ is a unique solution to a set of equations given as the following:

$$x_i^* = \arg \max_{x_i} \{f_i(x_i) - \rho_i \cdot x_i\} \quad (2.91)$$

where $\rho_i = \left.\frac{d g(x)}{dx_i}\right|_{\mathbf{x}^*}$, which is assumed to be given ahead of time, so that it is a constant.\(^8\) Further suppose that $\mathbf{x}^\dagger = [x_1^\dagger, x_2^\dagger, \ldots, x_N^\dagger]^\prime$ is a unique solution to the following:

$$\mathbf{x}^\dagger = \arg \max_{\mathbf{x}} \left\{\sum_i f_i(x_i) - g(\mathbf{x})\right\} \quad (2.92)$$

Then, $\mathbf{x}^* = \mathbf{x}^\dagger$.

**Proof:** Suppose $\mathbf{x}^* \neq \mathbf{x}^\dagger$. Then,

$$\sum_i f_i(x_i^*) - g(\mathbf{x}^*) < \sum_i f_i(x_i^\dagger) - g(\mathbf{x}^\dagger) \quad (2.93)$$

Otherwise, $\mathbf{x}^\dagger$ is not a unique solution to the optimization problem in Eq. (2.92). Then, there exists at least

\(^8\) This is a strange assumption here because $\rho_i$ is seemingly a function of $x_i$. However, this assumption is used only here to prove the lemma and becomes unnecessary once the constraints accompanying the problem are introduced, and it is replaced by the Lagrangian multiplier in the actual formulation later.
one such \( i \) that

\[
\frac{df_i(x_i)}{dx_i} \bigg|_{x_i^*} - \frac{dg(x)}{dx} \bigg|_{x^*} \neq 0
\]

Since \( f_i(x_i^*) - g(x^*) \) is not maximum, we have

\[
\frac{df_i(x_i)}{dx_i} \bigg|_{x_i^*} = \frac{dg(x)}{dx} \bigg|_{x^*} \neq 0
\]

However, this is a contradiction because \( x_i^* \) is an unique solution to the maximization problem in Eq. (2.91) because

\[
\frac{f_i(x_i)}{x_i} \bigg|_{x_i^*} = \rho_i
\]

\[
= \frac{dg(x)}{dx} \bigg|_{x^*}
\]

Therefore, \( x^* = x^\dagger \). The theoretical basis for this proof is given in [54] earlier.

From the perspective of the vertically integrated utility, the utility chooses the investments in generation and transmission, the amount of control and maintenance effort, and the generation output of each generator within the region to maximize its profit while following the steps outlined in Figures 2-4 and 2-5. The rate at which the utility charges for its electric services determines the allowed revenue, \( Y[n] \), for the entire year \( n \) and is traditionally based on the utility’s total investment cost, i.e.,

\[
Y[n] = (1 + r_{cos}) \sum_{k=(n-1)T_T+1}^{nT_T} (1 - \xi)^k \left( \sum_{g_i} C_{g_i}^G(K_{g_i}^G[k], I_{g_i}^G[k], k) + \sum_{l} C_{l}^T(K_{l}^T[k], I_{l}^T[k], k) \right)
\]

where \( r_{cos} \) is the allowed rate of return on investment. In Eq. (2.97) we use the fact that typically \( T_T = 1 \) year. The profit of the utility is, then, determined by

\[
\Pi_{utility}[n] = T[n] - \sum_{k=(n-1)T_T+1}^{nT_T} (1 - \xi)^k \left( \sum_{g_i} c_{g_i}(x_{g_i}[k], u_{g_i}[k], Q_{g_i}[k], k) + \sum_{g_i} C_{g_i}^G(K_{g_i}^G[k], I_{g_i}^G[k], k) + \sum_{l} C_{l}^T(K_{l}^T[k], I_{l}^T[k], k) + v_{tech}(e_{tech}[k]) + v_m(e_m[k]) \right)
\]

\[
= \sum_{k=(n-1)T_T+1}^{nT_T} (1 - \xi)^k \left[ r_{cos} \left( \sum_{g_i} C_{g_i}^G(K_{g_i}^G[k], I_{g_i}^G[k], k) + \sum_{l} C_{l}^T(K_{l}^T[k], I_{l}^T[k], k) \right) - \sum_{g_i} c_{g_i}(x_{g_i}[k], u_{g_i}[k], Q_{g_i}[k], k) - v_{tech}(e_{tech}[k]) - v_m(e_m[k]) \right]
\]

by substituting the expression in Eq. (2.97) for \( T[n] \). This problem formulation is a generalization of an earlier formulation in [54] after including the cost of the control effort and the maintenance effort. The profit
maximization of the vertically integrated utility leads to the following optimization problem:

\[
\left[ \mathbf{I}_G^* \right] = \mathbf{Q}_G^* = \arg \max_{\mathbf{I}_G[n], \mathbf{I}_T[n], \mathbf{I}_{tech}[n], \mathbf{e}_m[n], \mathbf{u}_G[n], \mathbf{Q}_G[n]} \sum_{n=1}^{5} \mathcal{E} \{ \Pi_{utility}[n] \} \tag{2.99}
\]

where \( \mathbf{Q}_G[n] \) (and similarly \( \mathbf{u}_G[n] \)) is defined as the following:

\[
\mathbf{Q}_G[n] = [Q_G[k = nT], Q_G[k = nT - 1], \cdots, Q_G[k = (n-1)T + 1]]
\]

\[
\mathbf{Q}_G[k] = [Q_{g_1[k]}, Q_{g_2[k]}, \cdots, Q_{g_{NG}[k]]}] \tag{2.100}
\]

The summation is from \( n = 1 \) to \( n = 5 \) based on the 5 year planning assumption as with block (a) in Figure 2-4.

According to blocks (a) through (e) in Figures 2-4 and 2-5, there is a natural time scale separation in the decision variables. On one hand, the decision variables related to the investment and the operation of the network, \( \mathbf{I}_G, \mathbf{I}_T, \mathbf{e}_{tech}, \) and \( \mathbf{e}_m \) are associated with the slow dynamics in the yearly time scale due to the practical considerations and simplifying assumption made earlier in the chapter. On the other hand, the decision variables related to the scheduling and dispatch of generation, \( \mathbf{u}_G \) and \( \mathbf{Q}_G[n] \), are connected to the fast dynamics in the hourly time scale. In addition, there is a natural time scale separation in the state variables. Consider the state variables, the generation capacity (\( \mathbf{K}^G[k] \)), the transmission capacity (\( \mathbf{K}^T[k] \)), the status of the generator (\( \mathbf{x}_G[k] \)) and the elements in utility functions associated with the uncertainties defined in Eqs. (2.28) through (2.31). Since the decision variables, \( \mathbf{I}_G, \mathbf{I}_T, \mathbf{e}_{tech}, \) and \( \mathbf{e}_m \), are determined in the slow time scale, we notice that the corresponding state variables, \( \mathbf{K}^G[k] \) and \( \mathbf{K}^T[k] \), also evolve in the same slow time scale because \( \mathbf{K}^G[k+1] = \mathbf{K}^G[k] \) unless \( \mathbf{I}_G[k] \neq 0 \) and similarly, \( \mathbf{K}^T[k+1] = \mathbf{K}^T[k] \) unless \( \mathbf{I}_T[k] \neq 0 \). Then, under these conditions the optimization problem in Eq. (2.99) can be separated into

\[
\left[ \mathbf{I}_G^* \right] = \arg \max_{\mathbf{I}_G[n], \mathbf{I}_T[n], \mathbf{I}_{tech}[n]} \sum_{n=1}^{5} \mathcal{E} \left\{ (1-\xi) \mathbf{n} \mathbf{r}_\mathbf{G}_\mathbf{G} \left( \sum_{g_i} \mathbf{C}_{g_i}^G \mathbf{K}_{g_i}^G[n], \mathbf{I}_{g_i}^G[n], n \right) + \sum_{n=1}^{5} \mathcal{E} \left\{ (1-\xi) \mathbf{n} \mathbf{r}_\mathbf{G}_\mathbf{G} \left( \sum_{g_i} \mathbf{C}_{g_i}^G \mathbf{K}_{g_i}^G[n], \mathbf{I}_{g_i}^G[n], n \right) \right\} \tag{2.101}
\]

where

\[
\mathbf{I}_{g_i}^G[n] = [I_{g_i}^G[k = nT], I_{g_i}^G[k = nT - T_G], \cdots, I_{g_i}^G[k = (n-1)T + T_G]]
\]

\[
\left[ \mathbf{u}_G^*, \mathbf{Q}_G^* \right] = \arg \max_{\mathbf{u}_G[n], \mathbf{Q}_G[n]} \mathcal{E} \left\{ - \sum_{k=1}^{7 \times 24} (1-\xi)^k \sum_{g_i} c_{g_i} \mathbf{x}_{g_i}[k], u_{g_i}[k], Q_{g_i}[k], k \right\} \tag{2.103}
\]
subject to

\[ \begin{cases} I_G[n] \\ I_T[n] \end{cases} \] Approval by the regulator \hspace{1cm} (2.104)

\[ \sum_{g_i} K_{g_i}[n] \geq \hat{K}_G[n]: \text{ according to block (b) in Figure 2-4} \] \hspace{1cm} (2.105)

\[ K_T^n[n] \geq \hat{K}_T[n]: \text{ according to block (d) in Figure 2-4} \] \hspace{1cm} (2.106)

\[ u_{g_i}[k+1] = u_{g_i}[k] \text{ if } t_{g_i, dn} < x[k] < t_{g_i, up} \] \hspace{1cm} (2.107)

\[ \sum_{g_i} Q_{g_i}[k] = \sum_{d_j} Q_{d_j}[k]: \lambda[k] \] \hspace{1cm} (2.108)

\[ 0 \leq Q_{g_i}[k = nT + mT_G + q] \leq K_{g_i}^G[n] + \varepsilon'_m \cdot I_{g_i}^G: \eta_{g_i}[k] \] \hspace{1cm} for \( m = 1, 2, \cdots (T_R/T_G) - 1 \), and for \( q = 1, 2, \cdots T_G - 1 \) \hspace{1cm} (2.109)

where \( \varepsilon_m \) is a vector of all zeros except for the ith element which is equal to 1

\[ F_i(Q_G[k], Q_D[k]) \leq F_i^{\text{max}}(F[k], K_T^n[n], e_{\text{tech}}[n], \varepsilon_m[n]): \mu[k] \] \hspace{1cm} (2.110)

It is interesting to note that the slow dynamics and the fast dynamics are linked through the constraints defined in Ineq. (2.109) and (2.110) since the supply of generation or transmission at any hour \( k \) is limited by the investment induced capacity.

Based on the optimization problem defined in Eq. (2.101) and the constraints in (2.104) it is recognized that the regulator plays a critical role in determining the level of investment. Under the cost-of-service regulation the regulator serves as a social planner and needs to solve an optimization problem similar to the one in Eq. (2.13), modified to include the effect of the guaranteed return on investment.

As described in the previous section, the guaranteed return on investment has an effect when there is a difference in the revenue collected by the utility and the revenue allowed by the regulator. We employ the modeling simplification of treating the process of making up the difference in the revenue collected and that allowed as an exclusive process between the regulator and loads as in [51]. The key notion is the shadow cost of public funds, \( \lambda_f \), in linking the actual process of making up the revenue to this modeling simplification.

The total cost to the regulator for year \( n \) using this model is given as the following [51]:

\[ TC_{\text{reg}}[n] = (1 + \lambda_f) \left( T[n] - \sum_{k=(n-1)T_R+1}^{nT_R} (1 - \xi)^k \sum_{d_j} \rho[n] \cdot Q_{d_j}[k] \right) \] \hspace{1cm} (2.111)

\[ = (1 + \lambda_f) \left( (1 + r_{\text{cos}}) \sum_{k=(n-1)T_R+1}^{nT_R} (1 - \xi)^k \left( \sum_{g_i} C_{g_i}^G(K_{g_i}^G[k], I_{g_i}^G[k], k) \right) \right) \]
\[ + \sum_{i} C_i^T (K_i^T [k], I_i^T [k], k) - \sum_{k=(n-1)T_T+1}^{nT_T} (1 - \xi)^k \sum_{d_j} \rho [n] \cdot Q_{d_j} [k] \]

Again, the expression given in Eq. (2.97) is substituted for \( \Upsilon [n] \). In controlling the cost, \( T C_{reg} \), the regulator has the control variables of the investment in generation and in transmission as well as the rate at which the electric services are provided to the loads at its disposal. This is due to the particular structure of the cost-of-service regulation scheme considered here where the regulator has the sole authority to approve the proposed investment and the rate.

The optimization problem associated with the regulator is then formulated by combining Eqs. (2.88), (2.101), (2.103), and (2.111) given as the following:

\[
\begin{align*}
\left[ \bar{I}_G, \bar{I}_T, \bar{e}_{tech}, \bar{e}_m, \bar{\rho} \right]^T &= \arg \max \sum_{l} \sum_{\bar{e}_{tech}[l], \bar{e}_m[l]} \sum_{\bar{\rho}[l]}^{5} \mathcal{E} \left\{ (1 - \xi)^nT_T \left[ r_{cos} \left( \sum_{g_i} C_{g_i}^G (K_{g_i}^G [n], I_{g_i}^G [n], n) \right. \right. \\
& \quad + \left. \sum_{i} C_i^T (K_i^T [n], I_i^T [n], n) \right) - v_{tech}(e_{tech}[k]) - v_m(e_m[k]) \right] - (1 + \lambda_f) \times \\
& \quad \left[ (1 - \xi)^nT_T (1 + r_{cos}) \left( \sum_{g_i} C_{g_i}^G (K_{g_i}^G [n], I_{g_i}^G [n], n) + \sum_{i} C_i^T (K_i^T [n], I_i^T [n], n) \right) \right. \\
& \quad \left. - \sum_{k=(n-1)T_T+1}^{nT_T} (1 - \xi)^k \sum_{d_j} \rho [n] \cdot Q_{d_j} [k] \right) \right\} \\
\end{align*}
\]

subject to the following constraints:

\[ I_{g_i} [k] \geq 0 \]  
(2.113)

\[ I_{i} [k] \geq 0 \]  
(2.114)

\[ e_{tech} [k] \geq 0 \]  
(2.115)

\[ e_m [k] \geq 0 \]  
(2.116)

\[ \sum_{g_i} K_{g_i}^G [n] \geq \hat{K}_{G} [n]: \text{ according to block (b) in Figure 2-4} \]  
(2.117)

\[ K_i^T [n] \geq \hat{K}_i^T [n]: \text{ according to block (d) in Figure 2-4} \]  
(2.118)

Plus, an accurate assessment of \( Q_{d_j} [k] \) in Eq. (2.112) requires the regulator to solve an optimization problem of the following form:

\[
\begin{align*}
\left[ u_G, \bar{Q}_G, \bar{Q}_D \right]^T &= \arg \max u_G[k], \bar{Q}_G[k], \bar{Q}_D[k] \mathcal{E} \left\{ (n-1)T_T+\bar{\eta} \left( \sum_{k=(n-1)T_T+\bar{\eta}}^{nT_T} (1 - \xi)^k \left( \sum_{d_j} U_{d_j} (Q_{d_j} [k], k) - \rho [n] \cdot Q_{d_j} [k] \right) \right) \right\} \\
\end{align*}
\]

(2.119)
\[- \sum_{g_i} c_{g_i}(x_{g_i}[k], u_{g_i}[k], Q_{g_i}[k], k) \}

where \([q, \bar{q}]\) is the interval over which the unit commitment decision by the vertically integrated utility is made. The optimization problem in Eq. (2.119) is subject to the following constraints, as is the case with the utility:

\[ u_{g_i}[k + 1] = u_{g_i}[k] \text{ if } t_{g_i, dn} < x[k] < t_{g_i, up} \]  \hspace{1cm} (2.120)

\[ \sum_{g_i} Q_{g_i}[k] = \sum_{d_j} Q_{d_j}[k] : \lambda[k] \]  \hspace{1cm} (2.121)

\[ 0 \leq Q_{g_i}[k] = (n - 1)T + mT_G + q \leq K_{g_i}^G[n - 1] + \varepsilon_m I_{g_i}^G T_G - 1 \]
for \(m = 1, 2, \cdots, T_T - 1\), and for \(q = 1, 2, \cdots, T_G - 1\) \hspace{1cm} (2.122)

where \(\varepsilon_m\) is a vector of all zeros except for the \(i\)th element which is equal to 1

\[ F_i(Q_G[k], Q_D[k]) \leq F_i^{\text{max}}(F[k], K_i^T[n], e_{\text{tech}}[n], e_m[n]) : \mu_i[k] \]  \hspace{1cm} (2.123)

Compared to Eq. (2.7) we have defined the systemwide social welfare associated with the slow dynamics, \(SW_{\text{utility},1}[n]\), to be the following:

\[ SW_{\text{utility},1}[n] = r_{\cos} \left( \sum_{g_i} C_{g_i}^{G}(K_{g_i}^G[n], I_{g_i}^T[n], n) + \sum_l C_l^T(K_I^T[n], I_I^T[n], n) \right) - v_{\text{tech}}(e_{\text{tech}}[k]) - v_m(e_m[k]) \]

\[-(1 + \lambda_T) \left( (1 + r_{\cos}) \left( \sum_{g_i} C_{g_i}^{G}(K_{g_i}^G[n], I_{g_i}^T[n], n) + \sum_l C_l^T(K_I^T[n], I_I^T[n], n) \right) \right) \]

\[-\sum_{k=1}^{T_T}(1 - \xi)^k \sum_{d_j} \rho[n] \cdot Q_{d_j}[k] \]

and the systemwide social welfare associated with the fast dynamics, \(SW_{\text{utility},2}[k]\), as:

\[ SW_{\text{utility},2}[k] = \sum_{d_j} (U_{d_j}(Q_{d_j}[k], k) - \rho[n] \cdot Q_{d_j}[k]) - \sum_{g_i} c_{g_i}(x_{g_i}[k], u_{g_i}[k], Q_{g_i}[k], k) \]  \hspace{1cm} (2.125)

under the rate-of-return regulation imposed on the vertically integrated utility.

The optimization problem as expressed in Eqs. (2.112) and (2.119) becomes a tremendous burden for the regulator especially because various functions necessary for solving the problem are proprietary to the vertically integrated utility by nature. For example, the various options associated with network expansion and generation addition are not readily available to the regulator, so that estimating the cost functions of investment in generation and in transmission is not a trivial task. The cost functions associated with the control effort and the maintenance effort are even more difficult to estimate than those of investment. Plus, the utility functions of the individual loads may be deduced much more easily by the vertically integrated
utility than by the regulator.\(^9\) Admittedly the vertically integrated utility is in a much better position to approach the optimization problem in Eqs. (2.112) and (2.119). However, as described earlier, the incentive structure of the rate-of-return regulation is such that the optimization problem associated with the vertically integrated utility is the one given in Eqs. (2.101) and (2.103).

It can be seen from comparing the optimization problem of the vertically integrated utility in Eqs. (2.101) and (2.103) and the one of the regulator in Eqs. (2.112) and (2.119) that a significant difference can arise in the decisions about investment in generation and in transmission. There is a good chance that the much debated incident of regulation, often referred to as Averch-Johnson effect, will take place [5]. The Averch-Johnson effect observes that the regulated firm is more likely to opt for the capital investment when possible. This is stated more formally in the following proposition.

**Proposition 2:** First, we make a conventional assumption that \(F_i^{\max}(\mathbf{F}[k], K_i^T[n], e_{tech}[n], e_m[n])\) in Ineq. (2.123) is an increasing function with respect to \(K_i^T[n]\).\(^{10}\)

Suppose there exist two alternatives for satisfying the constraints listed in Ineqs. (2.113) through (2.118), namely \(u_{utility}^{\star} = [\mathbf{I}_G^+, \mathbf{I}_T^+, e_{tech}^+, e_m^+]\) and \(u_{utility}^{\dagger} = [\mathbf{I}_G^+, \mathbf{I}_T^+, e_{tech}^+, e_m^+]\). Further suppose that \(u_{utility}^{\star}\) is a part of the solution to the optimization problem in Eqs. (2.112) and (2.119) and \(u_{utility}^{\dagger}\) is the solution to the optimization problem in Eqs. (2.101) and (2.103). Then,

\[
u_{utility}^{\star} = u_{utility}^{\dagger}
\]

only if

\[
\begin{align*}
\nu_{tech}(e_{tech}^+) &= \nu_{tech}^{\min}(e_{tech}^{\min}) \\
\nu_m(e_m^+) &= \nu_m^{\min}(e_m^{\min})
\end{align*}
\]

where \(\nu_{tech}^{\min}(e_{tech}^{\min})\) and \(\nu_m^{\min}(e_m^{\min})\) are the smallest costs possible associated with the control effort and with the maintenance effort, respectively, while not violating the constraints listed in Ineqs. (2.113) through (2.118). This is because if we assume that \(u_{utility}^{\star} = u_{utility}^{\dagger}\) holds true but

\[
\begin{align*}
\nu_{tech}(e_{tech}^+) &\neq \nu_{tech}^{\min}(e_{tech}^{\min}) \\
\nu_m(e_m^+) &\neq \nu_m^{\min}(e_m^{\min})
\end{align*}
\]

\(^9\)To be more accurate, the utility function of each load is revealed to the vertically integrated utility in the form of the demand function for electricity, \(D_d(Q_d[k], k)\), i.e.,

\[
D_d(Q_d[k], k) = \frac{dU_d}{Q_d[k]}(Q_d[k], k)
\]

\(^{10}\)This is a typical assumption that the flow limit on transmission line \(l\) can be increased with more investment in the same line.
Then, by direct substitution and (2.129), we have

$$\Pi_{utility}(\overline{I}_G^*, \overline{I}_T^*, e_{tech}^{min}, e_m^{min}) < \Pi_{utility}(\overline{I}_G^*, \overline{I}_T^*, e_{tech}^{min}, e_m^{min})$$  \hspace{1cm} (2.130)

Now, $u^{\dagger}[\overline{I}_G^*, \overline{I}_T^*, e_{tech}^{min}, e_m^{min}]$ may not satisfy all the constraints. Among the constraints listed in Ineqs. (2.113) through (2.118), the only constraint that can be violated is when $u^{\dagger}$ is the constraint in Ineq. (2.118). However, this constraint can be satisfied by increasing $R_T^T$ by increasing the investment in transmission, $\overline{I}_T^*$, as suggested at the beginning of this proposition. Plus, when the constraint is satisfied, the resulting profit, $\Pi_{utility}$, is at least as great as $\Pi_{utility}(\overline{I}_G^*, \overline{I}_T^*, e_{tech}^{min}, e_m^{min})$. Thus, $u^{\dagger}_{utility}$ is not the solution to the optimization problem in Eqs. (2.101) and (2.103) if $v_{tech}(e_{tech}^{*}) \neq v_{tech}^{min}(e_{tech}^{min})$ or $v_m(e_m^{*}) \neq v_m^{min}(e_m^{min})$.

Given the difficulties above, a market mechanism is introduced through the restructuring process in order to improve (specifically long-term) efficiency. In the subsequent section we describe so-called functional unbundling fundamental to the restructuring process of the electric power industry, and defer to Chapter 3 for a detailed discussion of improved efficiency.

### 2.3 Restructuring process and functional unbundling

With the introduction of market mechanisms to the electric power industry, the generation sector of a vertically integrated utility is changed from a regulated industry to a competitive industry. This means that instead of having the vertically integrated utility as the only supplier, there are many new companies that produce and market wholesale electric power. These new companies are in direct competition with one another in supplying electricity to the loads. In addition to wholesale competition, retail competition has been initiated in some regions of the US, and for the first time in the industry, the consumers in those regions have a choice of electricity suppliers. The introduction of wholesale and retail competition to the electric power industry has produced and will continue to produce significant changes in the industry. These changes are collectively referred to as the restructuring process [78].

Historically, independent power suppliers are denied from using transmission systems by the vertically integrated utility who owns and operates the network. In some instances, the utility indirectly refuses to make the network accessible to the independent power suppliers by withholding certain types of important transmission services from them. In other instances, when the suppliers are allowed to use the network, the services are provided to the utility owned generators first before serving independent power suppliers when a scarcity in transmission services is created by transmission congestion. The growth of a competitive power generation market is significantly impeded due to these practices by the utility [79].

To address such hindrances to flourishing competition in the energy market, the vertically integrated utilities are required by the regulator - in the case of the US, the Federal Energy Regulatory Commission
(FERC) - to undergo a functional unbundling of the transmission sector from the rest. The functional unbundling of transmission induces the implementation of the same tariff for using the network to all users regardless of their affiliations. The detached entity which offers only transmission services is often called a transmission provider (TP). As a result, separate rates are established for wholesale generation and for transmission. Based on the transmission rates the network users are impartially provided with various network services including network capacity, scheduling, dispatch and system control. In addition, a public (electronic) information system is created so that any network users can obtain data about prices and the available capacity of the transmission systems in the same way at the same time. The key notion of functional unbundling is to avoid discriminatory practices within a vertically integrated utility by completely separating the functions related to network services from the rest of the utility.

As a consequence of this functional unbundling, a new set of objectives are created for all individual entities, including suppliers, consumers and the TP. In the following section we briefly describe these objectives.

### 2.3.1 Defining the objectives of the transmission provider (TP)

A TP operates and plans for the regional electric power network so that the system integrity is preserved while a fair and equitable access to the network is provided to all users. In the short term, a TP is responsible for managing congestion and ensuring systemwide reliability through generation schedule, dispatch and system control. In the long term, a TP is responsible for maintaining an adequate level of network capacity by investment in transmission expansion, technology and maintenance. A TP is considered to be a natural monopoly due to its high sunk cost, economies of scale and economies of scope [59] [63]. Given its monopoly status, the TP assumes the authority, under strict oversight by the regulator, to administer transmission tariffs and congestion charges instead of the users determining the value of the services provided. This is an important function not only from the perspective of meeting traditional revenue requirements (the transmission comprises about 2% of the major utilities’ operating expenses) but also from the perspective of providing proper price signals. It is important to realize that, with functional unbundling, ensuring systemwide reliability and maintaining an adequate level of network capacity become increasingly difficult since explicit coordination is lost for utilizing generation and transmission resources. However, since the incentives to add new generation and transmission capacity and to locate future loads, are often created based on the congestion charges, some level of coordination may be achieved through these price signals.

To summarize, the main objectives of the TP are two-fold; (1) to invest so that its allowed profit is maximized under the regulations and (2) to operate so that the systemwide reliability is maintained at an acceptable level. To achieve these objectives, providing the proper price signals, through transmission tariffs and congestion charges, becomes critical.
2.3.2 Defining the objectives of network users

Network users can be categorized as either suppliers or consumers.

Objectives of suppliers

There are many different sizes and types of suppliers, ranging from small photovoltaic ones to large nuclear plants. Each supplier is unique in that, depending on its type, size and location, it provides various services. For example, the units equipped with automatic generation control devices can provide a real-time balancing of generation and fluctuating demand, often referred to as regulation services. Other units with fast-start, fast-ramping capability can afford additional capacity that can be switched on in response to a contingency, such as the loss of on-line generation, within a short period. This service is typically called operating reserves.

The profit of a supplier is determined by the revenue from offering generated output and the corresponding generation cost from the wholesale market. Given that the price on the wholesale electricity market changes widely from hour to hour, the supplier is responsible, in the short term, for determining the best time to turn on and off its generating unit and the desirable production level and, in the intermediate term, for deciding upon a suitable maintenance schedule for the unit, such as determining when and how long to take the unit out of the network for services. In addition, the price varies from one region to another within the same network due to the effect of transmission congestion and, in the long term, the supplier is responsible for the most expedient future expansion, considering both the intertemporal and locational variation of market conditions. Given that there are also longer term financial tradings through which energy can be sold, the main objectives of suppliers are optimal production and marketing decisions that maximize their respective profits.

Objectives of the consumers

In the wholesale market, most of the consumers are distribution companies connecting retail customers of various classes to the transmission network. Generally, 4 classes of customers are defined - residential, commercial, industrial, and "other" - primarily based on the power consumption (demand patterns and load usage) level and distribution voltage level. These classifications are currently used to determine the appropriate fees to be charged for delivering, from the wholesale market, bundled electric services: the cost of the energy, transmission, and distribution plus some other charges such as billing. As the retail competition gains greater measure, part of or all of the fee is expected to go through the unbundling process. The main objective of the consumers continues to be the maximization of their respective consumer surplus.
Chapter 3

Measuring the performance of the transmission provider (TP) using a systemwide social welfare function under uncertainties

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

First, the benchmark performance measure is defined while accounting for the subtlety of functional unbundling described in Chapter 2. This benchmark performance measure may be compared to the systemwide social welfare function for the benign social planner given in Eq. (2.14). The maximization of benchmark performance yields an optimal level of investment, control effort and maintenance effort for transmission [9]. It is shown that under certain conditions, optimizing the benchmark performance leads to solving the optimization problem defined in Eq. (2.13).

Following the formulation of the benchmark performance measure we describe two possible regulation schemes to be imposed on the TP, namely the cost-of-service regulation and the price-cap-regulation (PCR). The TP remains a monopoly through the restructuring process due to the assumption that there exist high sunk cost, economies of scale and economies of scope for the network similar to the argument presented in Chapter 2. The main function of the TP is to provide adequate transmission capacity for participants to trade electricity in the electric energy market.
Then, a systemwide social welfare function is developed under the cost-of-service regulation, this time imposed on the TP. The restructuring of the electric power industry is still a relatively recent process at the time of this writing, and there is yet to be a consensus on an actual implementation scheme for regulating the TP based on guaranteed cost-of-service. In this chapter, four of the more common implementation schemes are described and examined using the corresponding systemwide social welfare functions. The formulated systemwide social welfare functions can be compared to what is given in Eqs. (2.124) and (2.125).

It is shown that even though each scheme has a few distinct peculiarities that separate it from the others, they all suffer from a shortcoming similar to that suffered by the cost-of-service regulation imposed on the vertically integrated utility, most notably the burden put on the regulator in eliciting from the regulated firm, in this chapter the TP, the social welfare optimizing behavior given in Eqs. (2.112) and (2.119).

The PCR is proposed as a possible alternative regulation scheme to be imposed on the TP [15]. Starting from one of the possible regulation schemes described under cost-of-service regulation, we describe the systemwide social welfare function associated with the newly developed PCR in the context of the industry, and show that the main difference between these two regulation schemes is not in the functional form of the systemwide social welfare but is in the party responsible for solving the optimization problem [18].

3.1 Benchmark performance measures for the transmission provider (TP)

Following the restructuring process the electricity is provided to the load by the generators through market mechanisms, and the vertically integrated utility is divided into generation, transmission and distribution/load sectors as described in Chapter 2.

Under each functional unbundling we assume that the actual utility functions of the loads and the actual cost functions (related to both operations and investment) of the generators within a given regional electric power network become highly guarded private information and are only revealed in the form of demand functions and supply functions respectively through market participants’ overall market activities regardless of the actual market implementation of a particular region for energy.\(^1\) We denote the demand and supply functions as \(D_d(Q_d[k], k)\) and \(S_{g_i}(Q_{g_i}[k], k)\). On the other hand, the cost functions associated with a TP - the investment cost into transmission, \(C^T(K^T_k[k], I^T_k[k], k)\), the control cost, \(v_{tech}(e_{tech}[k], k)\), and the maintenance cost, \(v_m(e_m[k], k)\) - are assumed to be available so that the only uncertainties related to these functions are the stochastic nature of their future values.

By combining the demand and supply functions and the cost of operating, maintaining and investing in

\(^1\)This is not to be confused with the demand and supply functions that are required to be bidden to so-called spot markets in some regions in the US which have gone through the restructuring process. The demand and the supply functions in this chapter refers 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.
a transmission network, the new systemwide social welfare function after the functional unbundling process is given as the following:

\[
SW_{TP}[k] = \sum_{d_j} \int_{Q_{d_j}[k]=0}^{Q_{d_j}[k]} D_{d_j}(\bar{Q}_{d_j}[k], k)d\bar{Q}_{d_j}[k] - \sum_{g_i} \int_{Q_{g_i}[k]=0}^{Q_{g_i}[k]} S_{g_i}(\bar{Q}_{g_i}[k], k)d\bar{Q}_{g_i}[k] - \sum_{l} C_{l}^T(K_{l}^T[k], I_{l}^T[k], k)
- \nu_{tech}(e_{tech}[k], k) - \nu_{m}(e_{m}[k], k)
\]

(3.1)

The subscript \(TP\) indicates that only information regarding the TP is assumed to be fully available. The benchmark performance measure is then formulated as the problem of maximizing the systemwide social welfare function given in Eq. (3.1), and is given as the following:

\[
\left[ I_{T}^*, e_{tech}^*, e_{m}^*, Q_{G}^*, Q_{D}^* \right]' = \text{arg} \max_{I_{T}^*[k], e_{tech}[k], e_{m}[k], Q_{G}[k], Q_{D}[k]} \sum_{k=1}^{T} (1 - \xi^k) E \{ SW_{TP}[k] \}
\]

(3.2)

subject to

\[
I_{l}[k] \geq 0
\]

(3.3)

\[
e_{tech}[k] \geq 0
\]

(3.4)

\[
e_{m}[k] \geq 0
\]

(3.5)

\[
\sum_{g_i} Q_{g_i}[k] = \sum_{d_j} Q_{d_j}[k] : \lambda[k]
\]

(3.6)

\[
Q_{g_i}^{\text{min}}[k] \leq Q_{g_i}[k] \leq Q_{g_i}^{\text{max}}[k] : \eta_{g_i}[k]
\]

(3.7)

\[
F_{l}(Q_{G}[k], Q_{D}[k]) \leq F_{l}^{\text{max}}(F[k], K_{l}[k], e_{tech}[k], e_{m}[k]) : \mu_{l}[k]
\]

(3.8)

where

\[
K_{l}^T[k+1] = (1 - \rho_{T}) K_{l}^T[k] + I_{l}[k]
\]

(3.9)

and \(Q_{g_i}^{\text{min}}[k]\) and \(Q_{g_i}^{\text{max}}[k]\) are the minimum and the maximum output of generator \(g_i\) and are also revealed through their overall market activities. The maximization is solved for some very long time scale because of the continuous intertemporal interplay.

It is worthwhile comparing the benchmark performance measure associated with the stand-alone TP after the functional unbundling process with the optimization problem of a benign social planner for the purposes of relating the optimality of the solutions to the underlying market structure.

We see that 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. (2.13), the utility functions of the loads and the cost functions of the generators are assumed to be available to a benign social planner. The assumption of knowing these functions in detail allows for an explicit determination of the optimal control
variables related to the loads and the generators, including the investment in generation, the generation scheduling and the unit commitment decisions. The result is the true optimality of the systemwide social welfare function.

On the other hand, as apparent from Eq. (3.2), the utility functions of the loads and the cost functions of the generators, in the stand-alone TP environment, are assumed to be unavailable but can only be inferred from their market activities. The result is a loss of optimality in its solution but, at the same time, a gain from the simplification in solving the optimization problem. The control variables derived from solving Eq. (3.2) guarantee the optimality of the systemwide social welfare function up to the optimality of the market activities of the loads and the generators. In order to ensure that the unique solution\(^2\) to Eq. (3.2) matches the unique solution to Eq. (2.13), three conditions need to be met, namely (1) the demand and supply functions are equal to the marginal utility and the marginal (operating) costs of the corresponding loads and generators, respectively; (2) the minimum and maximum generation limits under the stand-alone TP scheme are the same as those under the benign social planner scheme when the generator is on, or are zero if the generator is off; and (3) the initial capacities of each transmission line within the network are the same under the stand-alone TP scheme and under the benign social planner scheme. This is stated more formally in the following lemma.

**Lemma 2:** Suppose \(\overline{u}_{\text{system}}\) and \(\overline{u}_{\text{TP}}\) are the unique solutions to the optimization problems stated in Eqs. (2.13) and (3.2) and the systemwide social welfare functions, \(SW^*_{\text{system}}[k]\) and \(SW^*_{\text{TP}}[k]\), are the corresponding results to the solutions, \(\overline{u}_{\text{system}}\) and \(\overline{u}_{\text{TP}}\), respectively.

Let \(\overline{u}_{\text{system}} = [I_G^*, I_T^*, e_{\text{tech}}^*, e_m^*, \overline{u}_G^*, Q_G^*, Q_D^*]'\) and \(\overline{u}_{\text{TP}} = [I_T^*, e_{\text{tech}}^*, e_m^*, Q_G^*, Q_D^*]'\). Then,

\[
\overline{I}_T^* = I_T^* \tag{3.10}
\]

\[
e_{\text{tech}}^* = e_{\text{tech}}^* \tag{3.11}
\]

\[
 e_m^* = e_m^* \tag{3.12}
\]

\[
 Q_G^* = Q_G^* \tag{3.13}
\]

\[
 Q_D^* = Q_D^* \tag{3.14}
\]

if and only if

\[
D_{d_j}(Q_{d_j}[k], k) = \frac{dU_{d_j}}{dQ_{d_j}[k]}(Q_{d_j}[k], k) \tag{3.15}
\]

\[
S_{g_i}(Q_{g_i}[k], k) = \frac{\partial C_{g_i}}{\partial Q_{g_i}[k]}(x_{g_i}[k], u_{g_i}[k], Q_{g_i}[k], k) \tag{3.16}
\]

\(^2\)The convexity of functions within the optimization problem assumed in Chapter 2 ensures the uniqueness condition of the solution and is extended in this chapter.
\[ Q_{gi}^{\min}[k] \text{ (in Eq. (3.7))} = \begin{cases} Q_{gi}^{\min}[k] \text{ (in Eq. (2.25) if } u_{gi}^*[k] = 1 \text{ from } \overline{u}_{\text{system}}^*),} \\
0 \text{ otherwise} \end{cases} \quad (3.17) \]

\[ Q_{gi}^{\max}[k] \text{ (in Eq. (3.7))} = \begin{cases} K_i^G[k] \text{ (in Eq. (2.25) if } u_{gi}^*[k] = 1 \text{ from } \overline{u}_{\text{system}}^*),} \\
0 \text{ otherwise} \end{cases} \quad (3.18) \]

\[ K_i^T[0] \text{ (in Eq. (3.9))} = K_i^T[0] \text{ (in Eq. (2.6))} \quad (3.19) \]

**Proof:** Suppose the conditions specified in Eqs. (3.15) through (3.19) hold true. Then, the control variables of the optimization problems in Eqs. (2.13) and (3.2) are subject to the same set of constraints, since the constraints on \( Q_{gi}^*[k] \) imposed by Eqs. (2.5) and (2.23) and Ineq. (2.25) put the same constraints on \( Q_{gi}^{\dagger}[k] \) by Eqs. (3.17) and (3.18). Plus,

\[ SW_{TP}^{\dagger}[k] = SW_{\text{system}}^*[k] + \text{Const.} \quad (3.20) \]

due to Eqs. (3.15) and (3.16), where Const. is the constant cost related to the systemwide operating cost of the generators. For example,

\[ \text{Const.} = \sum_{g_i, f} c_{g_i,f} + \sum_{g_i, k} T_{g_i} + \sum_{g_i} S_{g_i} \quad (3.21) \]

where \( c_{g_i,f}, T_{g_i}, \text{ and } S_{g_i} \) are as defined in Eq. (2.2). Thus, the maximizing \( SW_{TP}^{\dagger}[k] \) and \( SW_{\text{system}}^*[k] \) lead to the same set of solutions given that the initial conditions match. This restriction on the initial condition is enforced due to Eq. (3.19).

Suppose, on the other hand, the conditions specified in Eqs. (3.10) through (3.14) hold true for any convex functions of \( U_{d_i}(Q_{d_i}[k], k) \) 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. (2.13) and (3.2) 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 since the constraints on the other control variables are already the same due to Ineqs. (3.3) through (3.6) and (3.8) through (3.9), which are identical to Ineq. (2.17), (2.20) through (2.22), (2.24), and (2.26). If \( u_{g_i}[k]^* = 0 \), then \( Q_{g_i}^*[k] = 0 \) while if \( u_{g_i}[k]^* = 1 \), then

\[ Q_{g_i}^{\min} \leq Q_{g_i}^*[k] \leq K_i^G[k] \quad (3.22) \]

This implies that the constraints for \( Q_{g_i}^{\dagger}[k] \) must be defined as the following:

\[ Q_{gi}^{\min}[k] \text{ (in Eq. (3.7))} = \begin{cases} Q_{gi}^{\min}[k] \text{ (in Eq. (2.25) if } u_{gi}^*[k] = 1 \text{ from } \overline{u}_{\text{system}}^*),} \\
0 \text{ otherwise} \end{cases} \quad (3.23) \]
and
\[ Q_{g_i}^{\text{max}}[k] \text{ (in Eq. (3.7))} = \begin{cases} K_{g_i}[k] \text{ (in Eq. (2.25) if } u_{g_i}[k] = 1 \text{ from } \bar{u}_{\text{system}}^*) \\ 0 \text{ otherwise} \end{cases} \] (3.24)

In order to match the objective functions within the optimization problems in Eqs. (2.13) and (3.2),
\[ U_{d_j}(Q_{d_j}[k], k) = \int D_{d_j}(Q_{d_j}[k], k)dQ_{d_j}[k] + \text{Const1.} \] (3.25)
\[ c_{g_i}(x_{g_i}[k], u_{g_i}[k], Q_{g_i}[k], k) = \int S_{g_i}(Q_{g_i}[k], k)dQ_{g_i}[k] + \text{Const2.} \] (3.26)
where Const1 and Const2 are some constants. Finally, by imposing that the initial conditions are the same:
\[ K_i^T[0] \text{ (in Eq. (3.9))} = K_i^T[0] \text{ (in Eq. (2.6))} \] (3.27)

the optimization problems in Eqs. (2.13) and (3.2) are equivalent, as specified by the conditions in Eqs. (3.10) through (3.14). Thus, the constraints defined in Eqs. (3.23) through (3.27) must be satisfied, which are identical to the constraints in Eqs. (3.15) through (3.19).

Since the energy portion of the electricity is provided through market mechanisms, the decisions for unit commitment and investment into generation are determined through decentralized optimization by individual generators. Suppose the unit commitment and investment decisions determined by each supplier are the same as the ones obtained by solving the centralized optimization problem in Eq. (2.13) [1]. Then, the conditions given in Eqs. (3.17) and (3.18) are satisfied. Plus, the conditions defined in Eqs. (3.15) and (3.16) 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 perfect market assumptions. Invoking Lemmas 1 and 2, we conclude that the market activities modeled by the optimization problem in Eq. (3.2) lead to the same optimal solution, with respect to systemwide social welfare, as the one obtained by solving the centralized optimization problem in Eq. (2.13).

Some modifications are necessary to the optimization problem in Eq. (3.2) in order to apply it to an actual system because of the high sunk cost and economies of scale assumed for the investment in transmission described in Proposition 1. Similar to the formulation developed in Chapter 2, we introduce the regulation imposed on the TP as a mechanism for inducing the operation and planning of an electric power network close to the systemwide optimal social welfare function. This implies that the energy portion of the electricity is being provided through market mechanisms while the transmission portion of it is being provided through regulation. In the following sections various regulation schemes that may be imposed on the TP are discussed.
3.2 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 3-1. The TP 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 the various electric services (generation and transmission) while the generators are the suppliers of the energy portion of the electric services. The marketers often participate in the trading of electric services on behalf of the loads or generators, and typically do not own or operate generation, transmission or distribution systems. The function of the marketers is largely ignored in this chapter, but is re-visited in Chapter 6.

The regulator is typically a government agency whose responsibility is to oversee, directly and/or indirectly, the operations and the planning of the network by the TP. This regulation is necessary even after the restructuring process since the TP provider remains a monopoly largely due to the high sunk cost and economies of scale. As a monopoly the TP charges for the transmission portion of electric services above the marginal cost of the network capacity so that the TP may continue to support the network as a viable business while ensuring a reasonable return on its investment. The regulation determines what the degree

Figure 3-1: Composition of the electric power network economics after the restructuring process
of reasonable return is and limits the TP from charging more than is reasonable. The cost-of-service regulation guarantees a return on all investments that made with an approval, up to the amount allowed by the regulator. We examine the effect of this type of regulation imposed on the vertically integrated utility in Chapter 2.

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 electric service is now provided through market mechanisms. Only the transmission portion of the electric service is under the direct control of the regulator, through rate approval. However, there is a significant expansion of the regulator’s function in terms of its indirect control over the electric power network economics. This is due to the fact that the particular form of market mechanism governing the energy market needs to be approved of by the regulator before implementation. The role of the regulator is two-fold: (1) establishing (approving) roles, rights and responsibilities within a chosen market mechanism for energy market, and (2) approving the rational rates for transmission capacity, so that the overall operation and planning of the electric power network approaches the systemwide social welfare optimization described in Eq. (3.2).

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

3.2.1 Effect of the spot market and the cost-of-service regulation imposed on a 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 value 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 the consumer, each load \( d_j \) chooses the optimal level of its consumption, \( Q_{d_j}[k] \), at each hour \( k \) in the spot market based on the maximization function, often referred to as the competitive consumer surplus function, given as the following:

\[
Q_{d_j}^*[k] = \arg \max_{Q_{d_j}[k]} \mathcal{E} \left\{ \int_{Q_{d_j}[k]}^{Q_{d_j}^*[k]} D_{d_j}(Q_{d_j}[k], k) dQ_{d_j}[k] - \rho_{e,d_j}(Q_D[k], Q_G[k], k) \cdot Q_{d_j}[k] \right\} - \hat{p}_{t,d_j}(Q_D[k], Q_G[k], k) \cdot Q_{d_j}[k] \tag{3.28}
\]

where \( \rho_{e,d_j}(Q_D[k], Q_G[k], k) \) and \( \hat{p}_{t,d_j}(Q_D[k], Q_G[k], k) \) are the prices for the energy and the expected charge.

3The restrictions of having cost-of-service regulation and of the existence of a spot market are relaxed in the later sections and the later chapters.
for the transmission (delivery) portions of electric services at load \(d_j\), respectively.

Compared to the consumer surplus function defined in Eq. (2.88), there are three noticeable differences introduced in Eq. (3.28). 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 described in Chapter 2. Since the vertically integrated utility is unbundled from a sole electric service provider into generation and transmission, the charge associated with the 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-to-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 the varying costs 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 the generators with different ownerships and, at the same time, induce economic efficiency. The third difference is related to the output dependence of each charge. Because it may cost differently to provide a different amount 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 when the electricity usage is high, a number of more expensive generators may need to be utilized in order to meet the demand, and this raises the overall energy price.

Quantity dependent pricing for transmission capacity [54] is particularly important. 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, especially at certain peak demand hours. As discussed in Chapter 2 the electric power flows on the transmission lines are limited by the transfer capacity. The dispatch in generation and load are subject to some restrictions due to this inability to conduct a transfer of electricity through a particular path in the electric power network. 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:

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

(3.29)

Thus, the prices for transmission capacities, \(\hat{\mu}_{i,d_j}(Q_D[k], Q_G[k], k)\) and \(\hat{\mu}_{i,g}(Q_D[k], Q_G[k], k)\), need to be chosen such as to give proper incentive 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, especially congestion pricing, becomes significant for achieving economic efficiency while conforming to operational limits on the power transfer through each transmission line.

Mirroring the formulation of the competitive consumer surplus function in Eq. (3.28), from the perspective of the supplier, each generator \(g_i\) chooses the optimal level of its production, \(Q_{g_i}[k]\), at each hour \(k\) in
the spot market based on the maximization function, often referred to as the competitive supplier surplus function, given as the following:\(^4\):

\[
Q^*[k] = \arg \max_{Q_g[k]} \mathcal{E} \left\{ \rho_{g_i}(Q_D[k], Q_G[k], k) \cdot Q_g[k] - \hat{\rho}_{g_i}(Q_D[k], Q_G[k], k) \cdot Q_g[k] \right\}
\]

\[
- \int_{Q_g[k]=0}^{Q_g[k]} S_{g_i}(Q_g[k], k)dQ_g[k]
\]

(3.30)

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

Since the energy portion of the electricity is provided through a market mechanism, under perfect competition with free entry assumption, the corresponding price at each bus is identical throughout the network, i.e., \(\rho_{g_i}[k] = \rho_{c,d_j}[k] = \rho_{c,g_i}[k]\). Then, the result of decentralized optimizations by all loads and generators in Eqs. (3.28) and (3.30) yields the same solution as the result of the following optimization problem:\(^5\):

\[
[Q_G^*[k], Q_D^*[k]] = \arg \max_{Q_G[k]} \mathcal{E} \left\{ \sum_{d_j} \left( \int_{Q_{d_j}[k]=0}^{Q_{d_j}[k]} D_{d_j}(Q_{d_j}[k], k)dQ_{d_j}[k] - \rho_{c,d_j}(Q_D[k], Q_G[k], k) \cdot Q_{d_j}[k] \right) \right\}
\]

\[
- \sum_{g_i} \left( \int_{Q_{g_i}[k]=0}^{Q_{g_i}[k]} S_{g_i}(Q_{g_i}[k], k)dQ_{g_i}[k] + \rho_{c,g_i}(Q_D[k], Q_G[k], k) \cdot Q_{g_i}[k] \right)
\]

(3.31)

subject to

\[
\sum_{g_i} Q_{g_i}[k] = \sum_{d_j} Q_{d_j}[k] : \lambda[k]
\]

(3.32)

\[
Q_{g_i}^{\min}[k] \leq Q_{g_i}[k] \leq Q_{g_i}^{\max}[k] : \eta_{g_i}[k]
\]

(3.33)

\[
F_l(Q_G[k], Q_D[k]) \leq F_l^{\max}[k] : \mu_l[k]
\]

(3.34)

where \(\rho_{c,d_j}(Q_D[k], Q_G[k], k)\) and \(\rho_{c,g_i}(Q_D[k], Q_G[k], k)\) replace \(\hat{\rho}_{c,d_j}(Q_D[k], Q_G[k], k)\) and \(\hat{\rho}_{c,g_i}(Q_D[k], Q_G[k], k)\) respectively so that the compensation associated with transmission congestion is expressed separately through the constraint defined in Ineq. (3.34) under centralized optimization. The equivalence of the decentralized decision making of Eqs. (3.28) and (3.30) and the centralized optimization in Eq. (3.31) is the result of Lemma 1.

The optimization problem given in Eq. (3.31) can be expressed as a linear programming (LP) problem.

---

\(^4\)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 its supply function in the market, and when the cost function of the supplier is revealed, the unit commitment decision is already internalized in its supply function.

\(^5\)Here we assume that the price for transmission charged to the market participants needs to be consistent regardless of how the contracts are structured. This assumption is very important when the market participants are involved in spot market activities as well as in various bilateral contracts, and it can be achieved by mapping the transmission usage by individual participants to each contract through which their energy needs are met, as is done in [76] [42] [43] for example.
The Lagrangian function is constructed as the following:

\[
\mathcal{L}(Q_G, Q_D, \lambda, \mu_L) = \mathcal{E} \left\{ \sum_{d_j} \left( \int_{Q_{d_j}[k]}^{Q_{d_j}[k]} D_{d_j}(Q_{d_j}[k], k) dQ_{d_j}[k] - \rho_{t,d_j}(Q_D[k], Q_G[k], k) \cdot Q_{d_j}[k] \right) \\
- \sum_{g_i} \left( \int_{Q_{g_i}[k]}^{Q_{g_i}[k]} S_{g_i}(Q_{g_i}[k], k) dQ_{g_i}[k] + \rho_{t,g_i}(Q_D[k], Q_G[k], k) \cdot Q_{g_i}[k] \right) \\
+ \lambda[k] \left( \sum_{d_j} Q_{d_j}[k] - \sum_{g_i} Q_{g_i}[k] \right) + \sum_{g_i} \eta_{g_i}[k] \left( Q_{g_i}[k] - Q_{g_i}^{\text{max}}[k] \right) + \sum_{l} \mu_l[k] \left( F_l(Q_G[k], Q_D[k]) - F_l^{\text{max}}[k] \right) \right\}
\] (3.35)

Based on the Lagrangian function defined in Eq. (3.35), it is evident that the Lagrangian multipliers \( \lambda[k] \) and \( \eta_{g_i}[k] \) are the shadow costs related to the energy portion of the electric services, and the Lagrangian multiplier \( \mu_l[k] \) is related to the transmission portion [13]. Then, the total transmission revenue collected at time \( k \) is given by:

\[
TR[k] = \sum_{d_j} \rho_{t,d_j}(Q_D[k], Q_G[k], k) \cdot Q_{d_j}[k] + \sum_{g_i} \rho_{t,g_i}(Q_D[k], Q_G[k], k) \cdot Q_{g_i}[k] + \sum_{l} \mu_l[k] \cdot F_l^{\text{max}}[k] \quad (3.36)
\]

If the price of the transmission capacity other than the shadow cost related to transmission congestion is set to be equal to zero, i.e., \( \rho_{t,d_j}(Q_D[k], Q_G[k], k) = 0 \) and \( \rho_{t,g_i}(Q_D[k], Q_G[k], k) = 0 \), then the Lagrangian multipliers, \( \lambda[k] \), \( \eta_{g_i}[k] \) and \( \mu_l[k] \) in the optimization problem in Eq. (3.35) match the shadow costs arising from the Lagrangian formulation of the optimization problem in Eq. (3.2) after time scale separation. Thus, by Proposition 1, the transmission capacity pricing of \( \rho_{t,d_j}(Q_D[k], Q_G[k], k) \), \( \rho_{t,g_i}(Q_D[k], Q_G[k], k) \) \( \neq 0 \) are the result of the high sunk cost and economies of scale assumed for the investment in transmission. It is shown in [7] that \( \rho_{t,d_j}(Q_D[k], Q_G[k], k) \), \( \rho_{t,g_i}(Q_D[k], Q_G[k], k) \) \( \neq 0 \) is equivalent to \( \rho_{t,d_j}(Q_D[k], Q_G[k], k) \) \( \neq 0 \) while \( \rho_{t,g_i}(Q_D[k], Q_G[k], k) = 0 \) by appropriately adjusting the supply function, \( S_{g_i}(Q_{g_i}[k], k) \). Therefore, for the rest of the thesis we adopt the convention of assigning the transmission charge, other than the shadow cost related to transmission congestion, only to the load. In the following section we describe four commonly proposed schemes for assigning the transmission charge.

### 3.2.2 Transmission charge under cost-of-service regulation

In order to ensure that the TP continues to support the energy market as a viable business with a reasonable expected return on its investment, the high sunk cost and economies of scale need to be addressed through the transmission charge for investment in a network.

Under the cost-of-service regulation imposed on a TP, the regulator guarantees a reasonable rate of return
on all of the approved investment in transmission made by a 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:

$$\Upsilon[n] = (1 + \rho_{\cos}) \sum_{k=(n-1)\tau_{T}}^{n\tau_{T}} \sum_{i} (1 - \xi)^k C_i^T(K_i^T[k], I_i^T[k], k)$$  \hspace{1cm} (3.37)

where $\rho_{\cos}$ is the rate of return on the investment allowed by the regulator. In Eq. (3.37) we use the fact that typically the time scale for investment in transmission is a year, i.e., $\tau_{T} = 1$ year. From the perspective of a TP, the profit is, then, determined by:

$$\Pi_{TP}[n] = \Upsilon[n] - \sum_{k=(n-1)\tau_{T}+1}^{n\tau_{T}} (1 - \xi)^k \left( \sum_{i} C_i^T(K_i^T[k], I_i^T[k], k) + \nu_{tech}(e_{tech}[k]) + \nu_{m}(e_{m}[k]) \right)$$ \hspace{1cm} (3.38)

$$= \sum_{k=(n-1)\tau_{T}+1}^{n\tau_{T}} (1 - \xi)^k \left( \rho_{\cos} \sum_{i} C_i^T(K_i^T[k], I_i^T[k], k) - \nu_{tech}(e_{tech}[k]) - \nu_{m}(e_{m}[k]) \right)$$

where the expression in Eq. (3.37) is substituted for $\Upsilon[n]$. The decision to dispensing the efforts into control and maintenance, $e_{tech}$ and $e_{m}$ respectively, are assumed for simplicity to be made only once at the beginning of each year, as was done in Chapter 2. Accordingly, the profit maximization of the TP under the rate-of-return regulation is given as the following:

$$[\bar{\Upsilon}_T^*, \bar{e}_{tech}^*, \bar{e}_{m}^*] = \arg \max_{\Upsilon_T[n], e_{tech}[n], e_{m}[n]} \sum_{n=1}^{T_I/\tau_{T}} E \{ \Pi_{TP}[n] \}$$ \hspace{1cm} (3.39)

where we make another simplifying assumption that the investment decision is made considering not an infinite time horizon but rather over the time scale of $T_I$ as was done in Chapter 2.

From the perspective of a regulator, the associated cost, $TC_{reg}[n]$, for year $n$ includes the expense arising from compensating for the difference between the revenue collected from the loads and generators and the revenue guaranteed to the TP. Again, by employing the modeling simplification in Chapter 2 and in [51] of treating the process of making up the difference in the revenue collected and allowed as an exclusive process between the regulator and the loads, the expression of this cost is given as the following:

$$TC_{reg} = (1 + \lambda_f)(\Upsilon[n] - TR[n])$$ \hspace{1cm} (3.40)

where $\lambda_f$ 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.,

$$TR[n] = \sum_{k=(n-1)\tau_{T}+1}^{n\tau_{T}} (1 - \xi)^k E \{ TR[k] \}$$ \hspace{1cm} (3.41)
derived from Eq. (3.36).

Based on the cost associated with the regulator in Eq. (3.40), it is evident that there is a significant weight placed on the transmission charge levied on the loads other than the shadow costs related to transmission congestion. As before, this is due to the high sunk cost and economies of scale assumed for the investment in transmission, which, without the additional transmission charge, leads to a considerable difference in the revenue collected and 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 optimal transmission pricing is based on 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 in transmission is made with prudence, although not necessarily as optimally as possible, and that the relative prices for the transmission portion of the electricity services are much lower than those for the energy, i.e.,

$$\sum_{k=1}^{T_T} v_\rho \cdot d_j \cdot Q_{d_j}[k] \approx \sum_{k=1}^{T_T} \sum_{d_j} \rho_i \cdot d_j \cdot Q_{d_j}[k]$$  \hspace{1cm} (3.42)

This is usually satisfied in most regions in the U.S. 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 the behavior of the parties affected by the transmission charge depends on their respective utility functions, and is thus quite system specific. A comparison is made with the solution to the optimization problem in Eq. (3.2) in which no transmission charges are imposed on the loads. At the time of this writing, no generalized result exists for quantifying the effect of the transmission charge. The third criterion is a delicate standard by which different schemes are judged since the notion of fairness tends to be highly subjective. Nevertheless, we pay particular attention to the fairness criterion when we describe the following four schemes more commonly used to assign the transmission charge, namely (1) the fully ex post allocation scheme, (2) the ex ante access fee and ex post settlement scheme, (3) the ex ante injection tax on loads and the ex post settlement scheme and (4) the ex ante flow tax on loads and the ex post settlement scheme [54] [57].

**Full ex post allocation scheme**

Under the full 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 and the collected for the TP by the regulator is computed. Part of this difference is 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. Taking into account the profit of the TP and the cost of the regulator given in Eqs. (3.38) and (3.41), the systemwide social welfare defined under the scheme may be computed by solving the optimization problem given as the following:

\[
[\mathbf{I}_T^*, \mathbf{\varepsilon}_{\text{tech}}^*, \mathbf{\varepsilon}_m^*] = \arg \max_{\mathbf{I}_T[n], \mathbf{\varepsilon}_{\text{tech}}[n], \mathbf{\varepsilon}_m[n]} \sum_{n=1}^{T_f/T_R} E \left\{ \Pi_{TP}[n] - (1 + \lambda_f)(Y[n] - TR[n]) \right\} \tag{3.43}
\]

\[
= \arg \max_{\mathbf{I}_T[n], \mathbf{\varepsilon}_{\text{tech}}[n], \mathbf{\varepsilon}_m[n]} \sum_{n=1}^{T_f/T_R} (1 - \xi)^n T_R E \left\{ r_{cos} \sum_l C_l^T(K_l^T[n], I_l^T[n], n) - v_{\text{tech}}(\mathbf{\varepsilon}_{\text{tech}}[n]) \right. \\
- v_m(\mathbf{\varepsilon}_m[n]) - (1 + \lambda_f) \left( \sum_l C_l^T(K_l^T[n], I_l^T[n], n) \right) \\
- \sum_{k=-(n-1)T_R+1}^{nT_R} (1 - \xi)^k \sum_l \mu_l[k] \cdot F_l^{\text{max}}[k] \right\}
\]

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

\[
[\mathbf{Q}_G^*[k], \mathbf{Q}_D^*[k]] = \arg \max_{\mathbf{Q}_G[k], \mathbf{Q}_D[k]} E \left\{ \sum_l \int_{Q_{d_l}[k]=0}^{Q_{d_l}[k]} D_{d_l}(\mathbf{Q}_{d_l}[k], k) dQ_{d_l}[k] \right\} \\
- \sum_{g_l} \int_{Q_{g_l}[k]=0}^{Q_{g_l}[k]} S_{g_l}(\mathbf{Q}_{g_l}[k], k) dQ_{g_l}[k] \right\}
\tag{3.44}
\]

subject to

\[
\sum_{g_l} Q_{g_l}[k] = \sum_{d_l} Q_{d_l}[k] : \lambda[k] \tag{3.45}
\]

\[
Q_{g_l}^{\text{min}}[k] \leq Q_{g_l}[k] \leq Q_{g_l}^{\text{max}}[k] : \eta_{g_l}[k] \tag{3.46}
\]

\[
F_l(\mathbf{Q}_G[k], \mathbf{Q}_D[k]) \leq F_l^{\text{max}}[k] : \mu_l[k] \tag{3.47}
\]

Computing the systemwide social welfare requires, as indicated before, solving two separate optimization problems of different time scales, one dynamics evolving at the slow rate of \( T_R \), and the other dynamics at the fast rate of an hour, as given in Eqs. (3.43) and (3.44), respectively.

At first glance, the optimization problem in Eq. (3.44) appears to be identical to the fast dynamics counterpart in Eq. (3.2) which represents the benchmark performance measure associated with the TP. 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 its consumer surplus as given in Eq. (3.28). For example, suppose it is possible that a transmission charge reaches a sustainable steady state after this particular scheme
has been in effect, a number of years, and we denote that price as $\hat{p}_t$. Then, the competitive consumer surplus function is given by:

$$
Q_{d_j}^*[n] = \arg \max_{Q_{d_j}[k]} \sum_{k=(n-1)T+1}^{nT} \mathcal{E} \left\{ \int_{Q_{d_j}[k]}^{Q_{d_j}^{\ast}[k]} D_{d_j}(Q_{d_j}^{\ast}[k], k) dQ_{d_j}[k] - \rho_{c,d_j}(Q_D[k], Q_G[k], k) \cdot Q_{d_j}[k] \right\}
$$

\begin{equation}
\hat{p}_t \cdot \max_{Q_{d_j}[k]} \left( Q_{d_j}((n-1)T+1), Q_{d_j}((n-1)T+2), \ldots, Q_{d_j}[nT]\right) \right\}
\end{equation}

thus there exists a clear distortion in the behavior of load $d_j$ compared to the benchmark performance measure given in (3.2) although the extent to which the distortion is introduced is not clear since it depends on various factors including the actual functional form of $D_{d_j}(Q_{d_j}[k], k)$ in Eq. (3.1) and the relative size of $\hat{p}_t, d_j$ to $\rho_{c,d_j}(Q_D[k], Q_G[k], k)$. Eq. (3.48) further suggests that the optimization problem in Eq. (3.44) needs to be viewed with an implied assumption that the distortion in the behavior of the load, due to the ex post transmission charge, is already reflected in the demand function observed from the perspective of the systemwide social welfare, i.e., $D_{d_j}(Q_{d_j}[k], k)$ in Eq. (3.44) is different from $D_{d_j}(Q_{d_j}[k], k)$ in Eq. (3.1).

In addition, the difference in revenue between the collected and the allowed is largest under the full ex post allocation scheme. This is due to the fact that no transmission charge is allotted in advance for offsetting the regulator’s cost, $T_{C_{ref}}[n]$, defined in Eq. (3.40). This large difference results in a considerable deviation of the solution to Eq. (3.43) from the slow dynamics counterpart in Eq. (3.1).

Similarly, some variations of the full 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 the variations are subject to, perhaps more importantly, the largest deviation permitted in the slow dynamics since no effort is made to offset the difference in revenue between the collected and the allowed. The variations of the scheme refer 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.

**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 an access fee at the beginning of the year, and then if there exists a difference in revenue between the collected through the access fee 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 the various forms as discussed in the previous section. We make one simplifying assumption that, from the sense of expected value, an adequate ex ante access fee can be determined so that no ex post charge is necessary.

Given that assumption, taking into account the profit of the TP and the cost of the regulator given in
Eqs. (3.38) and (3.41), the systemwide social welfare defined under the scheme may be computed by solving the optimization problem given as the following:

\[
\left[ \mathbf{r}_T, \mathbf{e}_{tech}^*, \mathbf{e}_m^* \right] = \arg \max_{\mathbf{r}_T[n], \mathbf{e}_{tech}[n], \mathbf{e}_m[n]} \sum_{n=1}^{T_f/T_T} \left( 1 - \xi \right)^{nT_T} \mathcal{E} \left\{ \sum_{l} C_{l}^{T} (K_l^T[n], I_{l}^T[n], n) - \nu_{tech}(e_{tech}[n]) \right. \\
\left. - \nu_{m}(e_{m}[n]) - (1 + \lambda_f) \left( \sum_{l} C_{l}^{T} (K_l^T[n], I_{l}^T[n], n) \right) \\
\sum_{k=(n-1)T_T+1}^{nT_T} \left[ \mathbf{\mu}_k \cdot F_{l}^{max}[k] \right] \right\}
\]

(3.49)

where \( \hat{\rho}_l \) 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. (3.44) subject to the constraints given in Eq. (3.45) and Ineqs. (3.46) and (3.47).

Once again, given that the optimization problem given in Eq. (3.44) needs to be solved, the solution is expected, at first glance, to be the same as the fast dynamics counterpart of the benchmark performance measure which resulted from solving the optimization problem given in Eq. (3.2). However, the access fee charged at the beginning of the year may distort the behavior of the loads; we see this by examining the surplus function given in Eq. (3.28). 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:

\[
Q_{d_j}^*[n] = \arg \max_{\mathbf{Q}_{d_j}[k]} \sum_{k=(n-1)T_T+1}^{T_T} \mathcal{E} \left\{ \int_{Q_{d_j}[k]=0}^{Q_{d_j}[k]} D_{d_j}(Q_{d_j}[k], k, Q_{d_j}[k]) dQ_{d_j}[k] - \rho_{c, d_j}(Q_D[k], Q_G[k], k) \cdot Q_{d_j}[k] - \hat{\rho}_l \right\}
\]

(3.50)

Suppose there exists a self-supplying load which needs to draw a small amount of electricity from the network only at the times when its supplier system fails to produce. As the rate of failure decreases, at some point it is possible that the following condition is met:

\[
\max_{Q_{d_j}[k]} \sum_{k=(n-1)T_T+1}^{nT_T} \mathcal{E} \left\{ \int_{Q_{d_j}[k]=0}^{Q_{d_j}[k]} D_{d_j}(Q_{d_j}[k], k, Q_{d_j}[k]) dQ_{d_j}[k] - \rho_{c, d_j}(Q_D[k], Q_G[k], k) \cdot Q_{d_j}[k] - \hat{\rho}_l \right\} < 0
\]

(3.51)

even though \( \rho_{c, d_j} \ll D_{d_j}(0, k) \) for all \( k \)'s, and the assumption in Ineq. (3.42) holds true. In this case, the load in question is better off (in the purely economic sense) because of the high transmission charge, by being disconnected from the network and by not purchasing the electricity from the network even when its supply system is not operational. Thus, there exists a clear distortion of the behavior of the load \( d_j \) compared to the benchmark performance measure given in (3.2) although, again, the extent to which the distortion is introduced is not clear since it depends on various factors including, perhaps most importantly, the actual
functional form of \(D_d(Q_d[k], k)\) in Eq. (3.1). A further inference can be made from Ineq. (3.51) that the optimization problem in Eq. (3.44) needs to be viewed with an implied assumption that the distortion of the behavior in the load is already reflected in the number of loads participating in the electricity market, i.e., the \(d_j\)'s in (3.44) are different from the \(d_j\)'s in Eq. (3.1) under the *ex ante* access fee scheme.

**Ex ante injection tax on load and the ex post settlement scheme**

Under the *ex ante* injection tax on load and the *ex post* settlement scheme, the tax rate for allowing injection\(^6\), \(\hat{\nu}_t[n]\), is determined first. Then, the transmission charge is levied on the load in the form of an 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 load. The *ex post* charge can again take on the various forms discussed earlier. We make the simplifying assumption that, from the sense of expected value, an 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, taking into account the profit of the TP and the cost of the regulator given in Eqs. (3.38) and (3.41), the systemwide social welfare defined under the scheme may be computed by solving the optimization problem given as the following:

\[
\left[ I^+_{T^+}, e_{tech^+}, e_{m^+} \right]' = \arg \max_{I^{T^+}_{T^+}, e_{tech^+}, e_{m^+}} \sum_{n=1}^{T_f/T_T} (1 - \xi)^n T_T E \left\{ r_{cos} \sum_l C_l^T (K_l^T[n], I_l^T[n], n) - \nu_{tech}(e_{tech}[n]) - \nu_m(e_m[n]) - (1 + \lambda_f) \left[ \sum_l C_l^T (K_l^T[n], I_l^T[n], n) - \sum_{k=(n-1)T_T+1}^{nT_T} (1 - \xi)^k \times \left( \sum_{d_j} \hat{\nu}_t[n] \cdot Q_d_j[k] + \sum_l \mu_l[(n-1)T_T + k] \cdot F_l^{max}[k] \right) \right] \right\} 
\]

(3.52)

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

\[
\begin{align*}
\left[ Q_{G^*}[k], Q_D^*[k] \right]' = \arg \max_{Q_{G^*}[k], Q_D^*[k]} E \left\{ D_d(Q_{d_j}[k], k) dQ_{d_j}[k] - \hat{\nu}_t[n] \cdot Q_d[k] \right\} 
\end{align*}
\]

(3.53)

subject to the constraints in Eq. (3.45) and Ineqs. (3.46) and (3.47).

A distortion in the behavior of each load is introduced to the market mechanism due to the transmission

---

\(^6\)In the case of the load, withdrawal is perhaps a 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 the more conventional usage of the term in the electric power industry at the time of this writing.
charge in the form of the injection tax. This distortion is evident from the new term, \( \hat{\rho}_t[n] \cdot Q_{d_j}[k] \), inserted in the optimization problem in Eq. (3.53), contrast to the fast dynamics counterpart in Eq. (3.2), which represents the benchmark performance measure associated with the TP. Although the extent to which this distortion affects the systemwide social welfare depends on the system, if \( \rho_{e, d_j} \ll D_{d_j}(0, k) \) for all \( k \)'s, and the assumption in Ineq. (3.42) holds true, then it may be inferred that the distortion under the \( \text{ex ante} \) injection tax scheme is smaller than under the \( \text{ex ante} \) access fee scheme since the number of loads is preserved when compared to that from Eq. (3.1).

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

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

\[
F_l[k] = F_l(Q_G[k], Q_D[k])
\]  
(3.54)

for an existing network. The vectors, \( Q_G[k] \) and \( Q_D[k] \), designate the amount of electricity injected into the network by the 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 \( \text{ex ante} \) 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:

\[
f_{l,d_j}[k] = F_l(Q_{G_{d_j}}[k], Q_{D_{d_j}}[k])
\]  
(3.55)

where \( Q_{G_{d_j}}[k] \) and \( Q_{D_{d_j}}[k] \) are given by:

\[
Q_{G_{d_j}}[k] = \left( \frac{Q_{d_j}[k]}{\sum_{d_j} Q_{d_j}[k]} \right) \cdot Q_G[k]
\]  
(3.56)

\[
Q_{D_{d_j}}[k] = [0, \ldots, Q_{d_j}[k], 0, \ldots, 0]'
\]  
(3.57)

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 \(q_{i,j}^+ [k]\) and \(q_{i,j}^- [k]\) denote the positive and the negative directional flow of \(f_{i,j} [k]\), i.e.,

\[
q_{i,j}^+ [k] = \begin{cases} f_{i,j} [k] & \text{if } f_{i,j} [k] \geq 0 \\ 0 & \text{otherwise} \end{cases}
\]

(3.58)

\[
q_{i,j}^- [k] = \begin{cases} -f_{i,j} [k] & \text{if } f_{i,j} [k] \leq 0 \\ 0 & \text{otherwise} \end{cases}
\]

(3.59)

For example, the apparent flow through transmission line \(l\), \(F_l [k]\), is the difference between the positive directional flow, \(q_{i,j}^+ [k]\), and the negative directional flow, \(q_{i,j}^- [k]\), caused by supplying the individual demand \(Q_{d,j}\), summed over all loads given by:

\[
F_l [k] = \sum_{d,j} (q_{i,j}^+ [k] - q_{i,j}^- [k])
\]

(3.60)

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 [76] [42] and [43].

Using the decomposition method in Eq. (3.55) and accounting for the profit of the TP and the cost of the regulator given in Eqs. (3.38) and (3.41), the systemwide social welfare defined under the scheme may be computed by solving the optimization problem given as the following:

\[
\begin{aligned}
\left[I_T^*, e_{tech}^*, e_m^*\right]^t &= \arg \max_{I_T[n], e_{tech}[n], e_m[n]} \sum_{n=1}^{T_T/T_S} \left(1 - \xi\right)^{nT_T} \mathcal{E} \left\{ r_{cos} \sum_l C_l^T (K_l^T [n], I_l^T [n], n) - v_{tech}(e_{tech}[n]) \right. \\
&\left. - v_m(e_m[n]) - (1 + \lambda_f) \left[ \sum_l C_l^T (K_l^T [n], I_l^T [n], n) - \sum_{k=(n-1)T_T+1}^{T_T} \int_l (1 - \xi)^{k-T_T} \times \right. \\
&\left. \left( \hat{\mu}_l[n] \sum_{d_j} (q_{i,j}^+ [k] + q_{i,j}^- [k]) + \mu_l[n] \sum_{d_j} (q_{i,j}^+ [k] - q_{i,j}^- [k]) \right) \right] \right\}
\end{aligned}
\]

(3.61)

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

\[
[Q_G^*[k], Q_D^*[k]]^t = \arg \max_{Q_G[k], Q_D[k]} \mathcal{E} \left\{ \sum_{d_j} \left( \int_{Q_{d,j}[k]=0}^{Q_{d,j}[k]} D_{d_j} (Q_{d,j}[k], k) dQ_{d,j}[k] - \sum_l \hat{\mu}_l[n] (q_{i,j}^+ [k] + q_{i,j}^- [k]) \right) \right. \\
\left. - \sum_{g_i} \int_{Q_{g_i}[k]=0}^{Q_{g_i}[k]} S_{g_i} (Q_{g_i}[k], k) dQ_{g_i}[k] \right\}
\]

(3.62)
subject to the constraints in Eq. (3.45) and Ineqs. (3.46) and (3.47).

Similar to the optimization problem in Eq. (3.53), there is a clear distortion of the behavior of each load according to the optimization problem in Eq. (3.62) under the \textit{ex ante} flow tax scheme when compared to the fast dynamics counterpart in Eq. (3.1), which represents the benchmark performance measure associated with the TP. It may be inferred that the distortion under the \textit{ex ante} flow tax scheme is smaller than under the \textit{ex ante} access fee scheme assuming \( \rho_{e,d} \ll D_{d}(0,k) \) for all \( k \)'s, and that Ineq. (3.42) holds true although it is not known the exact extent to which this distortion affects the systemwide social welfare.

\textbf{Comparison of transmission charging methods}

All of the schemes described above have advantages and disadvantages when compared to the other methods. From the expected value sense the methods described above satisfy the sufficient revenue collection criterion similarly well. As described before, each method introduces some distortions to the market activities although the degree to which these distortions are introduced by the different methods is highly system specific. Thus, we next resort to the highly subjective fairness criterion to compare the methods.

With respect to the fairness criterion the \textit{ex ante} flow tax scheme seems to be preferable to 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., \( Q_{\text{peak}} = Q_{d_i}^{\text{peak}} = Q_{d_j}^{\text{peak}} = Q_{d_k}^{\text{peak}} \). 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 at 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 \( \tau \neq t \), \( Q_{\text{peak}} = Q_{d_i}[\tau] = Q_{d_j}[\tau] \) and \( Q_{d_k}[\tau] = 0 \). For \( t \), \( Q_{\text{peak}} = Q_{d_i}[t] = Q_{d_j}[t] = Q_{d_k}[t] \). Plus, load \( d_i \) is located near cheaper generation sources while \( d_j \) and \( d_k \) are located near expensive sources. Load \( d_i \) is located quite far away from the other loads. Then, under the full \textit{ex post} scheme and under the \textit{ex ante} access fee scheme described above, each load in the network pays the same transmission charge. Under the \textit{ex ante} injection tax scheme, loads \( d_i \) and \( d_j \) pay the same transmission charge while load \( d_k \) pays much less. Under the \textit{ex ante} flow tax scheme, load \( d_j \) pays the most and load \( d_k \) pays the least transmission charge. Considering that the value of the network is different to each load based on the corresponding characteristics, the \textit{ex ante} flow tax scheme seems more equitable than the other methods although this matter is somewhat subjective and is debatable. Nevertheless, the full \textit{ex post} scheme favors the load with a consistent usage pattern, the \textit{ex ante} access fee scheme favors the native load with a large demand, and the \textit{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 cost-of-service regulation, all suffer similarly from several defects of inefficiency. The optimization problem for the regulator as expressed in Eq. (3.43), (3.49), (3.52), or (3.61) becomes a tremendous burden because the 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 of the cost associated with investment in transmission may be possible, it is unlikely that the regulator is able to evaluate accurately the costs associated with the control effort and the maintenance effort since these costs tend to be highly dependent on the constantly evolving network operating conditions [36]. Suppose the regulator makes the estimate with help from the TP. From the optimization problem for the TP given in Eq. (3.39), it is clear that the incentive structure is such that the TP favors expanding the investment in transmission over increasing either the control effort or the maintenance effort. Given this incentive structure, the cost estimates for the control effort and 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 Chapter 2 [5].

Then, there are the transmission related benefits to be evaluated. This requires forecasting the demand and supply functions of the 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 a substantial amount of knowledge about the loads and the generators is it possible to make an accurate forecast for the demand and supply functions. Unfortunately, the regulator lacks this knowledge because the regulator seldom actually participates in the market process. 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 the transmission portion of the electric services. As discussed earlier, because each scheme under cost-of-service regulation unfortunately carries a safety net of the ex post settlement so that the revenue requirement for the TP is always fulfilled even when the forecast is not made precisely, 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, the cost-of-service regulation needs to be replaced with a more appropriate regulatory structure. The PCR is one form of performance-based regulation (PBR) commonly suggested as an alternative to the cost-of-service 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. (3.43), (3.49), (3.52), or (3.43) on the regulated firm, in this case the TP, and the result is an increase in the systemwide social welfare function [67] [68]. In the following section we examine the application of a possible PBR scheme to the TP.

### 3.3 Performance-based-regulation (PBR)

Under 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 following its restructuring process, this means the price charged for providing transmission capacity by the TP is strictly based on the cost of the investment in the transmission network. As was pointed out in earlier discussions, in this environment there is little or no incentive for the TP to reduce costs through improving productivity.

The PBR is a regulatory structure where this linkage between the cost and the price of service is broken
by offering instead financial incentives to the regulated firm, the TP, to lower costs. Under the PBR, an improvement in efficiency by the TP is rewarded with higher profit while a loss in productivity is penalized with lower profit. Thus, the PBR is viewed as an attractive alternative to the traditional cost-of-service regulation to be placed on the TP; there exist many success stories related to applying the PBR in the telecommunications and railroad industries [67]. One of the main advantage of adopting the PBR is its capability to encourage the implementation of new and more advanced technology in providing services because the PBR creates incentives that are similar to those inherent in competition.

It is possible to devise different approaches to applying the 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 savings, 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 step of setting the initial price is similar to that under cost-of-service regulation. Once the initial price is set, then the regulator decides on the various indices to be used to compute the ceiling prices for the specified period into the future. These indices include changes in productivity and unanticipated changes in costs not under the control of the regulated firm. This 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, or Z, factor and include such elements as low-income program expenditures and sometimes research and development (R&D) costs [68] [7].

The firm’s incentives to reduce costs come from the higher profit expected under this approach. So long as there is a sustained demand for the service, 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 (typically 5 years) over which the price ceilings are determined is usually much longer than the price review by the regulator (1 year) under cost-of-service regulation. Such stability in regulation also induces higher efficiency since the firm is assured of keeping the additional profits realized from cost reduction without the risk of 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 to the revenue cap is subject to factors pertaining to the price as well as to the quantity. The factors pertaining to the price are the same as the factors described earlier under the price cap approach, including the X factor.

---

7There is an inherent problem with the method because the set of optimization problems defined in the previous section needs to be solved by the regulator in order to determine a meaningful initial price for enforcing the PCR scheme. Some simplification to the computation may be possible as explored in [57] after this problem is made evident but this is beyond the scope of this thesis.
and the Z factor. The factors pertaining to the quantity are mainly related to customer growth.

A noted feature of the revenue cap approach is the flexibility for the service in determining the overall output level endowed to the firm, compared to the relative inflexibility of 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 higher profit.

The sliding scale PBR may be considered a refinement of either the revenue cap or price cap approaches. 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 invokes 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 in its sharing mechanism especially if the economics related to the service being provided by the regulated firm are highly uncertain. Both the firm and the customers are protected by sharing the risks above and below certain thresholds.

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.

### 3.3.1 Price-cap regulation

Consider the ex ante flow tax scheme discussed earlier. From Eq. (3.61) the TP’s revenue for year $n$ is given by:

$$TR[n] = \mathcal{E} \left\{ \sum_{k=(n-1)T+1}^{nT} \sum_{l} (1 - \xi)^k \left( \hat{\mu}_t[n] \sum_{d_j} (q_{i,d_j}^+[k] + q_{i,d_j}^-[k]) + \mu_k[k] \sum_{d_j} (q_{i,d_j}^+[k] - q_{i,d_j}^-[k]) \right) \right\}$$  \hspace{1cm} (3.63)

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

$$[Q_G^*[k], Q_D^*[k]]' = \text{arg max}_{Q_G[k], Q_D[k]} \quad \mathcal{E} \left\{ \sum_{d_j} \left( \int_{Q_{d_j}^-[k]}^{Q_{d_j}^+[k]} D_{d_j} (Q_{d_j}^*[k], k) dQ_{d_j}^*[k] - \sum_{l} \hat{\mu}_t[n] (q_{i,d_j}^+[k] + q_{i,d_j}^-[k]) \right) \right\}$$ \hspace{1cm} (3.64)

subject to the constraints in Eq. (3.45) and Ineqs. (3.46) and (3.47). Suppose the rate of the flow tax, $\hat{\mu}_t[n]$, is allowed to vary hour-by-hour, denoted as $\hat{\mu}_t[k]$. By rearranging the expression inside (·) on the
right-hand-side (RHS) of Eq. (3.63) and substituting ⃗{\hat{\varphi}}_t[k] for ⃗{\hat{\varphi}}_t[n] we have

\[ TR[k] = E \left\{ \sum_l \sum_{d_j} \left[ (\hat{\varphi}_t[k] + \mu_t[k]) q_{t,d_j}^+ [k] + (\hat{\varphi}_t[k] - \mu_t[k]) q_{t,d_j}^- [k] \right] \right\} \]  

(3.65)

where

\[ TR[n] = \sum_{k=(n-1)T+1}^{nT+1} (1 - \zeta)^k TR[k] \]  

(3.66)

From Eq. (3.65) 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, \( q_{t,d_j}^+ [k] \) and \( q_{t,d_j}^- [k] \), and the transmission rent, \( (\hat{\varphi}_t[k] + \mu_t[k]) \) and \( (\hat{\varphi}_t[k] - \mu_t[k]) \), respectively.

Before introducing the PCR scheme to be applied to regulating the transmission business, it is noted here that the structure of the transmission rent as defined in Eq. (3.65) is not the usual multi-part tariff [4] [25] [49]. 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 to this service by charging a fixed price plus a variable price which depends on the overall quantity of the service provided. Nevertheless, it is true that the pricing appears to have two components, \( \hat{\varphi}_t[k] \) and \( \mu_t[k] \). Furthermore, it is also true that \( \hat{\varphi}_t[k] \) is placed so that the high fixed cost (and/or economies of scale) is recovered for the TP, and \( \mu_t[k] \) is dependent on the quantity of the transmission capacity being demanded. However, \( \mu_t[k] \) is zero unless the transmission line \( l \) is congested, and it 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 [8] [53] [64].

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

First, the regulator defines the initial ceiling prices, \( \hat{\varphi}_t[1] \) and \( \mu_t[1] \). 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 year 5 are determined by:

\[ \hat{\varphi}_t[n+1] = \hat{\varphi}_t[n] (1 + i_{\varphi} - X_{\varphi}) + Z_{\varphi} \]  

(3.67)

\[ \mu_t[n+1] = \mu_t[n] (1 + i_{\mu} - X_{\mu}) + Z_{\mu} \]  

(3.68)

for \( n = 1, 2, \ldots, 4 \). If there is a significant effect from exogenous factors, which requires an adjustment to the price before the end of the review period, the \( Z \) factor is defined for each price element.

76
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 to the regulation firm, in this case the TP, and completely away from the regulator so long as the following constraints are met:

\[ \hat{\mu}_t[k] \leq \hat{\mu}_t[n] \]  
\[ \mu_t[k] \leq \mu_t[n] \]  

(3.69)  
(3.70)

where \( k = (n-1)T_T + 1, (n-1)T_T + 1, \ldots, nT_T \). However, \( \mu_t[k] \) reflects the value of scarcity in transmission capacity and is determined exogenously through solving the optimization problem in Eq. (3.64). 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 [20].

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

It is recognized that each of the two price elements in Eq. (3.65) has a different impact on the operation and planning of the electric power network because the rate for flow tax, \( \hat{\mu}_t[k] \), is closely related to the recovery of investment cost in the presence of economies of scale, while the congestion price, \( \mu_t[k] \), is intimately associated with the allocation of transmission capacity at the time of scarcity. On one hand, the rate for flow tax, \( \hat{\mu}_t[k] \), is necessary for assuring the recovery of the investment cost in transmission for the optimal systemwide social welfare even when there exists a high degree of economies of scale in the network. On the other hand, the congestion price, \( \mu_t[k] \), 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 near real time operation, and the effect of the accumulated congestion price over time, \( \sum_{n=1}^{T_1/T_T} \sum_{k=(n-1)T_T}^{T_T} (1 - \xi)^k \mu_t[k] \) for each line \( l \), determines the optimal investment strategy for planning.

Suppose we make an assumption that there exists a set of prices, \( \hat{\mu}_t[l] = 0 \) and \( \mu_t[l] \neq 0 \) if and only if \( F_l([Q_G[k], Q_D[k]]) = F_l^{max}(F[k], K_l[k], e_{tech}[k], e_m[k]) \) such that there is a unique vector, \([Q_G[k], Q_D[k]]\)', which solves the optimization problem in Eq. (3.64) for given \( K_l[k] \), \( e_{tech}[k] \), and \( e_m[k] \). The required condition for this assumption may be roughly described as that there exists an adequate supply of the energy portion of the electric services [62] [13]. Fortunately this condition is always satisfied under the perfect competition with free entry assumption made earlier for the energy market. The vector, \([Q_G[l], Q_D[l]]'\)', is often referred to as the optimal power flow solution [28]. 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, \( \hat{\mu}_t[n] \) and \( \mu_t[n] \), are determined by the initial ceiling prices and the indices for price adjustment.
imposed by the regulator. Then, in near real time operation, the TP may choose a set of prices, $\hat{\rho}_l^T[k] \leq \hat{\rho}_l[n]$ and $\mu_l^T[k]$, so that there exists a solution, $[Q_{G}^T[k], Q_{D}^T[k]]'$, to the optimization problem in Eq. (3.64) 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 by setting $\hat{\rho}_l^T[k] = \hat{\rho}_l[n]$ and $\mu_l^T[k] = \mu_l[n]$.

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

$$\mu_l^T[k] \leq \mu_l[n] \tag{3.71}$$

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 to the transmission network since the TP is a monopoly. If the condition in Ineq. (3.71) is required to be satisfied absolutely, then there may be no vector, $[Q_{G}^T[k], Q_{D}^T[k]]'$, that solves the optimization problem in Eq. (3.64) while satisfying the constraints in Eq. (3.45) and Ineq. (3.46) and (3.47). Thus, some modifications to the enforcement of the PCR scheme on the TP are required.

The following modifications are proposed to the conventional PCR scheme to be imposed on the TP. On one hand, for the rate of the flow tax, $\hat{\rho}_l[k]$, a strict ceiling price, $\hat{\rho}_l[n]$, applies so that $\hat{\rho}_l[k] \leq \hat{\rho}_l[n]$ for all $k = (n - 1)T_T + 1, (n - 1)T_T + 2, \ldots, nT_T$. 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 of the energy market that there is a set of prices, $\hat{\rho}_l^T[k] \leq \hat{\rho}_l[n]$ and $\mu_l^T[k]$, such that there exists a solution, $[Q_{G}^T[k], Q_{D}^T[k]]'$, to the optimization problem in Eq. (3.64) 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. (3.65) amended as the following:

$$TR^T[k] = \begin{cases} 
\sum_i \sum_{d_j} \left[ (\hat{\rho}_l^T[k] + \mu_l^T[k]) q_{i,d_j}[k] + (\hat{\rho}_l^T[k] - \mu_l^T[k]) q_{i,d_j}^-[k] \right] & \text{if } \hat{\rho}_l^T[k] \leq \hat{\rho}_l[n] \text{ and } \mu_l^T[k] \leq \mu_l[n] \\
(1 - r_{penalty}) \sum_i \sum_{d_j} \mu_l[n] (q_{i,d_j}^-[k] - q_{i,d_j}^+[k]) & \text{otherwise, i.e. } \hat{\rho}_l^T[k] = 0 \text{ and } \\
\mu_l^T[k] > \mu_l[n] \text{ for any } l 
\end{cases} \tag{3.72}$$

where $r_{penalty}$ is the penalty rate imposed on the TP for poor performance.\(^8\) Poor performance of the TP refers to ill conceived decisions on the amount of investment, the control effort and the maintenance effort in 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 collected and the retained by the

\(^8\)The penalty for poor performance is imposed on the TP only under extreme situations because implementing this penalty scheme can be very cumbersome and is, thus, unattractive.
TP at hour \( k \) when \( \hat{\mu}_l[k] = 0 \) and \( \mu_l[k] > \mu_l[n] \) for any \( l \) is given by

\[
\sum_l \sum_{d,j} \left[ \left( \hat{\rho}_l^x[k] + \mu_l^x[k] \right) q_t^{d,j}[k] + \left( \hat{\rho}_l^i[k] - \mu_l^i[k] \right) q_t^{-d,j}[k] - \mu_l[n](1 - r_{\text{penalty}})(q_t^{d,j}[k] - q_t^{-d,j}[k]) \right]
\]  

(3.73)

and is assumed to be returned to the loads in the spot market\(^9\) indirectly through the regulator based on a modeling simplification similar to those in Chapter 2 and in [51]: treating the process of making up the difference in the revenue collected and allowed as an exclusive process between the regulator and the loads.

According to Eq. (3.72) the penalty imposed on the TP for violating the ceiling price on the 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 the flow tax is set to zero as indicated by \( \hat{\rho}_l^i[k] = 0 \). Plus, the poor performance penalty factor, \( r_{\text{penalty}} \), 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 to below the allowed level. Considering that the level of congestion is inversely correlated to the reliability of the network operation, maintaining a congestion price below the desired level is equivalent to maintaining reliability above some advisable level.

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

\[
\left[ I^{xT}, e_{\text{tech}}^{xT}, e_m^{xT} \right] = \arg \max_{I_T^{xT}, e_{\text{tech}}^{xT}, e_m^{xT}} T \left\{ \sum_{n=1}^{T_T/T_R} \left( 1 - \xi \right)^{nT_R} TS^*[n] - u_{\text{tech}}(e_{\text{tech}}[n]) - u_m(e_m[n]) \right\} 
\]  

(3.74)

where

\[
TR^*[n] = \sum_{k=(n-1)T_R+1}^{nT_R} (1 - \xi)^k TR(\hat{\rho}_l^x[k], k)
\]  

(3.75)

The complementing optimization problem for the fast dynamics is given by

\[
\hat{\rho}_l^x[k] = \arg \max_{\hat{\rho}_l^x[k]} \sum_l \sum_{d,j} \left[ \left( \hat{\rho}_l[k] + \mu_l^x[k] \right) q_t^{d,j}[k] + \left( \hat{\rho}_l[k] - \mu_l^i[k] \right) q_t^{-d,j}[k] \right]
\]  

(3.76)

if \( \hat{\rho}_l[k] \leq \hat{\rho}_l[n] \) and \( \mu_l^x[k] \leq \mu_l[n] \), or

\[
\hat{\rho}_l^x[k] = 0
\]  

(3.77)

otherwise, i.e. \( \mu_l^x[k] > \mu_l[n] \) for any \( l \). The Lagrangian multiplier, \( \mu_l^x[k] \), is the result of solving the following

\(^9\)We emphasize that the refund is made only to the spot market participants here without further explanation. Such fine peculiarity is made clear in Chapter 6.

79
optimization problem:

\[
(Q_G^*[k], Q_D^*[k]) = \arg \max_{Q_G[k], Q_D[k]} \mathcal{E} \left\{ \sum_{d_j} \left( \int_{Q_{d_j}[k]=0}^{Q_{d_j}[k]} D_{d_j}(Q_{d_j}[k], k) dQ_{d_j}[k] - \sum_l \hat{\rho}_l^*[k](q_{l,d_j}[k] + q_{l,d_j}^-[k]) \right) \right.

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

(3.78)

subject to the constraints in Eq. (3.45) and Ineqs. (3.46) and (3.47). The revenue of the TP at each hour, \( TR(\hat{\rho}_l^*[k], k) \), is given by:

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

(3.79)

where \( r_{\text{penalty}} \) is, as before, the penalty rate imposed on the TP for poor performance.

According to the optimization problem in Eqs. (3.76) and (3.76), the incentive structure of the proposed PCR mechanism is such that the TP increases transmission revenue by reducing congestion within the network to the level determined by the regulator. This incentive structure allows responsibility to be placed on the TP, and not on the regulator: the responsibility for making decisions about the amount of investment, the control effort and the maintenance effort in the transmission network as identified in the optimization problem in Eq. (3.74), 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 operations, 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 regulator’s role in defining the quality of the service being provided by the regulated firm under the conventional PCR scheme.
Chapter 4

Current development of the restructuring process in some regions in the US with an emphasis on prudent transmission provision

In this chapter we survey the development of the restructuring process in some parts of the US, namely California, Pennsylvania-New Jersey-Maryland (PJM) and the Midwest. At the time of this writing, these regions are at the forefront of deregulation due to their diligent efforts to create active energy markets.\(^1\) Then, we review the simplifying models introduced in [54] in order to represent the markets in these regions. Using these models it is asserted that much effort in the restructuring process to date has been concentrated only on establishing the short term energy markets. This is followed by some critical observations on the unique characteristics of the electric power system. These observations clearly indicate that in order to achieve an adequate level of reliability and efficiency (long-term), the presence of well functioning long term energy markets is just as important (if not more so) as having effective short term energy markets, which explains many of the problems currently being experienced in these regions. In the subsequent chapters we describe how the establishing of well functioning long term energy markets is not a trivial matter. One of the difficulties is creating a transmission provider (TP) with the ability to offer easily tradable long term transmission rights to access the system. We close the chapter with a brief preview of the subsequent chapters.

\(^{1}\)In this thesis we only consider markets conducted in time scales longer than or equal to an hour. Markets other than the energy and transmission markets, such as ancillary services markets, are also beyond the scope of this thesis.
4.1 Progress in the restructuring process in some parts of the US

The restructuring process in the California, PJM and Midwest regions of the US is supported by the creation of active energy markets and a new entity called an Independent System Operator (ISO) which performs many functions of a TP. The ISO’s in these regions are put in charge of managing the regional energy markets and operating the overall system so that an acceptable level of efficiency and reliability is obtained while the divestiture of the generation and transmission assets of the vertically integrated utility takes place, as described in Chapter 2. By design, all three ISO’s in these regions are non-profit organizations and thus have no financial interests in the energy markets. The actual functions to be carried out by these ISO’s have been defined to fit their particular regional characteristics.

4.1.1 Restructuring process in the California region

In California electricity is provided through a close interaction among the regional ISO, the Power Exchange (PX) and new economic entities called scheduling coordinators (SC’s).\(^2\)

The PX forms a public market where the network users can trade energy short term in two different time scales: daily and hourly. The participation of the PX is intended to be entirely voluntary.\(^3\) At the beginning of each day, the market participants may bid some or all of their energy requirements to the PX. There are two types of bids: energy bids and adjustment bids. The energy bids represent the market participants’ initial preference for meeting their energy requirements. The adjustment bids, on the other hand, represent the market participants’ preference for meeting the operating constraints in the network. A brief description of the operating constraints is presented in Chapter 2. Once the bids are submitted, the PX clears the market by balancing the supply and demand using the energy bids, and submits the preferred generation schedule for each hour of the given day to the ISO along with the adjustment bids. The cleared bids at the PX represent the financial obligations of the market participants.

The SC’s are the marketers holding balanced bilateral contracts between the qualified suppliers and the consumers. The actual terms of the bilateral contracts are only known to the SC, the suppliers, and the consumers involved in the respective contracts. Similar to the PX, each SC submits a preferred generation schedule per day according to the balanced transactions that it has in its possession.

Once the ISO receives a preferred generation schedule from the PX and the various SC’s, the ISO conducts system reliability studies to determine any violations of the operating constraints in the network for the given schedule. If no operating constraints are violated according to the preferred schedules submitted by the PX and the SC’s, then all the preferred schedules are accepted, and the market participants are allowed physical access to the network accordingly. On the other hand, if there exists a number of violations of operating

\(^2\) Although PX can be considered one of the SC’s, for simplicity in describing the differences between the entities, we distinguish between the PX and SC’s.

\(^3\) In order to encourage the network users to participate in the PX the former utility companies are required to purchase energy through the PX for a certain period of time.
constraints for the given schedule, then the ISO informs the SC's of this, and suggests a set of revisions to their preferred schedules in order to eliminate these violations. The SC's may or may not adjust their initial schedules according to the suggested revisions and submit their revised schedules to the ISO. If they do, the ISO then conducts system reliability studies again using the revised schedules. If no operating constraints are violated in the revised schedules from the SC's, the ISO accepts the new schedule along with the preferred schedule submitted earlier by the PX. Once accepted, the market participants are allowed physical access to the network according to their new schedule. However, if there still exists any violation of the operating constraints, then the ISO executes the schedule adjustment process using the adjustment bids submitted by the PX. In the process the ISO constrains on or off certain generation units according to the respective adjustment bids until no operating constraints are violated throughout the network. The method used in this adjustment process is often referred to as the zonal pricing method [3]. In applying the zonal pricing method, the entire California region is divided into a number of zones. As the ISO clears the adjustment bids, the ISO determines the price for each zone so that when the suppliers produce by matching the zonal price to the amount of generation in their bids, the resulting operating conditions stay within the system constraints. The details of the zonal pricing method are referred to in [71] and [72]. In extreme cases, even if after exhausting all of the generation resources available through the adjustment bids, there continues to be violations of the operating constraints, then the ISO begins curtailing transactions by the SC's until no operating constraints are violated. All transactions of the SC's approved by the ISO are subject to a congestion charge based on the zonal prices.

If the market participants' energy requirements change within the day and/or there is any residual supply and demand, these can be accommodated in the hourly spot market conducted by the PX in conjunction with the ISO, and similar to the operation of the daily spot market. The only difference is that the cleared bids in the hourly market represent physical commitment as well as financial obligation.

To a degree, due to the presence of SC's there is considerable flexibility in terms of the customer choices and the ability to arrange the generation resources over the longer term. This is because consumers may choose to go through SC's and select particular suppliers by whom their energy needs are provided. In addition, the market participants may enter into longer term energy contracts through SC's without relying on the short term spot markets managed by the PX. However, as is clear from the description given above, there is a high dependency on the short term markets in the actual operation of the network since the ISO approves all of the proposed transactions only for the short term, after the adjustment process takes place on a daily (or hourly) basis, which assures no violation of the operating constraints. The reason for this inflexibility is because the ISO, as a non-profit organization, cannot offer transmission capacity ahead of time and take on the risk of physical access on behalf of the market participants involved in longer term transactions. This may explain why over 70% of the energy trade is currently being conducted through PX's rather than through SC's [3]. Recently some financial hedging against the congestion charge based on zonal prices has been possible by purchasing so-called financial transmission rights (FTR's) from the ISO,
although physical access to the network can still not be committed in advance.

With regards to charging for transmission, the market participants are initially required to pay for the fixed cost of the transmission network, and for the associated operating and maintenance costs, through a network tariff in the form of access fees.\footnote{The description of access fees is given in Chapter 3.} This network tariff is often referred to as a license plate transmission charge and is strictly regulated by the regulator. When there is a difference in zonal prices between the zones because of binding operating constraints, an additional transmission charge is collected by the ISO. This charge is referred to as the congestion charge and is not subject to regulation. However, the regional market mechanisms are designed such that the congestion charge collected by the ISO is transferred to the transmission asset owners, who then reduce the license plate transmission charge of the consumers after some portion of the charge is used to pay the holders of the FTR’s. This rebate process allows the transmission owners to be completely indifferent with respect to the congestion charge.

4.1.2 Restructuring process in the Pennsylvania-New Jersey-Maryland (PJM) region

In PJM the electricity is provided through the so-called multi-settlement system. Under the multi-settlement system the regional ISO conducts the (short term) spot markets twice: once at the beginning of the day and the other at the beginning of each hour. At the beginning of each day, the ISO collects bids from the market participants for the next 24 hours and clears the bids by balancing supply and demand at the lowest cost possible. At the beginning of each hour during the day the ISO again collects the bids from the market participants and clears the bids by again balancing supply and demand at the lowest cost possible. The hourly spot market gives the market participants a chance to make adjustments to their earlier bids in the daily spot market. The level of efficiency is measured by the overall generation cost of meeting the systemwide demand. Once both spot markets are cleared, the supplier produces and the load consumes energy according to the cleared bids.

While managing the spot markets, the ISO is also responsible for operating the overall system. This implies that when the ISO conducts the spot markets, the operation of the system based on the bids cleared in the markets must result in an acceptable level of reliability by satisfying all operating constraints present in the network. Since the ISO does not have direct control over the generation resources, the ISO satisfies the constraints by enforcing the so-called nodal pricing method. Applying the nodal pricing method, the ISO determines the price at each bus so that when suppliers produce and the loads consume according to the respective nodal prices, the resulting operating conditions stay within the operating constraints. For example, suppose an injection at a certain bus in the network affects some operating constraints negatively. Then, the price at that bus is set lower so that the production at that bus is small based on the submitted bids. The details of the nodal pricing method are given in [71] and [72].
It is evident that the ISO plays a very important role in the PJM system since the market participants are heavily relying on the ISO for meeting their energy requirements in the presence of congestion. When there is network congestion, no exclusive bilateral agreements between the suppliers and the loads requiring direct access to the network are binding since they are subject to the nodal pricing method of the spot markets. Thus, the consumers cannot designate particular generators as their exclusive suppliers, and vice versa. Because of this reason, the market participants in the PJM system have a limited number of options for satisfying their energy needs and may be excluded, to a large extent, from one of the fundamental benefits of the restructuring process: customer choices. This lack of choices typically results in an inefficacious level of costs savings possible and/or of new technology to be introduced in the system [58].\(^5\) In addition, the suppliers in the region cannot plan their generation in advance given the strong dependence on the short term spot markets. This results in a highly inefficient operation of generation resources since the unit commitment decisions have a strong temporal effect over a time scale longer than hourly or even daily, as described in Chapter 2.

Given these limitations, the market participants have created so-called contracts for difference [29] supported by the fixed transmission rights (FTR’s) [80] [83]. The FTR’s are offered by the ISO and are auctioned off at the beginning of each month. Using the contracts for difference and the FTR’s, the suppliers and consumers can execute bilateral contracts without requiring direct access to the network as follows. First, the interested parties enter into a bilateral contract by agreeing on a strike price at which a certain amount of energy is to be traded over a certain period of time. The strike price becomes the reference price for this bilateral contract. Then, the parties purchase FTR’s between the injection bus and the withdrawal bus for an amount of energy equaling the transaction agreed upon in the bilateral contract, which completes the initial transactions necessary for establishing a relationship according to the bilateral contracts.

On each day when the spot markets are conducted, the nodal prices become the prevailing market prices. First, the ISO pays the parties holding FTR’s the difference in nodal prices between the withdrawal bus and the injection bus. For simplicity without loss of generality we assume that the suppliers of the associated bilateral transaction are the holders of the FTR’s and thus receive the difference in nodal prices accordingly. Once the payment is made from the ISO to the suppliers based on the FTR’s, the price at which each supplier generates electricity becomes exactly equal to the nodal price at which each consumer of the associated contract buys electricity from the spot markets managed by the ISO. The suppliers then pays their counterpart consumers of the bilateral contract the difference between the prevailing market price at each load bus and the reference prices. If the prevailing market price at the load bus is higher than the strike price, the load receives the difference in prices from the suppliers as agreed through the contracts for difference. If the prevailing market price at the load bus is lower than the strike price, the load pays the suppliers the difference in prices. The resulting cash flow is then precisely as described by the bilateral

\(^5\)The effect of the inefficiency due to the lack of choices is actually quite difficult to measure especially at this early stage of the restructuring process.
contracts at the strike price between the supplier and the consumer.

However, as is clear from the description above, the exchange is purely financial among the parties, including the ISO, involved in the contracts for difference and the FTR's. These financial transactions take place regardless of whether the actual physical exchanges between the supplier and the consumers occur according to the bilateral contracts or not. Indeed, if the operating conditions of the system are such that some of the suppliers involved in the bilateral contracts are restricted from generating a sufficient amount of electricity as prescribed by the contracts, then the overall cash flow including the profit from generation may be quite different from the expected because there may be a significant loss of profit from not generating the sufficient amount of energy. This risk in profit cannot be eliminated by purchasing the contracts for difference and the FTR's because the combined contracts still do not guarantee physical access to the network. Therefore, even with these contracts in addition to the short term spot markets, there exists a deficiency customer choices and longer term planning for generation.

Similarly, in California, the market participants are initially required to pay the fixed cost of the transmission network and the associated operating and maintenance costs through the access fees, as described in Chapter 3; this portion of the network tariff is strictly regulated by the regulator. However, in the case that some of the operating constraints are binding, the systemwide revenue collected by the ISO from the consumers is always greater than the cost paid to the suppliers [65]. This difference is often considered an additional transmission revenue arising from transmission congestion and becomes larger when more operating constraints are binding. At the time of this writing, it has not been worked out how this portion of the network tariff should be regulated because the (transmission) congestion revenue may vary widely from one time to another depending on the operating conditions of the overall network. To make the matters worse, some of the congestion revenue is used to pay the holders of FTR's for the nodal price differences. When physical access is denied to any holders of FTR's, it may be very possible that the congestion revenue collected is smaller than the total payment owed to the holders of the FTR. In this case, it becomes necessary for the regulator to allow the ISO to pass on to the consumers the difference between the congestion revenue and the FTR payment even through an *ad hoc* manner since the ISO is a non-profit organization. In order to avoid such instances, when the FTR's are auctioned off, the ISO is required to assure that the physical access specified by all of the FTR's are simultaneous feasible for the entire month, which requires an accurate projection of the system usage. So long as the simultaneous feasibility condition is met, the congestion revenue collected is always higher than the payment required according to the FTR's. However, given that the ISO is completely indifferent to any financial gain or loss if an inaccurate projection is made, it is somewhat questionable how well the simultaneous feasibility condition may be satisfied without the ISO being overly conservative in projection.
4.1.3 Restructuring process in the Midwest region

In the Midwest region, the electricity is provided exclusively through bilateral transactions between the suppliers and the consumers. In contrast to the California and PJM regions, the Midwest region is composed of multiple control areas. At the time of this writing the regional ISO is not yet fully operational due to difficulties in integrating the multiple control areas. There are two definitions of control areas currently being used, defined by the Federal Energy Regulatory Commission (FERC) and the North American Electric Reliability Council (NERC). A control area is, according to FERC [84],

an electric power system or combination of electric power systems to which a common automatic generation control scheme is applied in order to: 1. match, at all times, the power output of the generators within the electric power system(s) and capacity and energy purchased from entities outside the electric power system(s), with the load within the electric power system(s); 2. maintain scheduled interchange with other Control Areas, within the limits of Good Utility Practice; 3. maintain the frequency of the electric power system(s) within reasonable limits in accordance with Good Utility Practice; and 4. provide sufficient generating capacity to maintain operating reserves in accordance with Good Utility Practice.

and according to NERC [85]

an electric system or systems, bounded by interconnection metering and telemetry, capable of controlling generation to maintain its interchange schedule with other Control Areas and contributing to frequency regulation of the Interconnection.

Regardless of whose definition is being used, the important notion of a control area is the balancing of supply and demand while maintaining an interchange schedule with other control areas so that the overall reliability of the interconnected network is preserved at an acceptable level.

The intended purpose of forming an ISO covering multiple control areas is to provide access to a larger regional transmission system for an increase in reliability as well as an increase in efficiency. The consumers in different control areas are allowed to choose their suppliers across the control area boundaries while paying only a single transmission charge assigned by the ISO. The ISO’s role is limited to approving or refusing proposed bilateral transactions based on so-called available transfer capability (ATC) computation. The ATC computation yields additional transfer capability across the control area boundaries on top of the current operating conditions by assessing the operating constraints for the entire interconnected network. Once the proposed transaction is approved by the ISO, the trade between the parties take place according to the proposed transaction, and the ISO imposes the transmission charge previously agreed upon. This transmission charge is subject to strict regulation. However, it turns out that approving proposed transactions based on the ATC is highly inefficient across multiple control area boundaries. The reason for the inefficiency is related to the fact that it is almost impossible to establish accurate operating constraints over the multiple
control areas. The details of this discussion are in to Chapter 7.

4.2 Simplifying models for representing the markets in the California, Pennsylvania-New Jersey-Maryland (PJM) and Midwest regions

The simplifying models for representing the markets in the California, Pennsylvania-New Jersey-Maryland (PJM) and Midwest regions are introduced in [54]. The models recognize that the grid operator exists as the final authority in administering the market activities among suppliers, consumers and other market participants. The models assume that, by having independence from the market participants, the grid operator is indifferent to the financial consequences in a market environment, which is the case with the ISO's described in the previous section. Depending on the market structures supported by the regional grid operator, the models are designated as the voluntary system operator model, the mandatory system operator model and the multilateral transaction model.

4.2.1 Voluntary system operator model; California region

The voluntary system operator model supports a multi-tiered market structure which is intended to minimize the grid operator's influence on market activities by network users while achieving an acceptable level of reliability [54]. Figure 4-1 shows the basic schematic of the model. In this model both explicit bilateral

![Diagram](image)

Figure 4-1: Voluntary System Operator Model

and centralized market-based trades are allowed. The presence of spot market transactions is desired due
to the need for a continual balance of instantaneous supply with uncertain demand, while direct access and customer choice are achieved via bilateral trades.

This is similar to the current market structure in the California region. On one hand, the short term energy trade in the region is conducted by the PX, represented as the market for energy in Figure 4-1. On the other hand, the longer term trade is agreed upon through bilateral contracts among the participants and is proposed for physical access to the ISO with the help of SC’s, which is represented by the bilateral contracts between the generator and the loads and the bids between the market participants and the grid operator in Figure 4-1. The transmission payment from the grid operator to the transmission owners in Figure 4-1 indicates the license plate transmission charge and congestion charge collected by the ISO and transferred to the transmission owners.

It is clear from the model that no commitment can be made for longer term physical access to the network because of the complete separation of transmission owners from the market participants. This is an inherent problem for having only a non-profit organization, the grid operator, oversees the systemwide market process.

4.2.2 Mandatory system operator model; Pennsylvania-New Jersey-Maryland (PJM) region

The mandatory system operator model is introduced in [54] based on the existing practices of tight power pools. In this model the grid operator exists as the sole centralized market for conducting economically and functionally bundled energy and transmission trades. Figure 4-2 describes the relation among the market participants. As shown in Figure 4-2 the market participants are required to submit bids to the grid operator,

![Diagram](image)

Figure 4-2: Mandatory System Operator Model

who also functions as the market for energy, in order to gain access to the network and meet their energy requirements. Based on bids, the grid operator determines simultaneously the generation dispatch and the
allocation of transmission capacity by finding the smallest generation cost operating conditions for meeting the load. The price for bundled energy and transmission is computed as a result, and the short term supply and demand are balanced systemwide until the next time the market is cleared. No physical trade among the participants is possible other than through the short term energy market managed by the grid operator in this model.

This is similar to the current market structure in the PJM region. The market participants in the region rely on the ISO for meeting their energy requirements, which are represented as the grid operator/market for energy in Figure 4-2. The payment from the grid operator to the transmission owners in Figure 4-1 indicates the transmission charges collected by the ISO, including the congestion charge.

It is clear from the model that no commitment can be made for longer term physical access to the network because of the complete separation of transmission owners from the market participants.

4.2.3 Multilateral Transaction model; Midwest ISO

The multilateral transaction model is based on bilateral transactions among the market participants [54]. The model consists of three stages for completing transactions. First, individual buyers and sellers enter into bilateral trades without disclosing the price to the grid operator and propose the agreed upon trades to the grid operator for physical implementation. The grid operator, upon receiving the proposed transactions, makes the decision whether to allow the transactions or not based on an analysis of the transmission network constraints. If the proposed transactions do not violate any constraints, then they are accepted without any modifications. This is the most desired case. If the proposed transactions result in a violation of constraints, then the grid operator accepts none or only a part of the proposed transactions and suggests necessary modifications in the form of public information called, the "loading vector" [69]. Based on this information, the market participants make a new set of trades to satisfy the unmet demand while observing the system limits. Figure 4-3 shows the interaction among the various market participants in this model. The function of the grid operator is limited to verifying whether the proposed transactions may result in a violation of the operating constraints. Any transaction approved by the grid operator is subject to a transmission charge, based on a prior agreement in the bilateral contracts between the market participants and the transmission owners.

This is similar to the market structure under development in the Midwest region. In this region the market participants are allowed to enter into various bilateral transactions across control area boundaries in order to meet their energy requirements. The agreed upon transactions are then proposed to the ISO for physical implementation. The ISO, upon receiving the proposed transactions, approves or refuses physical access depending on the system reliability impact. If the proposed transactions are refused, the ISO provides a set of suggested revisions to the transactions. This process is represented by the schedule request and loading vector information exchange between the market participants and the grid operator in Figure 4-3.
Once the transactions are approved, the market participants are allowed physical access to the network subject to the transmission charge.

Because the parties involved in the transaction receive physical access almost unconditionally according to the contract, once the ISO approves a transaction, the ISO may grant approval only for a short term since any longer term commitment requires risk taking. From the perspective of a non-profit organization by design, the ISO is not allowed to take on any risks. The transmission owners, on the other hand, have the necessary financial assets to take on the risk but may not be able to take on the risks of the market participants since the owners are completely separated from the market process by the grid operator, as is evident from Figure 4-3.

4.3 Some critical observations

As is well recognized, deregulation has brought about strong incentives, through the functional unbundling process, for individual entities with their own business objectives. The ultimate challenge is to have all the pieces work toward social welfare improvements over longer periods of time under various uncertainties when the individual businesses operate to meet their own objectives. It is in putting these pieces together, and understanding their interplay, that the proper market design rules and the right regulation play a fundamental role, which gives the incentive to offering value to others with carefully designed transmission provision.

Using the models given in Figures 4-1 through 4-3 we have asserted that the restructuring process to date has concentrated on establishing only the short term energy markets. However, several major features

---

6The measure of social welfare following deregulation is defined in Chapter 3.
unique to the electric power network highlight the importance of creating well functioning long term markets. These include the lack of practical means for storing, the lack of controllability of the transmission path and uncertainty in equipment availability.

**4.3.1 Observation 1: lack of practical means for storing electricity and its implication on the need for forward markets for energy**

In much existing literature, the argument involving the lack of practical means for storing electricity has been used primarily to support the need for well functioning short-term (spot) markets for balancing supply and demand when providing electricity competitively. It has become clear only very recently that a much more dominant effect of having no practical means for storing electricity is the long term shortages of capacity. For example, the California energy crisis in the spring of 2001 is due to a lack of generation planning based on long term market activities because it is very difficult, close to impossible, to build more capacity based on short-term spot electricity price signals. Industries with sizeable inventories (including the Federal reserve for gas and oil) are capable of alleviating a shortage with their inventories while new capacity is sought after and, moreover, the value of the storage is exactly dependent on the typical delay in developing new investment and additional capacity. Price elasticity on the demand side provides a temporary solution to a shortage, particularly in a society used to high quality, relatively inexpensive electricity service. However, the fundamental solution to the problem of long term shortages of capacity is to create well functioning long term markets for energy where a stable price signal is generated that gives real incentives for generation investment.

**4.3.2 Observation 2: lack of practical means for storing electricity and its implication on the need for forward markets for transmission**

At this point it must be recognized that it is practically impossible to have a well functioning forward market for energy without the judicious provision of transmission over the longer term parallel to the long term markets for energy. The judicious provision of transmission includes the ability to provide appropriate risk hedging mechanisms for eliminating price volatility as well as guaranteeing physical access when required. The argument for such mechanisms is two-fold and is directly related to the lack of practical means for storing electricity. First, short-term physical delivery is impossible and/or very inefficient without the right delivery infrastructure. Second, the forward markets for energy are very sensitive to the existence of meaningful long term markets for transmission. A void in the systematic delivery infrastructure is sure to impinge on the long-term market liquidity in energy markets (including local market power effects). Moreover, without long term mechanisms for valuing transmission service, there will be no investments in reinforcing the existing grid; the transmission businesses are likely to have existential problems. The main objective for creating long term markets for transmission should be to have designs which have a long-term positive impact on the
energy markets and create a basis for sustainable transmission business at the same time.

Once this basic objective of the long term markets for transmission is understood, we may then recognize further complexities unique to the transmission process. These complexities are related to the lack of controllability of the transmission path. Since the direct transfer of electricity through a particular path within an interconnected network is not possible, various boundaries existing in the network, including the state, ownership, and energy market boundaries, need to be respected under a well-defined regulatory setup [37]. This is a critical problem especially in grid systems composed of multiple transmission owners such as in the Midwest regions. As the role of the control areas begins to diminish under the open access requirement, it is clear that coordinating inter-regional transactions is a serious issue and should be responded to with a market mechanism for providing supply and demand curves for inter-regional transfers via tie-lines. In [40] a possible minimal coordination approach to this problem is provided; this is the first solution of its kind in which the imports/exports are viewed as commodities with their own supply and demand functions.

4.3.3 Observation 3: hidden reliability risks in current market development under the old regulatory paradigms

As the functional unbundling process takes place, resulting in various entities possessing distinct roles and rights, the reliability unbundling process needs to follow it in order to define market participants' respective responsibilities. Only then, is the true value assigned to many services provided by the entities, especially the transmission services of the TP. The current industry restructuring process, unfortunately, seems to miss this fact and, consequently, there has been very little effort toward considering the provision of transmission services to energy market participants at the proper value. Instead, the overall thinking has focused on cost minimizing under a guaranteed rate of return on capital investments. This is generally viewed as a "safe" no risk approach by the TP’s. Unfortunately, the same providers are not realizing that the regulatory requirements which require the TP’s to ensure (at least) short-term reliability are full of hidden risks in the current market development, and are very dependent on the contractual conditions under which the transmission capacity is provided to the system users.

Possibly the most notable example of these hidden risks is the adopting of the FTR’s sold by the ISO’s in the California and PJM regions on the part of the transmission owners. It is pointed out in [41] that in a combination of short-term congestion pricing using nodal (or zonal) pricing together with the FTR’s, the network related risks are improperly transferred from the holders of the FTR’s to the market participants and not to the ISO’s who are responsible for issuing the contracts. Thus, the actual short-term reliability related risks are borne by the customers served by these ISO’s. Specifically, whenever the committed FTR’s are not simultaneously feasible, the actual congestion costs will be higher than made available from the charge paid by the holders of the FTR’s. Since the FTR’s are intended to make the holders financially indifferent, they must be paid the congestion costs incurred. This charge is seen ex post by the customers or the transmission
owner as discussed in the previous sections. Either way, a TP is vulnerable because the delivery charges are increased and a transmission may lose customers, who may choose alternate, less expensive delivery routes such as distributed generation.

This particular possible mechanism for transmission provision illustrates the very critical regulatory role of approving one market solution over another. It is an example of typical issues which arise when proposing transmission provision while observing the reliability criteria developed under the old regulatory paradigms, the obsolete reliability standards and their mutual interplay [32], [46].

4.3.4 Observation 4: need for new reliability standards

In [32], a very careful examination is urged of the underlying paradigms of the reliability standards now that the industry is restructuring. It is strongly suggested that this should be done urgently and prior to proceeding with any rule-making and/or legislation regarding the reliability of the US Interconnected Grid. Once the first assessment is done, a carefully designed R&D agenda for understanding the interplay of regulatory, economic and engineering innovations should be established, possibly at the inter-agency level of several government agencies. The questions recently raised regarding the rule-making in reliability standards can no longer be answered in a meaningful way without this major effort [32].

The industry is undergoing a fundamental change in its operations affected by the newly evolving technologies and/or the regulatory changes. Currently, there is a tremendous misfit between what the operating and planning practices are and what might be possible under these changes. Deregulation has brought strong incentives, driven by market economics, to old and new business entities. However, only a carefully induced interplay among (partial) regulation, economic incentives (the pricing of products and services) and the engineering/technical innovations can lead to an overall gain (social welfare over prolonged periods of time), while leaving enough room for decentralized decision-making by various entities.

Possibly, the hardest connection to make concerns the relations between the market specifications (contracts for products and services in the new industry) and the traditional industry standards (operating and planning) developed under qualitatively different regulatory rules.

The most relevant change of paradigm has to do with how various uncertainties are presented when one operates and plans the system being managed [34]. It is quite striking to recognize that reliability-related risk management must go hand in hand with the contractual specifications for products/services in the new industry. Understanding this concept leads to the notion of reliability unbundling. The implications of this unbundling on business and the quality of electricity service in the eyes of the customers are considerable.

One can identify at least three qualitatively different sources of uncertainty in the changing industry: (A) regulatory uncertainties, (B) market designs and (C) equipment status/functionality.

Traditional reliability standards, the ones which the US Department of Energy wishes to enforce into law for example, concern only (C) without considering (A) and (B) [35]. At this point it should become clear
that the intended objectives of the restructuring process will not materialize unless a very serious look into the basic paradigms of unbundled reliability under competition is taken.

Here are a few key suggestions to support the fundamental problem at hand:

- Suggestion 1: The (N-1) reliability standard must be replaced by a qualitatively different standard [34].

- Suggestion 2: Reliability-related risks need to be taken by different entities instead of by utilities alone, which are often defined as so-called providers of the last resort. A portion of the electric service is likely to be provided through bilateral arrangements, in which adequate supply is ensured by the contractual agreements between the parties involved. The remaining users must be provided (as they are now) by the providers of the last resort, which are the remnants of the old utilities. This puts a tremendous burden on the providers of the last resort since, according to this old framework, they are expected to manage all uncertainties created by the market/regulation, without adequate financial incentives (less profit to be made on the supply side). This clearly implies an unbundling of reliability contributions.

- Suggestion 3: The market design should accommodate these suggested changes. How suggestions (1) and (2) are managed depends on the market design in place.

Based on this, it is fairly straightforward to understand that the short term reliability requirement imposed on the TP, for instance, cannot be met in an unconditional way unless the reliability-related risks are well understood, and the right incentives given to the parties to meet their share of the reliability risks.

It is not an overstatement that the California energy crisis in the spring of 2001, in particular the financial status of the transmission companies, is strongly impacted by the marketers not taking any (financial) responsibility for reliability-related risks. Instead, it is assumed the utilities will do so unconditionally, and without any financial compensation. If this issue is not sorted out as the overall market design, including the long term markets for transmission, is proceeded with, there will be no sustainable transmission business in the future. For further discussion of the operating and planning paradigms under open access, and the notion of reliability unbundling, refer to [46].

4.4 Proposed market design in this thesis

The proposal for the formation of overall markets with the new structure for the TP made in this thesis takes the basic observations described above into consideration. To start with, the proposal for the new TP differentiates between two qualitatively distinct cases: (1) in which all network assets are owned by one single owner, for instance, a traditional transmission company serving a single energy market; (2) in which different transmission owners facilitate transactions in a multi-market setup.

It turns out that what differentiates these two scenarios are the types of (transmission) services and the ownership of these services. We observe that it is not the size of an overall market, but this ownership
of network assets and the relation to the energy market(s), that ultimately determines how the markets for energy and transmission are designed and their performance is measured. In the subsequent chapters, we define the transmission services and their ownership. Then, we propose a possible design for the entire market (energy and transmission) that will provide systematic incentives to all parties.

4.4.1 Single transmission network ownership within a single energy market

This setup is the simplest one and is applicable to an island-type electric power grid, owned by a single transmission provider and electrically disconnected from the rest of the world (The National Grid in the UK is an example of this). The "product" being sold by the transmission owner to the system users is the total transmission capacity for each line in the network. The problem is to establish a mechanism for investing in a larger line capacity at the places where it is most valuable to the system users, operate the existing wires so that the most is made out of the existing designs short-term and have meaningful financial mechanisms that give the right incentives for this to take place. We stress that the basic role of a well functioning market is its ability to provide incentives for investment and longer-term delivery contract arrangements while possessing a system operator within the TP structure who is concerned primarily with short-term operations and in assisting with the projection of market demand for transmission planning purposes.

The basic setup of the entire market design proposed in this thesis is shown in Figure 5-1. The basic role of this entire market design is to accomplish a committed performance, and to nurture the long-term investments necessary to provide transmission access through the forward market, which is further traded through the secondary market. The new market design comprises the following entities:

- A) an Independent Transmission Company (ITC) selling its own products (the transmission capacity of each line). The income from these sales is used to invest in enforcing the transmission grid.

- B) a pro-active (Independent) System Operator (SO) who cooperates with the ITC to implement the contracts for delivery established between the ITC and the buyers of the product (the network users). The physical implementation of the contracts established by the ITC is likely to be best carried out by the SO to further ensure independence. The SO should also cooperate with (or assist) the ITC in determining how much line capacity is available for sale.

- C) an on-line information infrastructure providing well-defined specifications of the product availability, line by line. This information is continuously updated by a TP (an ITC/SO team).

- D) a secondary market for transmission in which multiple owners of each transmission product (portions of the total line capacity) meet under well-defined market rules to trade their products over time. They are generally trading jointly owned products purchased directly from the TP.
An ITC and markets for transmission  In Chapter 5 we describe a fundamental structure for the TP composed of the ITC and the SO. Under the proposed structure, the ITC and the SO are two entities working cooperatively to carry out the functions of the TP. The entities are differentiated through the ownership and the operational authority. Roughly speaking, the ITC owns the regional network, provides various services connected with the longer term (physical and financial) energy trade, and carries out related functions, including making investment decisions. The SO, on the other hand, has operational authority over the entire network, provides many services linked to the shorter term (physical) energy trade, and carries out associated functions, including managing transmission congestion.

At the minimum, there are three groups of entities and three infrastructures important for a proficient management of the electric power network. The three groups refer to the regulator; the TP, composed of the ITC and the SO; and the market participants consisting of generators, loads and marketers. The three infrastructures are the spot market for energy balancing, the forward markets for transmission and the open access same-time information system (OASIS). Chapter 5 describes the role of the TP with an emphasis on the ITC's functions and the forward markets for transmission.

Secondary market for transmission and supporting infrastructures  In Chapter 6 we discuss two infrastructures important for proficient management of the network, namely the secondary markets for transmission rights and the open access same time information systems (OASIS).

Following the restructuring process the participants in the electric power industry are engaging in complex market activities to meet their electricity needs. Hence, the value of the energy as well as the transmission portion of the electric services should be determined by employing the market mechanism. These values, once determined, must then be communicated to the market participants in order to price various contracts.

Many market participants enter into forward (delivery) contracts for energy. The forward price may be described as the spot market price for delivery of a commodity at a fixed time in the future. As a counterpart to the forward contract marketplace for energy, the secondary market for transmission provides the necessary mechanism for supporting the market activities so that a change in the value of the transmission portion of the electric services is readily conveyed to all of the market participants. Here the market participants may be the holders of the physical transmission rights, the holders of the financial transmission rights and/or the bidders in the spot market.

Without the presence of the secondary markets for transmission rights, the ITC relies solely on its expertise gained by observing the transmission charges imposed on the market participants in the spot market in order to determine the prices to be charged for various transmission rights. This creates the open loop computation of the charge. However, with the presence of the secondary market for transmission rights, the ITC can observe the change in prices in the secondary markets for equivalent rights and take this into consideration in determining the new prices, i.e., in a feedback fashion.

With the introduction of the secondary markets for transmission rights we can compare the workings of
the transmission rights in the form of the intermediate term transmission contracts proposed in Chapter 5 with the transmission congestion contracts (TCC) and the flowgate rights.

4.4.2 Multiple transmission owners, multiple energy markets: inter-regional
transmission organization (IRTO)

In Chapter 7, we describe the provision of transmission in a multiple control area setting. In each control area it is assumed that a different market structure and tariff system may exist. This is the case with the interconnected network of the New England, New York and PJM systems.

We first describe the advantages and disadvantages of having interconnections with neighboring control areas. We then describe the newly proposed market mechanisms (and transmission provision) for implementing inter-regional transactions. The proposed mechanisms are then contrasted to the methods under the vertically integrated utility scheme and under the present restructuring process. Finally, the mechanisms are compared to other methods recently proposed in the industry.

It is shown that a new structure is essential for fostering the efficient operation and planning of the interconnected electric power network while also ensuring reliability.
Chapter 5

Independent transmission company (ITC) for profit and markets for transmission

In this chapter we propose a fundamental structure for the transmission provider (TP) composed of the independent transmission company (ITC) and the system operator (SO) [6]. Under the proposed structure, the ITC and the SO are two entities working cooperatively to carry out the functions of a TP. The entities are differentiated through ownership and operational authority. Roughly speaking, the ITC owns the regional network, provides various services connected with the longer term (physical and financial) energy trade, and carries out the related functions, including the making of investment decisions [38]. The SO, on the other hand, has operational authority over the entire network, provides many services linked to the shorter term (physical) energy trade, and carries out the associated functions, including the managing of transmission congestion. Figure 5-1 shows the overall market composition under the newly proposed structure. At minimum, there are three groups of entities and three infrastructures important for proficiency in the electric power network functions. The three groups are the regulator; the TP, composed of the ITC and the SO; and the market participants, consisting of the generators, loads and marketers. The three infrastructures are the spot market for energy, the secondary market for transmission and the open access same-time information system (OASIS). This chapter describes the role of the TP, with an emphasis on the ITC, and the TP’s relations with the regulator, the market participants and the spot market for energy. The discussions of the secondary market for transmission and the OASIS is left to Chapter 6.

It is asserted that the new structure of the TP is essential for fostering the operation and planning of the electric power network by the TP with a desirable level of efficiency and reliability while also supporting the regional energy markets under the new price-cap-regulation (PCR) scheme proposed in this thesis [2].
5.1 Functions and services

The three principal functions of the TP are described in Chapter 3 during the development of the new PCR scheme to be imposed on the entity. They are (1) making investment decisions about transmission, (2) making expenditure decisions about the control and maintenance efforts and (3) making pricing decisions for congestion management. In that chapter these functions are, for example, expressed as the control variables to the optimization problem in Eq. (3.74), which describes the systemwide social welfare maximization problem for the TP following the restructuring process.

It is stressed in the same chapter that the advantage of the newly proposed PCR scheme over the traditional cost-of-service regulation concerns the transfer of the operation and planning authority to the TP from the regulator. Under the cost-of-service regulation scheme, it is the regulator who ultimately makes the investment and pricing decisions and entrusts the TP with making the desirable expenditure decisions, in order to achieve a desirable level of efficiency and reliability in the management of an electric power network as described by the optimization problem in Eq. (3.61) [66]. The key to the optimization problem is to create the essential network capacity using the most economical means and to distribute the existing capacity to the participants with a higher value for the network, which requires estimating the cost functions of the TP and forecasting the demand and supply functions of the loads and the generators, respectively. However, as
is pointed out, the TP is in a better position than the regulator to solve the optimization problem because not only does the TP know its cost functions without needing to estimate them, but also the TP is believed to possess a better understanding of the demand and supply functions since the TP is in the market dealing with the loads and the generators actively in order to provide them with their transmission portion of the electric services. Thus, by creating an incentive structure where the TP increases its profit by (1) increasing the systemwide social welfare and (2) transferring the operation and planning authority, the regulator may actually expect higher overall efficiency and reliability.

Once the demand and supply functions are known for each hour, then the actual amount of the transmission capacity to be distributed to individual participants can be readily computed by solving the optimization problems in Eqs. (3.76) and (3.77). If all trades among the market participants are conducted through the spot market for energy, then the TP can discover the demand and supply functions of the market participants by offering transmission capacity through hourly congestion pricing. That is to say, the congestion pricing mechanism described in Chapter 3 alone is adequate for dealing with the loads and the generators in the market under the proposed PCR scheme. As described in Chapter 4, however, the market participants are engaged in various market activities in order to offer and to acquire electricity according to their evolving needs. Most of these market activities are initiated as purely financial and thus actually have no immediate impact on the network operation. However, as some of these activities become physical exchanges requiring the actual transport of electricity from generation source to load sink, the accompanying transmission capacity needs to be available for purchase so that the participants can carry out these physical exchanges. This is where a TP may gain considerable insights into the demand and supply functions of the market participants by offering a transmission capacity matching these physical exchanges. It should be recognized that most financial contracts turn into physical exchanges on time scales much longer than an hour. Among many reasons, the time scale for which the unit commitment decisions are typically made (a day, to sometimes weeks, in advance) is why this is the case. Plus, not every financial contract requires the same type of transmission capacity services. Consequently the TP needs to offer more network services than just the hourly congestion pricing in order to participate in every phase of the energy market activities.

There are a couple of important features to be considered regarding a TP’s ability to offer transmission capacity should the need arise.

The first is related to instilling confidence in the market mechanisms by which electricity is provided from the generators to the loads. The core of the market for a commodity is that there is a set of pre-determined rules and that if the participants enter into various contracts following these rules, then the actual exchange of a commodity can take place according to these contracts whenever desired. Unless this core is satisfied, there is no binding principle of economics under which competitive market functions would provide the commodity in question with a desired level of efficiency. Thus, it requires a market mechanism by which the TP can offer network capacity and the market participants can acquire the transmission to match their physical exchanges whenever desired.
The second is related to establishing a process for the TP to estimate the market participants' overall demand for network capacity. The only viable process for such estimation is to allow the TP to also actively participate in the market process throughout so that there is a constant communication between the market participants and the TP concerning the transmission network. This requires a clear market mechanism that allows the TP to offer the desired transmission services to the market participants in the energy market without compromising the integrity of the market while remaining a natural monopoly.

Clearly this implies a different role for the TP than that under the vertically integrated utility structure or than under the current development as described by one of three structures in Chapter 4: the multilateral transaction model, the voluntary system operator model and the mandatory system operator model. Figure 5-2 shows this changing role of the TP in the evolving electric power industry [70]. In the dependent phase a

![Diagram](image)

Figure 5-2: Changing Role of Transmission Provider

TP functions as a part of vertically integrated utility as described in Chapter 2. In the passive phase a TP stands alone and oversees overall market activities under the rate-of-return regulation as described in the first half of Chapter 3. The market participants are required to submit, in the form of bids, their intended use of the system to the TP and, based on that information, the TP allocates the existing transmission capacity following the strict rules set by the regulators. The TP assumes no financial responsibilities and has minimal interactions with market participants while the regulator is responsible for approving the transmission charge, and therefore is ultimately responsible for the expansion of the network. In the active phase a market mechanism is set up so that a TP participates in every phase of the energy market activities. The TP actively learns the desired usage for the network by the participants rather than passively accepting the intended usage expressed in their submitted bids.

In the following sections we describe the minimum network services that must be provided by a TP for three different time scales: the long term (longer than one year), the intermediate term (a year to a season), and the short term.
5.1.1 Long term network services

Long term network services refer to any *point-to-point* network capacity offered through long term transmission contracts by the ITC in increments of a year starting from the year following the current one. These services are provided without the regulator imposing any direct regulations on ITC.

As mentioned in Chapter 4 the market participants enter into various forward contracts ranging from 1 year to 5 years in time for hedging purposes. Since the exact contents of these contracts are not of particular interest here, we make a simplifying assumption that there are two types of long term contracts, namely long term hub-based contracts and long term point-to-point contracts. The hub-based contracts are traded through an organized (power) exchange while the point-to-point contracts are entered into by two private parties. Here hub refers to a financial institution responsible for conducting the exchanges, rather than a specific physical location within the network, where the energy contracts can be offered by specifying either the location of the source bus or the location of the sink bus without specifying the location of the counterpart buses. This is one of the unique features proposed in this thesis that differentiates electric power network economics from that of other commodities. For example, hub in the crude oil trade may refer to the warehouse location to which the physical commodity is delivered and received during the duration of the actual exchange [56]. The difference chiefly arises from the lack of a practical means of storing electricity.\(^1\)

The long term hub-based contracts specify at least the following four elements: the location of the source bus, \(g_i\) (or the sink bus, \(d_j\)), the amount of energy to be delivered, \(Q_{gi}\) (or \(Q_{dj}\)), the price for the energy, \(p_{gi}(Q_{gi})\) (or \(p_{dj}(Q_{dj})\)), and the duration of the contract, \([T_s, T_e]\). The variables, \(T_s\) and \(T_e\), denote the beginning and end point in time for the exchange, respectively. In general, this information is publicly posted. Similarly, the long term point-to-point contracts include at least the following five specifications: the location of the source bus, \(g_i\), the location of the sink bus, \(d_j\), the amount of energy to be delivered, \(Q_{dj-gi, e}\), the price for the energy, \(p_{dj-gi, e}\), and the duration of the contract. This information, on the other hand, is usually proprietary only to the parties entering into the contract.

Suppose there are \(N_B\) buses in the network.\(^2\) Then, the ITC may offer up to \(2N_B(N_B - 1)\) long term transmission contracts\(^3\) for any given time. The coefficient of 2 accounts for the dual directionality of flow in point-to-point exchanges. The price at which each of these \(2N_B(N_B - 1)\) contracts is offered mainly depends on the expectations of the ITC about the transmission price to be charged in order to accommodate the transport of electricity specified by the amount of electricity, the location of the generation source, the location of the load sink and the duration of the exchange. The prices of transmission contracts for the same two locations in a point-to-point exchange may differ significantly depending on the direction of the

---

1 It becomes clear later in the chapter that not designating a physical location for the hub is important because, in this thesis, the transmission charges are not additive under the proposed PCR scheme.

2 The number of buses in a network, \(N_B\), is always less than or equal to the sum of the number of generator buses, \(N_G\), and the number of load buses, \(N_D\), since some buses may connect both a generator and a load to the network. This technical point becomes important in Chapter 6.

3 The number of possible contracts available may be much lower depending on the demand for the contracts and on the level of data aggregation. A discussion on data aggregation is held later in the chapter.
exchange.

The organized power exchange clears the hub-based contracts by matching the generation source with the load sink based on the bid price of each contract as well as the long term transmission contract offered by the ITC, but it does this without putting any obligation to purchase the transmission contracts on the participants. It is up to the participants whether they want to purchase the transmission contracts to hedge the network related risks described in Chapter 4. Whether the purchase is made or not, the cleared bid pairs, \( g_i \) and \( d_j \), should be such that the bid price at the generation source, \( \rho_{g_i} \), and the bid price at the load sink, \( \rho_{d_j} \), satisfy the following relationship:

\[
\rho_{d_j} = \rho_{g_i} + \rho_{d_j-g_i,t}
\]  

(5.1)

where \( \rho_{d_j-g_i,t} \) denotes the price of the transmission contract offered for the proposed exchange between buses \( g_i \) and \( d_j \). Similarly, two private parties can enter into a point-to-point contract and may hedge their network related risks by purchasing corresponding long term transmission contracts as available. Figure 5-3 shows the information exchange between the market participants and the ITC concerning long term transmission contracts.

![Diagram showing information exchange between market participants and ITC](image)

Figure 5-3: The information exchange between the market participants and the ITC for long term transmission contracts.

Under an organized power exchange scheme the ITC plays the important role of providing the critical information needed to conduct the forward markets for entering into long term energy contracts. In fact, based on the description above, the energy forward market cannot clear unless matching long term transmission contracts are offered so that the price at the generation source and the price at the load sink are evaluated correctly where the transmission charge is considered as given in Eq. (5.1). Since the ITC is under no obligation to offer long term transmission contracts, this may be cause for alarm at first glance. However,
such concern can be addressed in two ways.

One, when there is a high demand for certain point-to-point long term transmission contracts, it is in the ITC’s best interest to offer those contracts because the ITC can collect transmission revenue through the contracts while significantly reducing the associated volatility. The elements influencing the volatility are both financial and regulatory. The financial reason is related to the intrinsically stochastic nature of the demand and supply functions, and may be addressed through the expressions developed in Eqs. (3.15) and (3.16) in Chapter 3. The regulatory reason is related to the proposed PCR scheme. If there are no longer-term transmission contracts offered by the TP, then all the network capacity is allocated through the spot market, and the transmission revenue is collected solely based on Eq. (3.79) as described in Chapter 3. Accordingly, the transmission revenue may become extremely uncertain, switching back and forth between the normal rate basis and the penalty rate basis when the congestion prices experience high fluctuation. Plus, by offering long term transmission contracts the ITC learns the demand and supply functions of the loads and the generators over the longer period. Along with the reduced volatility in transmission revenue, this newly gained knowledge about the projected demand and supply functions is essential for carrying out the planning functions of the ITC. Therefore, it is likely that the ITC will offer long term transmission contracts voluntarily.

The other way to address this concern is to remember that when no long term transmission contracts are being offered by the ITC, any financial institution may continue to offer hedging contracts against the network related risks described in Chapter 4. The difference between the hedging contract being offered by the ITC and that offered by the financial institution is that the long term transmission contracts offered by the ITC are initially financial but they become pseudo-physical transmission contracts offered by the SO during the year for which the contract is written whereas the ones offered by the financial institutions remain financial. This is because of the unique feature of the ITC transmission contracts proposed here: the participants who own the energy contracts with matching transmission contracts may claim priority in utilizing the network without any obligations. We describe the aspects related to the physical transmission rights in more detail in the subsequent section.

Note: Recently proposed hedging mechanisms by the SO’s in [31] and [80] are quite misleading since no SO is in a position to compensate the holders of hedging contracts financially for not meeting a transmission contract. At present there is considerable confusion about this. Any risk related to the problems of simultaneous feasibility [31] with the FTR’s, for example, in Pennsylvania-New Jersey-Maryland (PJM) and in New York (NY), is currently reflected in ex post customer charges, while the issuers of FTR’s, SO’s, are risk free! This is not a sustainable arrangement, and it is possibly the single strongest argument against SO’s (not transmission owners) selling long term transmission contracts.
5.1.2 Intermediate term network services

Intermediate term network services refer to the link-based network capacity designed and offered by the ITC and the SO, respectively, through intermediate term contracts up to the end of the year (or the season) at any time within the current year (or season).

According to the framework developed in Chapter 3, the TP carries out most of its functions at the beginning of each year. That is to say, at the beginning of each year, the ITC makes the actual commitment to the investment decisions about network enhancement\(^4\) and the expenditure decisions about maintenance procedures. Plus, the SO determines the level of expenditure for the control effort, software tools in particular. These decisions rest on the knowledge gained by the ITC through offering the intermediate term transmission contracts and are based on the expertise obtained by the SO through operating the network in near real time. As two entities working cooperatively to carry out the functions of the TP, the ITC and the SO share their knowledge and expertise so that they can maximize their overall profit under the proposed PCR scheme. Once the decisions are made, the SO determines, with reasonable accuracy, the anticipated available transmission capacity and the prices to be charged for the capacity for the entire year (or season) [74]. The ITC then designs the intermediate term contracts for each transmission line within the year (or the season) to be auctioned off by the SO.\(^5\)

Suppose there are \(N_T\) transmission lines in the network. Then, the ITC designs up to \(2N_T\) intermediate term contracts for any given time\(^6\) and makes them available for purchase through the SO by posting the respective projected prices, \(\rho_i^-\) and \(\rho_i^+\), for the contracts on each link for both directions per day. Along with the prices, the SO posts the maximum flow limits, \(F_i^{\text{max}}\), and the so-called power transfer distribution factors (PTDF’s) for line \(l\) with respect to bus \(i\), \(H_{il}\). The PTDF\(^7\) of line \(l\) with respect to bus \(i\) is the sensitivity vector of the line flow on the injection into bus \(i\) within the network [85]. Under the proposed ITC and SO structure, it is required that the maximum flow limits, \(F_i^{\text{max}}\), and the PTDF’s stay invariant throughout the year (or the season).

Consider two prospective market participants with a proposed exchange of \(Q_{d_j - g_i, \epsilon}\) between the generation source at bus \(g_i\) and the load sink at bus \(d_j\) over the period of time between \(T_s\) and \(T_e\). Then, the participants may hedge their delivery related risks completely by purchasing intermediate term transmission contracts of the amount \(H_{l(d_j - g_i)} \cdot Q_{d_j - g_i, \epsilon}\) on each line \(l\) in the network at the price of \(\rho_i^+\) or \(\rho_i^-\) depending on the direction of the flow. Figure 5-4 shows the information exchange among the market participants, the ITC and the SO for the intermediate term transmission contracts. The participants with long term

---

\(^4\)The more sophisticated a TP is, the more the majority of decisions depends on the long term transmission contracts even before the actual (budgetary) decisions are committed to at the beginning of the year.

\(^5\)The delegation of this task to the SO is merely for convenience and stems from the assumption that the SO has greater expertise in dealing with market participants on a day-to-day basis in the actual implementation of physical transactions. In other words, there is no significance in requiring the SO, instead of the ITC, to provide the service.

\(^6\)Again, the actual number of contracts available may be much smaller depending on the demand for the contracts and the level of data aggregation.

\(^7\)With the introduction of the PTDF’s, the operation of the electric power network is performed in the linearized regime as viewed by the market participants. The details on PTDF’s and the linearization approximation is given in Appendix B.
transmission contracts are required to convert the current year portion of the point-to-point contracts into link based intermediate term transmission contracts at the beginning of the same year. The actual amount of the intermediate term transmission contracts to be issued when converting the long term transmission contracts is based on the PTDF published by the SO at the beginning of the year.

Under the proposed structure, the ITC and the SO are required by the regulator to offer intermediate term transmission contracts so that the market participants may purchase these contracts in order to hedge their network related risks. The actual implementation of the services, including the pricing, is not subject to regulatory oversight, however. This concurs with the notion that such contracts are essential for fostering the efficient use of the existing resources, as described in Chapter 4. The current trend in the regulatory structure is also to require the TP to offer some types of longer term transmission contracts.\(^8\)

Similar to the TP’s incentives for offering the long term transmission contracts, it is in the ITC’s best interest to offer those contracts because the ITC can collect the transmission revenue through the contracts while significantly reducing the associated volatility and also while learning the demand and supply functions of the loads and the generators over the longer period.

The proposed intermediate term transmission contract belongs to the categories of both physical transmission rights and the financial transmission rights. What distinguishes the physical transmission rights from the financial transmission rights is what priority and obligation to use the network capacity are assigned to the participants holding the transmission contracts [48]. Based on this distinction the intermediate term transmission contract proposed here is a pseudo-physical contract because the owner of the contract has

---

\[^8\]The exact division of functions between the ITC and the SO in intermediate term network services is not critical. Various arrangements between the two are possible, as long as the TP as a whole (ITC & SO) performs the proposed function.
priority but no obligation in utilizing the network.

The actual mechanism under which the transmission capacity is allocated is as follows. At the beginning of each day, the SO identifies the holders of intermediate term transmission contracts with matching energy contracts having generation capability. This is considered a request for priority in using the network. The available network capacity is then allocated to these holders first, and the suppliers involved in the allocation process are scheduled for dispatch. If the transmission contracts are assumed to be sold in units of days, the dispatch is performed for the entire day. The intermediate term transmission contracts without matching energy contracts become financial transmission rights and the contract holders use the contracts to claim their portion of the transmission revenue specified in the contract.

The residual transmission capacity, after the initial allocation based on the intermediate term transmission rights, is then distributed to the participants in the spot market and to the participants with bilateral energy contracts for energy without matching transmission contracts. This process is described in the following section.

5.1.3 Short term network services

Short term network services refer to the allocation of residual transmission capacity in the spot market for energy. The related function of a TP is near real time congestion management as represented through the optimization problem defined for the so-called fast dynamics in Eq. (3.78) associated with the TP under the newly proposed PCR scheme in Chapter 3. The actual allocation of residual transmission capacity in the spot market is based on this optimization problem, modified to account for the portion of transmission capacity allocated previously to the participants with physical transmission rights. We assume that the SO conducts the spot market for energy.\(^9\)

Suppose the SO receives the bids from the loads, \(D_{d_j}\), and the generators, \(S_{g_i}\), for trading energy in the spot market. In addition, some other bids are made as pairwise transactions, \(B_{d_j-g_i}\), for implementing bilateral trades without matching intermediate term transmission contracts. Then, the SO solves the optimization problem given as the following:

\[
[Q_{G*}[k], Q_{D*}[k], Q_{B*}[k]]' = \arg \max_{Q_{G}[k], Q_{D}[k], Q_{B}[k]} \left[ \sum_{d_j} \left( \int_{Q_{d_j}[k]=0}^{Q_{d_j}[k]} D_{d_j}(Q_{d_j}[k], k)dQ_{d_j}[k] \right) \right. \\
- \sum_{i} \hat{\rho}_i^n [n](q_{i,d_j}[k] + q_{i,d_j}[k]) - \sum_{g_i} \int_{Q_{g_i}[k]=0}^{Q_{g_i}[k]} S_{g_i}(Q_{g_i}[k], k)dQ_{g_i}[k] \\
\left. + \sum_{b_{ij}} \left( \int_{Q_{b_{ij}}[k]=0}^{Q_{b_{ij}}[k]} B_{b_{ij}}(Q_{b_{ij}}[k], k)dQ_{b_{ij}}[k] - \sum_{l} \hat{\rho}_l^n [n](q_{l,b_{ij}}[k] + q_{l,b_{ij}}[k]) \right) \right]
\]

\(^9\)This assumption can be easily relaxed as shown in Chapter 4.
where $q_{l,d_j}$ and $q_{l,b_j}$ are the electric power flow on line $l$ caused by meeting the demand $Q_{d_j}$ at bus $d_j$ from the spot market and by accommodating the bilateral exchange of $Q_{b_j}[k]$ between buses $i$ and $j$ as requested by the users not participating in the spot market, respectively. The optimization problem is subject to the constraints in Eq. (3.45) and Ineq. (3.46) as well as the modified transmission line flow limit given by:

$$F_l(Q_G[k], Q_D[k], Q_B^*[k], F_{\text{inter}}[k]) \leq F_{l,\text{max}}[k] : \mu_l[k] \quad (5.3)$$

where $F_{\text{inter}}[k]$ is the transmission capacity initially allocated to the participants with intermediate term transmission contracts and the matching energy contracts.

Figure 5-5 shows the information exchange among the market participants, and the SO for the allocation of residual transmission capacity in the spot market for energy. Unlike the long term network services and the intermediate term services, short term network service is provided under strict regulatory oversight. According to the newly proposed PCR scheme, this regulation is in the form of ceiling prices represented as $\rho_l[n]$ and $\mu_l[n]$, as shown in Figure 5-5.

Once the ITC and the SO define the PTDF and the flow limits, it is assumed that the SO allocates all of the residual transmission capacity available.

With the various services described above, the ITC and the SO can actively take part in the market.
process by offering these services to the market participants as they make various supply and demand decisions for energy at different times. First, the transmission capacity offered in the long term time scale is financial in nature and sends signals useful for making planning decisions to the ITC and the market participants. Next, the transmission capacity offered in the intermediate time scale is pseudo-physical in nature and sends signals useful for deciding the expenditures for the control effort and the maintenance effort. Finally, the transmission capacity offered in the short term time scale is physical in nature and sends signals useful for managing the congestion within the network by the SO. This is because the value of the network capacity is reflected in the price and the quantity at which these services are exchanged. The pricing of these services is, therefore, significant for discovering the demand and supply functions of the market participants. Incidentally, along with yielding useful economic signals, offering transmission capacity in the longer term significantly reduces the volatility in revenues of both the TP and the market participants. In the following section we discuss the pricing of the various services described above, starting with short term network services.

5.2 Services and pricing

In pricing network services, it is important to identify any attributes that limit the prices. Unlike the competitive firm whose marginal cost function serves as both upper bound and lower bound in the pricing of its service, the regulated firm under the PCR scheme only has an upper bound, which is defined by the regulator. In the case of the newly proposed PCR imposed on the TP, there are two price elements subject to ceiling prices, namely the *ex ante* flow tax and the congestion cost. Chapter 3 describes how these ceiling prices are employed to set the prices for the transmission capacity. We examine these ceiling prices in the context of the network services described above.

5.2.1 Pricing of short term network services

With the posting of the PTDF by the SO for the entire year, n, the apparent operation of the electric power network is performed in a simplified linear regime. Suppose we apply the so-called DC load flow assumption as a method for linear simplification. Then, the expression of the flow on transmission line \( l \), \( F_l(\mathbf{Q}_G[k], \mathbf{Q}_D[k]) \), is given by:

\[
F_l(\mathbf{Q}_G[k], \mathbf{Q}_D[k]) = \sum_{g_i} H_{lg_i} Q_{g_i} [k] - \sum_{d_j} H_{ld_j} Q_{d_j} [k]
\]  

(5.4)

where \( H_{li} \) denotes the PTDF of line \( l \) with respect to bus \( i \). The details of DC load flow simplification and the PTDF are given in Appendix B.

Given the DC load flow assumption, we develop the pricing of the residual transmission capacity by the SO. As described earlier, the pricing of a short term network service is subject to strict regulation in the
newly proposed PCR scheme. Thus, at each hour \( k \) the SO may charge the market participants in the spot market for the network capacity on line \( l \) up to the transmission price, \( \rho_l \), given by

\[
\rho_l[k] = \begin{cases} 
\hat{\rho}_l[k] + \mu_l[k] & \text{if } \hat{\rho}_l[k] \leq \hat{\rho}_l[n] \text{ and } \mu_\zeta[k] \leq \mu_\zeta[n] \text{ for all } \zeta \text{ in the network, including } l \text{ itself} \\
\mu_l[k] & \text{otherwise, i.e. if } \mu_\zeta[k] > \mu_\zeta[n] \text{ for any } \zeta \text{ in the network, including } l \text{ itself}
\end{cases}
\]

where the ceiling prices of the \textit{ex ante} flow tax and the congestion costs are denoted by \( \hat{\rho}_l[n] \) and \( \mu_l[n] \) respectively, and \( k = (n-1)T + 1, (n-1)T + 2, \ldots, nT \).

The revenue allowed to the TP at each hour \( k \) from the allocation of the residual transmission capacity is given by

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

under the proposed PCR scheme, where \( r_{\text{penalty}} \) denotes the penalty rate imposed on the TP for violating the ceiling price of the congestion charge. Thus, the maximum allowed revenue can be obtained by solving the following optimization problem:

\[
\hat{\rho}_l^*[k] = \begin{cases} 
\arg \max_{\hat{\rho}_l[k]} \sum_l \sum_{d_j} \left[ (\hat{\rho}_l[k] + \mu_l^*[k]) (q_{l,d_j}[k] + q_{l,b_{ij}}[k]) + (\hat{\rho}_l[k] - \mu_l^*[k]) (q_{l,d_j}[k] + q_{l,b_{ij}}[k]) \right] & \text{if } \hat{\rho}_l[k] \leq \hat{\rho}_l[n] \text{ and } \mu_l^*[k] \leq \mu_l[n] \text{ for any } l \\
0 & \text{otherwise, i.e. } \mu_l^*[k] > \mu_l[n] \text{ for any } l
\end{cases}
\]

Eq. (5.7) is identical to the set of Eqs. (3.76) and (3.77) in Chapter 3. The optimization problem in Eq. (5.7) is complementary to the market clearing process in the spot market modeled in Eq. (5.2). We make another simplifying assumption about the shape of the demand and supply functions, submitted to the SO by the market participants that they are linear, i.e.,

\[
D_{d_j}(Q_{d_j}[k], k) = 2\alpha_{d_j}[k] \cdot Q_{d_j}[k] + \beta_{d_j}[k]
\]

\[
S_{g_i}(Q_{g_i}[k]) = 2\alpha_{g_i} Q_{g_i}[k] + b_{g_i}
\]

\[
B_{d_j-g_i} = 2\alpha_{d_j-g_i}[k] \cdot Q_{d_j-g_i}[k] + \beta_{d_j-g_i}[k]
\]

Then, Eq. (5.2) reduces to the following:

\[
[Q_{G^*[k]}, Q_{D^*[k]}, Q_{B^*[k]}] = \arg \max_{Q_{G}[k], Q_{D}[k], Q_{B}[k]} \left[ \sum_{d_j} (\alpha_{d_j}[k] \cdot Q_{d_j}^2[k] + \beta_{d_j}[k] \cdot Q_{d_j}[k]) \right]
\]
\[-\sum_{d_j} \sum_{l} \tilde{\rho}_l[n](q_{l,d_j}^{-}[k] + q_{l,b_{ij}}^{-}[k]) - \sum_{g_i} (a_{g_i} Q_{g_i}^2[k] + b_{g_i} Q_{g_i}[k]) + \sum_{b_{ij}} (\alpha_{d_j-g_i}[k] \times \]

\[Q_{d_j-g_i}[k] + \beta_{d_j-g_i}[k] \cdot Q_{d_j-g_i}[k]) - \sum_{b_{ij}} \sum_{l} \tilde{\rho}_l[n](q_{l,b_{ij}}^{+}[k] + q_{l,b_{ij}}^{-}[k]) \]

subject to

\[\sum_{g_i} Q_{g_i}[k] = \sum_{d_j} Q_{d_j}[k] : \lambda[k ] \quad (5.12)\]

\[Q_{g_i}^{\text{min}}[k] \leq Q_{g_i}[k] \leq Q_{g_i}^{\text{max}}[k] : \eta_{g_i}[k ] \quad (5.13)\]

\[\sum_{g_i} H_{l,g_i} Q_{g_i}[k] - \sum_{d_j} H_{l,d_j} Q_{d_j}[k] + \sum_{b_{ij}} (H_{l,g_i} - H_{l,d_j}) Q_{b_{ij}}[k] + F_{l,\text{int}}[k] \leq F_{l}^{\text{max}}[n] : \mu_{l}[k ] \quad (5.14)\]

where \( F_{l,\text{int}}[k] \) are the portion of the transmission capacity allocated previously to the participants with physical transmission rights, and the Lagrangian multiplier \( \mu^*[k] \) matches the congestion cost in Eq. (5.7). We apply the DC load flow assumptions to the flow constraint in Ineq. (5.14).

It is clear from Eqs. (5.5) and (5.6) that there may be a significant difference between the revenue collected by the SO and the revenue received by the ITC when the congestion cost, for constraining the flows within the operationally acceptable limits, \( F_{l}^{\text{max}}[n] \), exceeds the corresponding ceiling price. As discussed in Chapter 3, this difference is returned through some form of the rebate mechanism\(^{10}\) to the market participants in the spot market since the PCR scheme, by definition, is intended to protect the market participants from being over-charged by the TP. The market participants who acquired their network capacity through the intermediate term transmission contracts, however, are excluded from the rebate process since those contracts are sold without any regulation and thus command no regulatory protection.

Note that in Chapter 3 we make the claim that the demand and supply functions in Eq. (3.78) actually represent the participants' preference in energy usage through their overall market activities. On the other hand, the demand and supply functions in Eq. (5.11) represent the bids submitted to the SO in the spot market for energy. This difference is rather delicate one but in the following section we claim that, at the equilibrium, the price of the transmission capacity revealed through solving the optimization problem in Eq. (5.11) approaches the solution to Eq. (3.78).

\(^{10}\)The actual "optimal" process for this rebate mechanism is beyond the scope of this thesis. Thus, it is sufficient to employ modeling simplifications in order to treat this process as exclusive between the regulator and the loads as done in Chapters 2 and 3 and in [51] and to avoid further discussion of the process for the rest of the chapter. Nevertheless, it is conjectured here without proof that the rebate is not a sustainable process because, if it were, the spot market participants' behavior would be altered by increasing the demand for transmission capacity with an anticipation of the rebate, and this would create a greater difference between the ceiling price and the apparent price, thus increasing the rebate. Therefore, the rebate mechanism is there for the occasional deviations from the expected in the demand for transmission capacity.
5.2.2 Pricing of intermediate term network services

Suppose the ITC carries out the computation of the optimization problem given in Eq. (5.7) and solves for the price of the network capacity on line $l$ using Eq. (5.5) but from the expected value sense for the entire year $n$, i.e.

$$
\hat{\rho}^*_l[k] = \begin{cases} 
\arg \max_{\hat{\rho}_l[k]} \sum_l \sum_i \mathbb{E} \left\{ (\hat{\rho}_l[k] + \mu^*_l[k]) (q_{l,dj}^*[k] + q_{l,bij}^*[k]) + (\hat{\rho}_l[k] - \mu^*_l[k]) (q_{l,dj}^-[k] + q_{l,bij}^-[k]) \right\} 
& \text{if } \mathbb{E} \{\hat{\rho}_l[k]\} \leq \hat{\rho}_l[n] \text{ and } \mathbb{E} \{\mu^*_l[k]\} \leq \mu_l[n] \text{ for any } l \\
0 & \text{otherwise, i.e. } \mathbb{E} \{\mu^*_l[k]\} > \mu_l[n] \text{ for any } l
\end{cases}
$$

(5.15)

$$
[Q_G^*[k], Q_D^*[k], Q_B^*[k]]' = \arg \max_{Q_G[k], Q_D[k], Q_B[k]} \mathbb{E} \left\{ \sum_{d_j} (\alpha_{d_j}[k] \cdot Q_{d_j}^2[k] + \beta_{d_j} [k] \cdot Q_{d_j}[k]) 
- \sum_{d_j} \sum_{l} \hat{\rho}^*_l[n] (q_{l,dj}^*[k] + q_{l,dj}^-[k]) - \sum_{g_i} (a_{g_i} Q_{g_i}^2[k] + b_{g_i} Q_{g_i}[k]) 
+ \sum_{b_{ij}} (\alpha_{d_j-g_i[k]} \cdot Q_{d_j-g_i[k]}^2[k] + \beta_{d_j-g_i[k]} [k] \cdot Q_{d_j-g_i[k]}[k]) 
- \sum_{b_{ij}} \sum_{l} \hat{\rho}^*_l[n] (q_{l,bij}^*[k] + q_{l,bij}^-[k]) \right\}
$$

(5.16)

for $k = (n - 1)T_l + 1, (n - 1)T_l + 2, \ldots, nT_l$ and subject to

$$
\sum_{g_i} \mathbb{E}\{Q_{g_i}[k]\} = \sum_{d_j} \mathbb{E}\{Q_{d_j}[k]\} : \mathbb{E}\{\lambda[k]\}
$$

(5.17)

$$
\mathbb{E}\{Q_{g_i}^{\min}[k]\} \leq \mathbb{E}\{Q_{g_i}[k]\} \leq \mathbb{E}\{Q_{g_i}^{\max}[k]\} : \mathbb{E}\{\eta_{g_i}[k]\}
$$

(5.18)

$$
\sum_{g_i} H_l g_i \mathbb{E}\{Q_{g_i}[k]\} - \sum_{d_j} H_{ld} \mathbb{E}\{Q_{d_j}[k]\} + \sum_{b_{ij}} (H_{l g_i} - H_{l d}) \mathbb{E}\{Q_{b_{ij}}[k]\} + \mathbb{E}\{F_{l, int}[k]\} \leq F_l^{\max}[n] : \mathbb{E}\{\mu_l[k]\}
$$

(5.19)

Then, for any hour $k$, the expected price of the network capacity is given by:

$$
\mathbb{E}\{\rho_l[k]\} = \mathbb{E}\{\hat{\rho}_l^*[k] + \mu_l[k]\}
$$

(5.20)

Consider the pricing, $\rho_l(T_s, T_e)$, of the network capacity on line $l$ for the intermediate term transmission contracts over the duration of $T_s$ and $T_e$. Based on the expression in Eq. (5.20) the intermediate term transmission contracts need to be priced for purchase at $k = T_s$ as the following:

$$
\rho_l(T_s, T_e) = \sum_{k=T_s}^{T_e} (1 - \xi)^{k-T_s} \mathbb{E}\{\rho_l[k]\}
$$

(5.21)
where ξ denotes the discount rate of risk-free investment, and \( \rho_l[k] \) is computed from the expression in Eq. (5.20). This is because any other pricing mechanism yields a risk-free opportunity for profit, known as the arbitrage in economics [14] [21], to some entities at the expenses of other entities. The anticipated behavior of the entities facing the expenses prevents such opportunities, and thus leads to the pricing of the intermediate term transmission contract given by the expression in Eq. (5.21). This is stated more formally in the following lemma.

**Lemma 3**: The profit maximizing ITC charges the profit maximizing market participants at the price given by Eq. (5.21) for the intermediate transmission contracts.

**Proof:**

Suppose the actual price of the intermediate term transmission contracts, \( p^+_l(T_s, T_e) \), is lower than the price given by Eq. (5.21). Then, the market participants are presented with an arbitrage opportunity since the holders of the intermediate term transmission contracts are entitled to the prevailing prices of network capacity in the cases when the network capacity is not used for physical exchanges by the holders. The holders gain, at the expense of the ITC, the difference between the price that is paid and the price that is prevailing in the spot market over the duration of the contract, \([T_s, T_e]\), i.e.,

\[
\pi = \sum_l (p^+_l(T_s, T_e) - p^+_l(T_s, T_e)) \cdot \sum_{k=T_s}^{T_e} F_{l,financial}[k]
\]  

(5.22)

where \( F_{l,financial}[k] \) is the network capacity not utilized after having been purchased through the intermediate term transmission contracts. Plus, given this arbitrage opportunity, the spot market participants are likely to enter into intermediate term transmission contracts so that they reduce their cost for transmission charges. As more spot market participants enter into contracts at a transmission price lower than the prevailing price in the spot market, the revenue collected from accommodating the physical exchanges by the ITC through the SO also becomes smaller. Therefore, this results in the profit maximizing ITC increasing the price of intermediate term transmission contracts.

Suppose the actual price of the intermediate term transmission contracts, \( p^+_l(T_s, T_e) \), is higher than the price given by Eq. (5.21). Then, the market participants are again presented with an arbitrage opportunity since the marketers may offer to sell financial transmission contracts at a slightly lower price than \( p^+_l(T_s, T_e) \) but still higher than the price given by Eq. (5.21), and they may receive the difference for profit. Assuming that there is a demand for intermediate term transmission contracts even at a high price, the financial transmission contracts issued by the marketers reduce the profits for ITC, and this results in the profit maximizing ITC decreasing the price of intermediate term transmission contracts.

Therefore, the price of the intermediate term transmission contract needs to be set according to the expression in Eq. (5.21).

Given that the prices of intermediate term transmission contracts approach the prices in the spot market for energy, the equilibrium price computed by solving the optimization problem in Eq. (5.16) also approaches
the solution to the optimization problem in Eq. (3.78). The demand and supply functions of the loads and the generators in Eq. (3.78) reflect the market participants’ preference in energy usage. Since all of the physical exchanges of electricity among the market participants are determined through purchasing physical transmission rights in the form of intermediate term transmission contracts and through trading in the spot market, then at the equilibrium, from the expected value sense, the consumption and the generation of electricity by the market participants revealed through Eq. (5.16) at the entirety need to be equal to the corresponding energy usage computed by solving the optimization problem in Eq. (3.78). Thus, the prices of network capacity computed through solving the optimization problems in Eqs. (3.78) and (5.16) are identical when there is a unique solution to Eq. (3.78) as assumed in Chapter 3.

The above implies that, in order to compute the prices to be charged the intermediate term transmission contracts, the ITC may need to solve the optimization problem in Eq. (3.78) while estimating the demand and supply functions of the loads and generators, $D_{d_i}(Q_{d_i}[k], k)$ and $S_{g_i}(Q_{g_i}[k], k)$ respectively, in their entirety, instead of solving the optimization problem in Eq. (5.16) while estimating the demand and supply functions in the spot market for energy, $2\alpha_{d_i}[k] \cdot Q_{d_i}[k] + \beta_{d_i}[k]$ and $2\alpha_{g_i}Q_{g_i}[k] + b_{g_i}$ respectively, separated from the bilateral preference function, $2\alpha_{d_i-g_i}[k] \cdot Q_{d_i-g_i}[k] + \beta_{d_i-g_i}[k]$ and the portion of the transmission capacity allocated previously to the participants with physical transmission rights, $F_{l, int}[k]$. In the following section we review an approximate computational method, known as the probabilistic optimal power flow (POPF), for pricing intermediate term network services [74].

Approximate computational method for pricing intermediate term network services

In [22] probabilistic optimal power flow (POPF) is introduced as a tool for evaluating the likely use of a transmission network. Using this novel method the binding transmission limits can be identified under normal operating conditions with some probability. It turns out that the value of the network capacity can also be deduced from solving POPF.

POPF uses a Monte Carlo-based method to efficiently solve optimal power flow (OPF), taking into account transmission line flow limits and generation capacity constraints over the possible range of load demands. For our purposes the complete formulation of the OPF can be expressed as the following:

$$Q^*_{G}[k] = \arg \min_{Q_{G}[k]} \sum_{g_i}(a_{g_i}Q^2_{g_i}[k] + b_{g_i}Q_{g_i}[k])$$

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

$$Q^\text{min}_{g_i}[k] \leq Q_{g_i}[k] \leq Q^\text{max}_{g_i}[k] : \eta_{g_i}[k]$$

$$\sum_{g_i} H_{l_{g_i}}Q_{g_i}[k] - \sum_{d_j} H_{l_{d_j}}Q_{d_j}[k] \leq F^\text{max}_l[n] : \mu_l[k]$$

115
where $Q_{d_j}[k]$ is assumed to be inelastic and thus given ahead of time.

First, we construct the probability density function of system demand. Traditional utilities have published what is referred to as a load duration curve, which depicts the cumulative probability of the system load as shown in Figure 5-6 [74]. The probability density function of system load, $f_{Q_D}$ is computed then by taking the first derivative of the load duration curve as follows:

$$f_{Q_D}(Q_D^{[i]}) = \left( \frac{dF_{Q_D}}{dQ_D} \right)_{Q_D = Q_D^{[i]}}$$

(5.27)

where $Q_D$ is the total system load, i.e. $Q_D = \sum_{d_j} Q_{d_j}$.

Given the probability density function, the individual load pattern is computed along the incremental increase in the system load starting from its minimum. Because of the metering problem, a probabilistic modeling of the individual load pattern may not be easily inferable in real-life systems. In [74] an approach is suggested to cope with this problem, which is based on finding the demand at each load bus within the network for any sum of systemwide load by meshing the load patterns pre-determined for the peak $[PK]$, off-peak $[OP]$ and normal $[N]$ loading conditions through the membership function, as shown in Figure 5-7. Employing fuzzy logic, these patterns are meshed to create typical individual load patterns which match the
probability density function of the system load as follows, \( Q_D^{[i]} = [Q_{d_1}^{[i]}, \ldots, Q_{d_{N_D}}^{[i]}] \):

\[
Q_D^{[i]} = Q_D^{[i]} \left( \eta^{[N]}(Q_D^{[i]}) \frac{Q_D^{[N]}}{1TQ_D^{[N]}}, \eta^{[OP]}(Q_D^{[i]}) \frac{Q_D^{[OP]}}{1TQ_D^{[OP]}}, \eta^{[PK]}(Q_D^{[i]}) \frac{Q_D^{[PK]}}{1TQ_D^{[PK]}} \right)
\] (5.28)

where \( \eta^{[i]}(Q_D^{[i]}) \) is the membership function expressed as a function of the total system load. Finally, OPF is solved as in Eq. (5.23), using individual load patterns along the system load from the minimum to the maximum.

These resulting solutions to the OPF problems yield much useful information. For example, computing flows on transmission lines after OPF and fitting the flows against the probability density function yield the cumulative probability of transmission system usage as follows [74]:

\[
\text{Prob} \{ F_l \leq \bar{F}_l \} = \text{Prob} \left\{ F_l \left( Q_D^{[i]} \right) \left( Q_D^{[i]} \right) \leq \bar{F}_l \right\} = \int_{0}^{\bar{F}_l} f_{Q_D}(Q_D^{[i]}) dF_l
\] (5.29)

For our purposes, the more interesting result is the derivation of the congestion charge from the OPF solutions. Suppose for a particular transmission line \( l \) we keep track of its shadow prices, \( \mu_l^{[i]} \), from the OPF solutions corresponding to the total system load \( Q_D^{[i]} \). Then, the value of the transmission line \( l \) for the duration of \( T_s \) and \( T_e \) is given as the following:

\[
\mu_l(T_s, T_e) = \mu_l^{[i]} \text{Prob} \left\{ Q_D[k] = Q_D^{[i]} \right\} = \sum_{[i]} \mu_l^{[i]} \int f_{Q_D}(Q_D^{[i]})
\] (5.30)

where \( Q_D[k] \) denotes the total system load at hour \( k \).

In order to match the results of Eqs. (5.21) and (5.30) exactly, the time value of money needs to be included in Eq. (5.30) and to the OPF formulation given in Eq. (5.23) needs to be replaced with the expression given in Eq. (5.16).

### 5.2.3 Pricing of long term network services

As evident from the expressions in Eqs. (5.15) through (5.21), the pricing of intermediate term network services is, to a large extent, simply repeating the computations used for the pricing of short term network services over a desired duration. The only difference is that in the short term the computation is deterministic whereas in the intermediate term it is stochastic. This is the result of Lemma 3 given that the maximum flow limits, \( F_l^{\max} \), and the PTDF's stay invariant throughout the year (or the season). With respect to the pricing of the long term network services, however, the result of Lemma 3 still holds true, but the maximum flow limits and the PTDF's may no longer be invariant because of the transmission investment allowed at the beginning of each year. Therefore, when pricing long term network services, we need to consider not
only the value of the transmission but also the investment decisions about transmission.

In Chapter 3 we introduced the optimization problem in Eq. (3.74) related to the investment decisions about transmission under the proposed PCR scheme. Suppose the ITC carries out the computation of this optimization problem and attains the solution \( I_T[n], e_{tech}[n] \) and \( e_m[n] \) by forecasting the supply and demand functions of the market participants for \( n = 1, 2, \cdots T_I \). We use the notation \( T_I \) to denote the time scale used to assess the value of new investments in transmission. Then, based on these investment decisions, the configuration of the overall network can be determined, which maximizes the expected value of ITC’s profit defined in Eq. (3.72) over the same time scale. Since we restrict the investment decisions about transmission to be made only once at the beginning of each year, the corresponding network configuration remains invariant throughout the year. Finally, once the configuration of the network is specified, the maximum flow limits, \( F_{l_{\text{max}}}[n] \), and the PTDF’s can be computed for each year.

Now consider the pricing of the network capacity on line \( l \) for the entire year \( n \). Since the maximum flow limits and the PTDF are given for the year, this is no different than solving the problem of pricing intermediate term transmission contracts as given in Eq. (5.21) by pretending that \( T_S \) and \( T_E \) denote the beginning and end of year \( n \). This process can be extended for multiple years and thus can be applied to the computing of long term network service prices. From the perspective of the SO, however, it may not be desirable to simply extend the method used to price intermediate term transmission contracts in order to price long term transmission contracts because the corresponding \( F_{l_{\text{max}}}[n] \) and the PTDF’s are very close to being deterministic in the intermediate term case, but are quite stochastic, depending on the actual investment decisions in the future, in the long term case. Since the actual operation of the network is restricted to conform to the pre-defined \( F_{l_{\text{max}}}[n] \), and the PTDF’s, as is evident from the expression in Ineq. (5.14), the ITC and SO may be exposed to an enormous amount of risks if \( F_{l_{\text{max}}}[n] \) and the PTDF’s are defined ahead of time based on a projection over many years. Thus, it is preferable to introduce long term transmission contracts based on point-to-point rather than link-based network capacity despite the potential liquidity problem described in Chapter 6. The resulting pricing of long term transmission contracts between generation bus \( g_i \) and load bus \( d_j \) over the duration between \( T_S \) and \( T_E \) is given by

\[
\rho_{g_i \rightarrow d_j}(T_S, T_E) = \frac{T_x}{k=T_S} (1 - \xi)^{k - T_S} \sum_l \mathcal{E} \{ \rho_l[k] (H_{lg_i}[k] - H_{ld_j}[k]) \}
\]  

(5.31)

### 5.3 Illustrative examples

Consider the 3-bus electric power network shown in Figure 5-8. The network is composed of 3 generating substations, 2 load centers and 3 transmission lines.

We illustrate the pricing of network services by examining the operation and planning of the network over a 4 year period, i.e., \( n = 1, \cdots, 4 \). In this example, a year consists of 2 seasons, each having 3 days, and depending on the demand of the loads, the seasons and the days are differentiated as peak, shoulder and
off-peak. Each day is composed of 2 hours.

There are 13 generating units available in this network: 4 thermal units at bus 1, 7 hydro units at bus 2, and 1 gas-turbine unit and 1 nuclear unit at bus 3. Table 5.1 summarizes the marginal costs of these units in the form of a linear function, i.e. \( S_{gi}(Q_{gi}[k]) = 2a_{gi}Q_{gi}[k] + b_{gi} \). These units are assumed to have infinite generation capacities.

The expected system conditions for \( n = 1, \ldots, 4 \) are given as follows: Figure 5-9 shows the load characteristics at each bus for the first year. 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 the loads in year 2 is same as shown in Figure 5-9. In year 3 (\( n = 3 \)), the projected
Figure 5-9: Load characteristics in years $n = 1, 2$

demand of the loads increases by 5% over that of the previous year throughout the network while no change is expected to take place in the generation. Figure 5-10 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 as the previous year. This is shown in Figure 5-10.

Suppose the ceiling prices for the ex ante flow tax and the congestion charge are set to be $6.92$ and $45$ respectively for years 2 and 3 on the transmission line between bus 2 and bus 3 with no additional penalty for exceeding this limit under the proposed PCR scheme. Then, based on the discussion in Chapter 3 and the projected system conditions for years 1, 2, 3, and 4, Table 5.2 summarizes the optimal transfer capacity for
each line given the initial transfer capacities of $F_{(1)}^{\text{max}} = 150\text{(MW)}$, $F_{(2)}^{\text{max}} = 150\text{(MW)}$, and $F_{(3)}^{\text{max}} = 60\text{(MW)}$. The change in transfer capacity on line (3) is accomplished by expanding the transmission line capacity.

<table>
<thead>
<tr>
<th>Line #</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transfer capacity in year 1 (MW)</td>
<td>150</td>
<td>150</td>
<td>60</td>
</tr>
<tr>
<td>Transfer capacity in year 2 (MW)</td>
<td>150</td>
<td>150</td>
<td>80</td>
</tr>
<tr>
<td>Transfer capacity in year 3 (MW)</td>
<td>150</td>
<td>150</td>
<td>85</td>
</tr>
<tr>
<td>Transfer capacity in year 4 (MW)</td>
<td>150</td>
<td>150</td>
<td>80</td>
</tr>
</tbody>
</table>

Table 5.2: Transfer capacity of each transmission line in years 1, 2, 3, and 4 in the 3-bus electric power network example

through investment and control effort.

By solving the optimization problem for the allocation of transmission capacity given in Eq. (5.7) for the projected system conditions, the price of network services can be computed as given in Tables 5.3 through 5.6. Based on Tables 5.3 through 5.6 the pricing of the long term network services and of the intermediate

<table>
<thead>
<tr>
<th>Year</th>
<th>Season</th>
<th>Day</th>
<th>Hour:</th>
<th>line (1)</th>
<th>line (2)</th>
<th>line (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1:</td>
<td>$6.92$</td>
<td>$6.92$</td>
<td>$6.92$</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2:</td>
<td>$6.92$</td>
<td>$6.92$</td>
<td>$6.92$</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1:</td>
<td>$6.92$</td>
<td>$6.92$</td>
<td>$6.92$</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1:</td>
<td>$6.92$</td>
<td>$6.92$</td>
<td>$6.92$</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2:</td>
<td>$6.92$</td>
<td>$6.92$</td>
<td>$6.92$</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1:</td>
<td>$6.92$</td>
<td>$6.92$</td>
<td>$6.92$</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>2:</td>
<td>$6.92$</td>
<td>$6.92$</td>
<td>$6.92$</td>
</tr>
</tbody>
</table>

Table 5.3: Expected prices for network services in year 1

term network services can be performed as follows. Suppose a market participant is interested in obtaining a long term transmission contract between bus 2 and bus 3 for years 2, 3, and 4 for the transfer of 10MW. Then the market participant is informed by the ITC that the price for obtaining such a transmission contract is $5,869.47. If the purchase is made at this price, the market participant is entitled to the physical access of 10MW between bus 2 and bus 3 to the network without having to be concerned about the volatility in network prices nor about possible curtailment. At the beginning of the year 2, the SO informs the market participants of the prices for intermediate term transmission contracts as given in Table 5.4. The market participant holding the long term transmission contract for the 10MW transaction between bus 2 and bus 3 then exchanges the portion of the contract covering year 2 for intermediate term transmission contracts of 3.33MW on line 1, of 3.33MW on line 2 and of 6.67MW on line 3. The pricing of the short term network
<table>
<thead>
<tr>
<th>Year</th>
<th>Season</th>
<th>Day</th>
<th>Hour:</th>
<th>line (1)</th>
<th>line (2)</th>
<th>line (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>$6.92</td>
<td>$6.92</td>
<td>$50.35</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>$6.92</td>
<td>$6.92</td>
<td>$50.35</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>$6.92</td>
<td>$6.92</td>
<td>$7.32</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>$6.92</td>
<td>$6.92</td>
<td>$7.32</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>$6.92</td>
<td>$6.92</td>
<td>$6.92</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>$6.92</td>
<td>$6.92</td>
<td>$6.92</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>$6.92</td>
<td>$6.92</td>
<td>$6.92</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>$6.92</td>
<td>$6.92</td>
<td>$6.92</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>$6.92</td>
<td>$6.92</td>
<td>$6.92</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>$6.92</td>
<td>$6.92</td>
<td>$6.92</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>$6.92</td>
<td>$6.92</td>
<td>$6.92</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>$6.92</td>
<td>$6.92</td>
<td>$6.92</td>
</tr>
</tbody>
</table>

Table 5.4: Expected prices for network services in year 2

<table>
<thead>
<tr>
<th>Year</th>
<th>Season</th>
<th>Day</th>
<th>Hour:</th>
<th>line (1)</th>
<th>line (2)</th>
<th>line (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>$6.92</td>
<td>$6.92</td>
<td>$50.38</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>$6.92</td>
<td>$6.92</td>
<td>$50.38</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>$6.92</td>
<td>$6.92</td>
<td>$6.92</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>$6.92</td>
<td>$6.92</td>
<td>$6.92</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>$6.92</td>
<td>$6.92</td>
<td>$6.92</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>$6.92</td>
<td>$6.92</td>
<td>$6.92</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>$6.92</td>
<td>$6.92</td>
<td>$6.92</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>$6.92</td>
<td>$6.92</td>
<td>$6.92</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>$6.92</td>
<td>$6.92</td>
<td>$6.92</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>$6.92</td>
<td>$6.92</td>
<td>$6.92</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>$6.92</td>
<td>$6.92</td>
<td>$6.92</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>$6.92</td>
<td>$6.92</td>
<td>$6.92</td>
</tr>
</tbody>
</table>

Table 5.5: Expected prices for network services in year 3

<table>
<thead>
<tr>
<th>Year</th>
<th>Season</th>
<th>Day</th>
<th>Hour:</th>
<th>line (1)</th>
<th>line (2)</th>
<th>line (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>$6.92</td>
<td>$6.92</td>
<td>$6.92</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>$6.92</td>
<td>$6.92</td>
<td>$6.92</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>$6.92</td>
<td>$6.92</td>
<td>$6.92</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>$6.92</td>
<td>$6.92</td>
<td>$6.92</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>$6.92</td>
<td>$6.92</td>
<td>$6.92</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>$6.92</td>
<td>$6.92</td>
<td>$6.92</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>$6.92</td>
<td>$6.92</td>
<td>$6.92</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>$6.92</td>
<td>$6.92</td>
<td>$6.92</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>$6.92</td>
<td>$6.92</td>
<td>$6.92</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>$6.92</td>
<td>$6.92</td>
<td>$6.92</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>$6.92</td>
<td>$6.92</td>
<td>$6.92</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>$6.92</td>
<td>$6.92</td>
<td>$6.92</td>
</tr>
</tbody>
</table>

Table 5.6: Expected prices for network services in year 4

122
services is then obtained on the _ex post_ basis depending on the bids made by the market participants in the spot market, as described earlier. This process is repeated every year.
Chapter 6

Secondary market for transmission and the supporting infrastructures

In this chapter we discuss two infrastructures important for the proficient management of the network, namely the secondary market for transmission rights and the open access same time information systems (OASIS)\(^1\).

6.1 Role of the secondary markets for transmission rights

Denote the price of the forward contract entered into at present time \(t\) for delivery on a specified date, \(\tau_{dd}\), supplied by the generator at bus \(g_i\) as \(\rho_{g_i}(t, \tau_{dd})\). This forward price may be described as the spot market price for delivery of a commodity at a fixed time in the future [56], i.e.,\(^2\)

\[
\rho_{g_i}(t, \tau_{dd}) = \mathcal{E}_t \{\rho_{g_i}[\tau_{dd}]\}
\]  

(6.1)

where

\[
\mathcal{E}_t \{\cdot\}
\]

denotes the expected value of \(\cdot\) computed given the information available up to the present time \(t\). Thus, at the time of the agreement the contract has a value of zero and remains at zero as long as the expected value of the spot market price on the delivery date stays unchanged. As the delivery date approaches, however, new information regarding market conditions emerges and may influence the expected value of the spot market

---

\(^1\)The term may be misleading here, as the OASIS described here does not entirely conform with the Federal Energy Regulatory Commission’s (FERC) definition of OASIS. It is more similar to the Pennsylvania-New Jersey-Maryland (PJM) system’s usage of OASIS for internal data posting.

\(^2\)The spot market price needs to be discounted at the rate of risk-free investment in order to reflect the present value of the contract as described in Chapter 4. This fine detail is not included here for the sake of simplicity.
price on the delivery date to move up or down. Suppose at time $t_1$, where $t < t_1 \leq \tau_{dd}$, the expected value of the spot market price on the delivery date changes from $\rho_{g_i}(t, \tau_{dd})$ to $\rho_{g_i}(t, \tau_{dd}) + \Delta$, i.e.,

$$\rho_{g_i}(t_1, \tau_{dd}) = \mathbb{E}_{t_1} \{\rho_{g_i}[\tau_{dd}]\} = \mathbb{E}_{t} \{\rho_{g_i}[\tau_{dd}]\} + \Delta$$

(6.2)

while including the information available now up to time $t_1$. Then, the value of the contract also changes from zero to $\Delta$. Here a forward contract market place plays a significant role in providing the mechanism for supporting the market activities so that the change in value of the contract is readily conveyed to all of the market participants. For example, with the rise in forward prices the buyer, whose demand is elastic, may want to reduce consumption and realize the profit on the sale of the original contract. The efficiency of the market mechanism depends on how effortlessly such market activities can be carried out.

As a counterpart to the forward contract marketplace for energy, the secondary market for transmission provides the necessary mechanism for supporting market activities so that a change in the value of the transmission capacity is readily conveyed to all of the market participants. Figure 6-1 shows the information exchange among the market participants, the system operator (SO) and the independent transmission company (ITC) for the intermediate term transmission contracts involving the secondary market. As described

![Diagram](image)

Figure 6-1: The information exchange among the market participants, the SO and the ITC for intermediate term transmission contracts involving the secondary market

in Chapter 5 the market participants can purchase intermediate term transmission contracts on each line $l$ in the network from the SO for any desirable duration within the year. First, at the beginning of each year
the total capacity available on individual transmission lines within the network is determined by the ITC for the entire year \(n\). When determining the capacity the ITC relies on the expertise of the SO regarding the operation of the network including the power transfer distribution factors (PTDF). Then, the ITC issues the intermediate term transmission contracts to be offered to the market participants and used as forward contracts for the transmission portion of the electric services. The price for each of these contracts, \(\rho_t(t_s, t_e)\), is determined by the ITC, initially, based on the expected value of the transmission charge,

\[
\mathcal{E}_{t=(n-1)T+1} \{\rho_t[k]\}
\]

over the interval \([t_s, t_e]\) so that the expected value of the overall transmission revenue is maximized while respecting the network constraints.

Following the issuance of the intermediate term transmission contracts, the SO conducts the spot market for energy at each hour \(k\), and the actual transmission charge for each line in the network, \(\rho_t[k]\), is determined and made available to the market participants. Here the market participants may be the holders of the physical transmission rights, the holders of the financial transmission rights and/or the bidders in the spot market based on the distinction made in Chapter 5.

Suppose the market conditions change so that the expected value of the transmission charge computed at the beginning of the year needs to be adjusted in order to reflect accurately the current state of the electricity market. Then, the ITC announces the adjusted prices for the transmission contracts and applies the new prices to the contracts from that point on. The market participants, in turn, may utilize the secondary markets to trade any outstanding contracts issued prior to the price adjustment at the newly updated price.

Besides functioning as the marketplace for buying and selling old intermediate term transmission contracts, the secondary market for transmission rights supports the trading of various financial derivatives written on the transmission rights. For examples, a market may issue options for network capacity backed by the intermediate term transmission contracts. Options come in two primary forms, namely calls and puts. A call option gives the holder the right, not the obligation, to buy a specified amount of the underlying commodity (in this case the network capacity) at a specified price and for a specified period of time. A put option gives the holder the right, not the obligation, to sell a specified amount of the underlying commodity at a specified price and over a specified period of time. The trading of various financial derivatives in the secondary markets for transmission rights is done independently from the ITC while the transmission charge determined in the spot market is the major driving force behind.

Without the presence of the secondary markets for transmission rights, the ITC relies solely on the expertise it has gained by observing the transmission charges being imposed on the market participants in the spot markets when determining the prices to be charged for the transmission rights. This creates the open loop computation of the charge. However, with the presence of the secondary market for transmission rights, the ITC can observe a change in prices for equivalent rights in the secondary markets and take
this into consideration when updating the prices, i.e. in the feedback fashion. The actual mechanism for determining the price while taking the prevailing price for transmission rights in the secondary markets is beyond the scope of this thesis.

Figure 6-2 shows the financial exchanges between the market participants and the secondary markets for transmission rights.

![Figure 6-2: The financial exchange between the market participants and the secondary markets for transmission rights](image)

6.2 Closer look at the proposed transmission rights

With the introduction of the secondary markets for transmission rights we can compare the workings of the transmission rights, in the form of the intermediate term transmission contracts proposed in this thesis, with the transmission congestion contracts (TCC) and flowgate rights introduced in Chapter 4.

6.2.1 Point-to-point transmission rights

The transmission congestion contracts (TCC's) proposed in [31] are the representatives of the point-to-point transmission rights being widely considered at the time of this writing as a possible form of allocating network capacity over the longer term. In order to understand the differences between the TCC's and the intermediate term transmission contracts proposed in this thesis we need to look at not only the actual mechanism for implementing the contracts but also the underlying market structures.

The underlying market structure assumed for the TCC's is the cost-of-service regulation imposed on the transmission owners and the operational authority given to the non-profit organization, called an Independent System Operator (ISO) as described in Chapter 4. Under this market structure, the market participants are allowed to submit bids for purchasing the TCC’s once, at the beginning of the year (or of the season).³

³The market participants can determine the amount and the price of the TCC’s for bidding purposes, either purely based on the expected value of financial transmission rights of this sort or based on a financial contract for energy, the so-called contract-for-difference (CFD). The CFD is an arrangement made between two or more participants for mimicking bilateral transactions under the TCC scheme. The details on the CFD are referred to in Chapter 4 and in [31].
The ISO then determines the price and the amount of TCC’s to be made available, and allocates network capacity corresponding to the contracts based on the bids. Each of the TCC’s issued to the participants specifies at least the following three elements: the location of the source bus, the location of the sink bus and the amount of the energy involved in the transaction.

Once the allocation of the TCC’s is concluded, all market participants are required to submit bids to the spot market in the same way whether a participant owns a TCC or not. The ISO then clears the spot market by solving the optimal power flow (OPF) problem reviewed in Chapter 5, and completes the dispatch schedules without any regards to the allocation of the TCC’s. As a result of the market clearing process, the combined price of the energy and transmission portions of the electric services at each bus is determined by the shadow price associated with the OPF problem as written in Eq. (5.23). The price at each bus is often referred to as the nodal price [31] or, more recently, the locational based marginal price (LBMP) [47]. The revenue is collected and distributed by the ISO, as a product of the injection, into the bus and the corresponding nodal price. For example, suppose an amount of electric power, $Q_{d_i}[k]$, is taken from the network at a nodal price of $\rho_{d_i}[k]$. Then the load at that bus pays $\rho_{d_i}[k]$ for each unit of power (a total of $\rho_{d_i}[k] \cdot Q_{d_i}[k]$) to the ISO. Analogously, if the amount of electric power, $Q_{g_i}[k]$, is injected into the network at the nodal price of $\rho_{d_i}[k]$, the generator at that bus is paid $\rho_{d_i}[k]$ for each unit of power (a total of $\rho_{d_i}[k] \cdot Q_{d_i}[k]$) from the ISO. The transmission charge collected by the ISO here is often referred to as the congestion charge and is the difference between the amount received from the loads and the amount paid to the generators. Finally, the holders of the TCC’s are paid the differential between the nodal price at the location of the sink bus and the nodal price at the location of the source bus specified in the contract. Throughout the process the transmission owners are not involved because the revenue received by the transmission owners is a guaranteed return allowed by the regulator and is not related to the TCC’s (and consequently to the transmission (congestion) charge).

Based on the implementation of the TCC scheme described above, it is evident that the TCC’s are purely financial transmission rights since the holders of the contract are not given priority in using the network. Indeed, the market clearing process is completely independent from the allocation of TCC’s. The network related risks are two-fold, as described in Chapter 5, namely the price volatility in transmission capacity and the actual dispatch schedule, but the TCC’s cover only the former.

When the financial relationship created by the TCC’s is examined, it is recognized that there is an apparent disconnect between the reward/penalty mechanism and the entities assuming the financial risks. Because it is the ISO issuing the TCC’s, in order to offer hedging opportunities to market participants against volatility in the transmission capacity prices, it appears that the ISO takes on the financial risks. However, as described in Chapter 4, the ISO cannot assume any financial responsibilities. Thus, this imposes a critical constraint (perhaps auditable by the regulator) on issuing the TCC’s, namely a revenue neutrality coming from the simultaneous feasibility criterion. Revenue neutrality refers to the sufficient transmission charge being collected by the SO so that all of the payment to the TCC holders can be made from the transmission
charge. Thus, the simultaneous feasibility criterion significantly limits the ability of the SO in issuing the amount of the contracts. If the contracts together specify an injection of $Q_{g_i}[k]$ at bus $g_i$, then at each hour from the beginning of the year to the end, the injection at bus $g_i$ needs to be at least $Q_{g_i}[k]$. Similarly, if the contracts together specify a withdrawal of $Q_{d_j}[k]$ at bus $d_j$, then at each hour from the beginning of the year to the end, the withdrawal at bus $g_i$ needs to be at least $Q_{d_j}[k]$. If there is a difference in the transmission charge collected by the SO in the spot market and the TCC payment made to the holders of the contracts, then the difference is handed over to or made up from the market participants through the regulators [51].

In comparison, the underlying market structure assumed for the intermediate term transmission contract is the newly proposed price-cap regulation (PCR) discussed in Chapter 3, imposed on the TP and composed of the ITC and the SO. Under this market structure, the market participants may purchase intermediate term transmission contracts at any time directly from the SO at the price determined by the ITC in conjunction with the SO for any duration within the year (or the season). The revenue collected by the SO for offering the contracts is given directly to the ITC. As described in Chapter 5 the intermediate term transmission contracts specify the following four elements: the designated transmission line, the network capacity offered on the line, the direction of the flow, and the duration of the contract.

Once the market participant makes the purchase of the intermediate term transmission contracts, the holder has the choice of exercising the contract as physical transmission rights or as financial transmission rights. For example, suppose the market participant has an energy contract for transporting an amount of electricity $Q_{d_j,g_i}$ between the generation source at bus $g_i$ and the load sink at bus $d_j$ starting at hour $t_s$ and ending at hour $t_e$. Then, the participant may go to the SO at any time before $t_s$ and purchase an intermediate term transmission contract on each network line $l$ over the entire period $[t_s,t_e]$ for the amount and the direction of the capacity determined by the product of $Q_{d_j,g_i}$ and the power transfer distribution factors (PTDF). Then, each day before the SO conducts the spot market, the holder of the energy contract with the matching intermediate term transmission contracts can submit the balanced transaction and receive priority in utilizing the network, so that when the SO clears the spot market, the balanced transaction is scheduled first for dispatch. The holders of the intermediate term transmission contracts do not pay any additional transmission charges for using the network other than the payment made to the SO initially for purchasing the contracts. Otherwise, the market participants may purchase the intermediate term transmission contracts as purely financial transmission rights. In that case, the SO clears the spot market, and the holders of the contracts without matching energy contracts are paid the transmission charge entitled to them according to the contracts, as described in Chapter 5.

### 6.2.2 Link-based transmission rights

The flowgate rights proposed in [17] are representatives of the link-based transmission rights being widely considered in the industry at the time of this writing as another possible form of allocating network capacity
over the longer term. Although the flowgate may refer to any transmission line in the system, in general the term refers only to the links associated with likely network congestion, as it does here.

Similar to the TCC's case, the underlying market structure assumed for the flowgate rights is the rate-of-return regulation imposed on the transmission owners and the operational authority given to the ISO as described in Chapter 4. Under this market structure, the market participants are allowed to submit bids for purchasing the flowgate rights, once at the beginning of the year (or of the season). The ISO then determines the price and the amount of flowgates to be made available and allocates the network capacity corresponding to the flowgate rights based on the bids. Each of the flowgate rights issued to the participants specify at least the following two elements: the designated flowgate (i.e., the likely congested line) and the network capacity offered at the flowgate.

Once the allocation of the flowgate rights is concluded, two separate markets, namely the forward market and the spot market, are conducted sequentially. First, the participants in the forward market arrange for transactions and acquire, from the current holders, the necessary flowgate rights for implementing the arranged transactions. In this process, if a participant arranges a transaction that reduces the congestion at the flowgate, then the participant receives newly created flowgate rights of the amount by which the congestion is reduced. The process continues until all the transactions arranged are covered by the flowgate rights. The network capacity of unused flowgate rights are then returned to the ISO, which conducts the spot markets next.

The market participants who do not want to participate in the forward market can submit various bids to the spot markets. The ISO then clears the bids in the spot market by solving the OPF problem, subject to the network capacity limits re-defined by the effect of the unused flowgate rights. Again, as a result of the market clearing process, the combined price of energy and of the transmission portions of electric services at each bus is determined by the nodal prices, and the revenue is collected and distributed by the SO as the product of the injection into the bus and the corresponding nodal price. A part of the congestion charge collected by the ISO is used to compensate for the unused flowgate rights reverted to the ISO.

In the following section we apply the three methods described above through a numerical example.

6.2.3 Illustrative examples

Consider the 3-bus electric power network again as shown in Figure 6-3. The transfer limits on the transmission lines are 150MW for lines 1 and 2 and 80MW for line 3. In the network there are 12 generation units, each owned by different suppliers. The marginal operating costs of these units are given in the form

---

4The market participants can determine the amount and the price of the flowgate rights for bidding purposes based on the expected value of physical transmission rights of this sort with the matching forward (and/or bilateral) contracts. As described in Chapter 4 the explicit bilateral transactions are assumed to be allowed under the flowgate scheme similarly to the way they are allowed under the proposed scheme.
of linear functions with respect to their corresponding generations, i.e.,
\[
S_{g_i}(Q_{g_i}[k]) = 2a_{g_i}Q_{g_i}[k] + b_{g_i}
\] (6.3)

At bus 1 there are 4 thermal units capable of generating electric power at the cost determined by
\[
\begin{array}{cccc}
a_{g_1} & a_{g_2} & a_{g_3} & a_{g_4} \\
0.300 & 0.300 & 1.250 & 0.613
\end{array}
\]

and \(b_{g_i} = 0\) for \(i = 1, 2, 3, 4\). At bus 2 there are 7 hydro units with
\[
\begin{array}{cccccccc}
a_{g_5} & a_{g_6} & a_{g_7} & a_{g_8} & a_{g_9} & a_{g_{10}} & a_{g_{11}} \\
0.125 & 0.125 & 0.013 & 0.350 & 0.350 & 0.400 & 0.400
\end{array}
\]

and again \(b_{g_i} = 0\) for \(i = 5, 6, \ldots, 11\). At bus 3 there is 1 gas-turbine unit is operating at
Assuming perfect competition conditions, the supply function at each bus is computed by aggregating the marginal operating costs of the generators, i.e.,

at bus 1

\[ S_{\text{bus } 1}(Q_{\text{bus } 1}[k]) = 0.2198Q_{\text{bus } 1}[k] \]  \hspace{1cm} (6.4)

at bus 2

\[ S_{\text{bus } 2}(Q_{\text{bus } 2}[k]) = 0.0187Q_{\text{bus } 2}[k] \]  \hspace{1cm} (6.5)

and at bus 3

\[ S_{\text{bus } 3}(Q_{\text{bus } 3}[k]) = 10Q_{\text{bus } 3}[k] + 1 \]  \hspace{1cm} (6.6)

For simplicity, let the entire year be composed of four hours, i.e., \( k = 1, 2, 3, \) and 4.

Suppose some of the market participants enter into various forward contracts in order to meet their electricity needs at the beginning of the year assuming that the expected demand functions of the load at bus 2 are given by

\[ D_{d_2}[k] = -2.5Q_{d_2}[k] + 48.15 \]  \hspace{1cm} (6.7)

and those of the load at bus 3 are given as the following:

\[ D_{d_3}[k] = -5.0Q_{d_3}[k] + 817.10 \]  \hspace{1cm} (6.8)

For comparison purposes we consider the following arrangement of forward contracts. First, the marketers at bus 2 and bus 3 agree on a forward contract for the transfer of 101.25MW covering the entire year, i.e., \( k = 1, 2, 3, \) and 4. The marketers at bus 1 and bus 3 then arrange for the transfer of 56.00MW for hours 1, 2 and 3, but not 4; i.e. \( k = 1, 2, \) and 3. Finally, the marketers at bus 1 and bus 2 arrange a forward contract of 18.50MW, this time covering the hours 2, 3, and 4 only, i.e., \( k = 2, 3, \) and 4. Based on the supply functions given in Eqs. (6.4) through (6.6) and the demand functions projected as given in Eqs. (6.7) and (6.8), the loads at bus 2 and at bus 3 are expected to pay 1.90 ($/MW) and 30.85 ($/MW) respectively. Figure 6-4 shows the physical exchange among the participants based on the arrangement in the forward contracts. Transaction 1 refers to the forward contract for the transfer of 101.25MW from bus 2 to bus 3 for \( k = 1, 2, 3, \) and 4. Transaction 2 refers to the forward contract for the transfer of 56.00MW from bus 1 to bus 3 also for \( k = 1, 2, \) and 4, and Transaction 3 refers to the forward contract for the transfer of 18.50MW from bus 1 to bus 2 for \( k = 2, 3, \) and 4.

Following the arrangement in the forward contracts a spot market is conducted each hour in order to
Figure 6-4: The physical exchange among participants for each transaction through forward contracts
meet the residual demand. Suppose that following the market clearing process in the spot market the actual demand functions of the loads are revealed as the following:

for the load at bus 2

\[ D_{d_2}[k] = -2.5Q_{d_2}[k] + 48.15 \quad (6.9) \]

where \( k = 1, 2, 3, \) and 4, and for the load at bus 3

\[ D_{d_3}[1] = -5.0Q_{d_3}[1] + 817.10 \quad (6.10) \]
\[ D_{d_3}[2] = -5.0Q_{d_3}[2] + 842.10 \quad (6.11) \]
\[ D_{d_3}[1] = -5.0Q_{d_3}[4] + 817.10 \quad (6.12) \]
\[ D_{d_3}[3] = -5.0Q_{d_3}[3] + 792.10 \quad (6.13) \]

As is evident from Eqs. (6.9) through (6.13), the actual demand function for the load at bus 2 stays invariant from the expected throughout the year, but the actual demand function for the load at bus 3 is identical only at hours 1 and 3 and deviates from the expected in hours 2 and 4.

The actual physical exchange among the participants is determined as a result of the arrangement in the forward contracts and the spot market and is, therefore, highly market structure dependent. Here the presence of transmission rights plays an important role in deciding the final outcome. For simplicity without the loss of generality, assume that the market participants involved in the forward contracts purchase the appropriate transmission rights available in order to hedge against price volatility in the transmission charges whenever possible.

Under the TCC scheme, this assumption implies that the marketers involved in Transaction 1, Transaction 2 and Transaction 3 will purchase the TCC of 101.25MW between bus 2 and bus 3 for \( k = 1, 2, 3, 4, \) the TCC of 56.00MW between bus 1 and bus 3 for \( k = 1, 2, 3, \) and the TCC of 18.50MW between bus 1 and bus 2 for \( k = 2, 3, 4, \) respectively. It is assumed that the appropriate arrangements are made between the market participants in the form of the contract-for-difference (CFD) in order to mimic the bilateral transfers specified in Transactions 1, 2 and 3. Denote the TCC corresponding to Transaction \( i \) as \( TCC_i. \) According to the supply and the projected demand functions in Eqs. (6.4) through (6.7), the proper valuation of each TCC leads to 28.95 ($/MW) for \( TCC_1[k], \) 14.48 ($/MW) for \( TCC_2[k], \) and -14.48 ($/MW) for \( TCC_3[k]. \) The negative value of \( TCC_3 \) arises due to the direction of the flow being opposite to the transmission congestion on line 3 caused by the transaction. Then, the expected profit, \( \mathcal{E} \{ \pi_i = \rho_i \cdot Q_i \} \) at each bus can be readily computed. For the suppliers at bus 1

\[ \mathcal{E} \{ \pi_1[1] \} = 30.85 \cdot 56.00 + 16.37 \cdot 18.50 - \frac{1}{2} (16.37 \cdot 74.50) - 14.48 \cdot 56.00 \quad (6.14) \]
\[ \mathcal{E} \{ \pi_1[2] \} = 30.85 \cdot 56.00 + 1.90 \cdot 18.50 - \frac{1}{2} (16.37 \cdot 74.50) - 14.48 \cdot 56.00 - (-14.48) \cdot 18.50 \]  
(6.15)

\[ \mathcal{E} \{ \pi_1[3] \} = 30.85 \cdot 56.00 + 1.90 \cdot 18.50 - \frac{1}{2} (16.37 \cdot 74.50) - 14.48 \cdot 56.00 - (-14.48) \cdot 18.50 \]  
(6.16)

\[ \mathcal{E} \{ \pi_1[4] \} = 1.90 \cdot 18.50 + 16.37 \cdot 56.00 - \frac{1}{2} (16.37 \cdot 74.50) - (-14.48) \cdot 18.50 \]  
(6.17)

The first term in Eqs. (6.14) through (6.16) is the expected revenue collected from realizing Transaction 2 of 56MW while the first term in Eq. (6.17) and the second term in Eqs. (6.15) and (6.16) are the expected revenue collected from realizing Transaction 3 of 18.50MW, all committed ahead of time in the forward contracts. The second term in Eqs. (6.14) and (6.17) is the expected revenue from offering 18.50MW and from offering 56.00MW in the spot market, respectively. The third term in Eqs. (6.14) through (6.17) is the expected cost of generating 74.50MW. The fourth term in Eqs. (6.14) through (6.16) is the cost of acquiring the TCC to eliminating the risks associated with the volatility in the transmission charge on realizing Transaction 2. Similarly, the fifth term in Eqs. (6.15) and (6.16) and the fourth term in Eq. (6.17) is the cost of acquiring the TCC associated with Transaction 3. It is noted that the expected value of the profit at each period is the same – $609.78 – even though at \( k = 2 \) and \( 3 \) the entire profit is through the forward contracts with matching transmission rights while at \( k = 1 \) and \( 4 \) the profit includes the hedged forward contracts and spot market activities.

For the suppliers at bus 2

\[ \mathcal{E} \{ \pi_2[k] \} = 30.85 \cdot 101.25 - \frac{1}{2} (1.90 \cdot 101.25) - 28.95 \cdot 101.25 \]  
(6.18)

for \( k = 1, 2, 3, \) and \( 4 \). The first term in Eq. (6.18) is the expected revenue collected from realizing Transaction 1 of the 101.25MW committed to in the forward contracts. The second term in Eq. (6.18) is the expected cost of generating 101.25MW. Finally, the third term is the cost of acquiring the TCC in order to eliminate the risks associated with volatility in the transmission charge on realizing Transaction 1. The expected profit of the suppliers at bus 2, $96.07, only includes the completely hedged forward contracts.

Since the transmission rights under the TCC scheme are purely financial, the market participants, at each hour \( k \), are required to submit the either their supply or demand bids for energy to the ISO, which is responsible for conducting the spot market. Once the spot market is cleared, the collected bids are used for scheduling and for settlement. We assume that the supply bids submitted to the ISO are identical to Eqs. (6.4) through (6.6) for the entire year, i.e., \( k = 1, 2, 3, 4 \). On the other hand, the demand bids are submitted according to Eqs. (6.9) through (6.13).

For \( k = 1 \) the result of clearing the spot market yields the prices of 16.37 ($/MW), 1.90 ($/MW), and 30.85 ($/MW) at bus 1, bus 2 and bus 3, respectively. The suppliers are then scheduled for 74.50MW, 101.25MW and 0MW at bus 1, bus 2 and bus 3, respectively. At the end of hour 1 (and 4), the ISO collects from the loads at bus 2 $35.12 (= 1.90 ($/MW) \times 18.50 (MW)) and from the loads at bus 3 $4,851.49
\(= 30.85 \, \text{($/MW)} \times 157.25 \, \text{(MW)}) \) for a total of $4,886.60. The ISO then pays to the suppliers at bus 1 $1,219.93 \(= 16.37 \, \text{($/MW)} \times 74.5 \, \text{(MW)}) \) and to the suppliers at bus 2 $192.13 \(= 1.90 \, \text{($/MW)} \times 101.25 \, \text{(MW)}) \) for a total of $1,412.06. Of the difference between the revenue collected from the loads and the cost paid to the suppliers, $3474.54, the suppliers at bus 1 receive $810.73 \(= (30.85 - 16.37) \, \text{($/MW)} \times 56 \, \text{(MW)}) \) and the suppliers at bus 2 receive $2,931.65 \(= (30.85 - 1.90) \, \text{($/MW)} \times 101.25 \, \text{(MW)}) \) as specified by the corresponding TCC’s [31].

For \(k = 2\) the result of clearing the spot market yields the prices of 18.24 \(\text{($/MW)}, 1.82 \, \text{($/MW)}, \text{and} \) 34.66 \(\text{($/MW)} \) at bus 1, bus 2 and bus 3, respectively. The suppliers are then scheduled for 82.98MW, 97.04MW and 0MW at bus 1, bus 2 and bus 3, respectively. At the end of hour 2, the ISO collects from the loads at bus 2 $33.70 \(= 1.82 \, \text{($/MW)} \times 18.53 \, \text{(MW)}) \) and from the loads at bus 3 $5,596.83 \(= 34.66 \, \text{($/MW)} \times 161.49 \, \text{(MW)}) \) for a total of $5,630.54. The ISO then pays to the suppliers at bus 1 $1,513.37 \(= 18.24 \, \text{($/MW)} \times 82.98 \, \text{(MW)}) \) and to the suppliers at bus 2 $176.50 \(= 1.82 \, \text{($/MW)} \times 97.04 \, \text{(MW)}) \) for a total of $1,689.87. Of the difference between the revenue collected from the loads and the cost paid to the suppliers, $3,940.67, the suppliers at bus 1 receive $615.73 \(= (34.66 - 18.24) \, \text{($/MW)} \times 56 \, \text{(MW)} + (1.82 - 18.24) \, \text{($/MW)} \times 18.5 \, \text{(MW)}) \) and the suppliers at bus 2 receive $3,244.94 \(= (34.66 - 1.82) \, \text{($/MW)} \times 101.25 \, \text{(MW)}) \) as specified by the corresponding TCC’s. In addition, the suppliers at bus 1 receive $1.46 \(= (1.90 - 1.82) \times 18.5 \, \text{(MW)}) \) from the loads at bus 2 and pay $213.11 \(= (34.66 - 30.85) \times 56 \, \text{(MW)}) \) to the loads at bus 3, and the suppliers at bus 2 pay $385.31 \(= (34.66 - 30.85) \times 101.25 \, \text{(MW)}) \) to the loads at bus 3 as specified by the corresponding CFD’s.

For \(k = 3\) the result of clearing the spot market yields the same prices and same scheduling as given for \(k = 1\). Thus, at the end of hour 3, the ISO collects from the loads at bus 2 $35.12 \(= 1.90 \, \text{($/MW)} \times 18.50 \, \text{(MW)}) \) and from the loads at bus 3 $4,851.49 \(= 30.85 \, \text{($/MW)} \times 157.25 \, \text{(MW)}) \) for a total of $4,886.60. The ISO then pays to the suppliers at bus 1 $1,219.93 \(= 16.37 \, \text{($/MW)} \times 74.5 \, \text{(MW)}) \) and to the suppliers at bus 2 $192.13 \(= 1.90 \, \text{($/MW)} \times 101.25 \, \text{(MW)}) \) for a total of $1,412.06. Of the difference between the revenue collected from the loads and the cost paid to the suppliers, $3474.54, the suppliers at bus 1 receive $542.90 \(= (30.85 - 16.37) \, \text{($/MW)} \times 56 \, \text{(MW)} + (1.90 - 16.37) \, \text{($/MW)} \times 18.5 \, \text{(MW)}) \) and the suppliers at bus 2 receive $2,931.65 \(= (30.85 - 1.90) \, \text{($/MW)} \times 101.25 \, \text{(MW)}) \) as specified by the corresponding TCC’s.

For \(k = 4\) the result of clearing the spot market yields the prices of 14.51 \(\text{($/MW)}, 1.98 \, \text{($/MW)}, \text{and} \) 27.05 \(\text{($/MW)} \) at bus 1, bus 2 and bus 3, respectively. The suppliers are then scheduled for 66.02MW, 105.46MW and 0MW at bus 1, bus 2 and bus 3, respectively. At the end of hour 2, the ISO collects from the loads at bus 2 $36.50 \(= 1.98 \, \text{($/MW)} \times 18.47 \, \text{(MW)}) \) and from the loads at bus 3 $4,138.42 \(= 27.05 \, \text{($/MW)} \times 153.01 \, \text{(MW)}) \) for a total of $4,174.92. The ISO then pays to the suppliers at bus 1 $958.08 \(= 14.51 \, \text{($/MW)} \times 66.02 \, \text{(MW)}) \) and to the suppliers at bus 2 $208.43 \(= 1.98 \, \text{($/MW)} \times 105.46 \, \text{(MW)}) \) for a total of $1,166.51. Of the difference between the revenue collected from the loads and the cost paid to the suppliers, $3,008.41, the suppliers at bus 1 return $231.90 \(= (1.98 - 14.51) \, \text{($/MW)} \times 18.5 \, \text{(MW)}) \) and the suppliers at bus 2 receive $2,538.34 \(= (27.05 - 1.98) \, \text{($/MW)} \times 101.25 \, \text{(MW)}) \) as specified by the
corresponding TCC's. Similarly as before, the suppliers at bus 1 pay $1.46 (= (1.98 - 1.90) x 18.5 (MW)) to the loads at bus 2, and the suppliers at bus 2 receive $385.32 (= (30.85 - 27.05) x 101.25 (MW)) from the loads at bus 3 as specified by the corresponding CFD's.

Tables 6.1 and 6.2 summarize the financial as well as physical exchanges among the market participants and the ISO. As evident from the example, the physical exchange among the participants may be different

<table>
<thead>
<tr>
<th></th>
<th>$g_{bus\ 1}$</th>
<th>$g_{bus\ 2}$</th>
<th>$d_{bus\ 2}$</th>
<th>$d_{bus\ 2}$</th>
<th>ISO</th>
</tr>
</thead>
<tbody>
<tr>
<td>for TCC$_1$:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$g_{bus\ 2 \rightarrow ISO}$</td>
<td>$11,726.58$</td>
<td>-$11,726.58$</td>
<td></td>
<td></td>
<td>$11,726.58$</td>
</tr>
<tr>
<td>for TCC$_2$:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$g_{bus\ 1 \rightarrow ISO}$</td>
<td>$2,432.18$</td>
<td>-$2,432.18$</td>
<td></td>
<td></td>
<td>$2,432.18$</td>
</tr>
<tr>
<td>for TCC$_3$:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$g_{bus\ 1 \rightarrow ISO}$</td>
<td>-$803.49</td>
<td>$803.49$</td>
<td></td>
<td></td>
<td>-$803.49</td>
</tr>
<tr>
<td>$k = 1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$d_{bus\ 2 \rightarrow ISO}$</td>
<td>$35.12$</td>
<td></td>
<td>-$35.12</td>
<td></td>
<td>$35.12$</td>
</tr>
<tr>
<td>$d_{bus\ 3 \rightarrow ISO}$</td>
<td>$4,886.60$</td>
<td></td>
<td>-$4,886.60</td>
<td></td>
<td>$4,886.60$</td>
</tr>
<tr>
<td>ISO $\rightarrow g_{bus\ 1}$</td>
<td>$1,219.93$</td>
<td>$1,219.93$</td>
<td>-$4,886.60</td>
<td>$1,219.93$</td>
<td></td>
</tr>
<tr>
<td>ISO $\rightarrow g_{bus\ 2}$</td>
<td>$192.13$</td>
<td>$192.13$</td>
<td>-$1,219.93</td>
<td>$192.13$</td>
<td></td>
</tr>
<tr>
<td>for TCC$_1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ISO $\rightarrow g_{bus\ 2}$</td>
<td>$2,931.65$</td>
<td>$2,931.65$</td>
<td>-$192.13</td>
<td>$2,931.65$</td>
<td></td>
</tr>
<tr>
<td>for TCC$_2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ISO $\rightarrow g_{bus\ 1}$</td>
<td>$810.73$</td>
<td>$810.73$</td>
<td>-$2,931.65</td>
<td>$810.73$</td>
<td></td>
</tr>
<tr>
<td>$k = 2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$d_{bus\ 2 \rightarrow ISO}$</td>
<td>$33.70$</td>
<td></td>
<td>-$33.70</td>
<td></td>
<td>$33.70$</td>
</tr>
<tr>
<td>$d_{bus\ 3 \rightarrow ISO}$</td>
<td>$5,596.83$</td>
<td></td>
<td>-$5,596.83</td>
<td></td>
<td>$5,596.83$</td>
</tr>
<tr>
<td>ISO $\rightarrow g_{bus\ 1}$</td>
<td>$1,513.37$</td>
<td>$1,513.37$</td>
<td>-$5,596.83</td>
<td>-$1,513.37</td>
<td></td>
</tr>
<tr>
<td>ISO $\rightarrow g_{bus\ 2}$</td>
<td>$176.50$</td>
<td>$176.50$</td>
<td>-$1,513.37</td>
<td>$176.50$</td>
<td></td>
</tr>
<tr>
<td>for TCC$_1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ISO $\rightarrow g_{bus\ 2}$</td>
<td>$3,324.94$</td>
<td>$3,324.94$</td>
<td>-$176.50</td>
<td>-$3,324.94</td>
<td></td>
</tr>
<tr>
<td>for TCC$_2$ and TCC$_3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ISO $\rightarrow g_{bus\ 1}$</td>
<td>$615.73$</td>
<td>$615.73$</td>
<td></td>
<td>-$615.73</td>
<td></td>
</tr>
<tr>
<td>for CDF$_1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$g_{bus\ 2 \rightarrow d_{bus\ 3}}$</td>
<td>$385.31$</td>
<td>-$385.31</td>
<td></td>
<td>$385.31$</td>
<td></td>
</tr>
<tr>
<td>for CDF$_2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$g_{bus\ 1 \rightarrow d_{bus\ 3}}$</td>
<td>$213.11$</td>
<td>-$213.11</td>
<td></td>
<td>$213.11$</td>
<td></td>
</tr>
<tr>
<td>for CDF$_3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$d_{bus\ 2 \rightarrow g_{bus\ 1}}$</td>
<td>$1.46$</td>
<td>$1.46$</td>
<td>-$1.46</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6.1: Exchanges among the market participants and the ISO under the TCC scheme for $k = 1, 2$

from the arrangement in the forward contracts under the TCC scheme depending on the system operating conditions. For instance, at hours 2 and 3, all of the transactions are initially committed to in the forward contracts for the energy and for the transmission portion of the electric services. When the bidding into the spot market by the market participants happens to be identical to the expected at the time of entering into forward contracts, the combination of the CFD's and the TCC's results in the actual physical exchanges, as is the case with the example at hour 3. However, if the bidding into the spot market deviates from the
<table>
<thead>
<tr>
<th>$d_{bus\ 2}$</th>
<th>$d_{bus\ 3}$</th>
<th>$d_{bus\ 2}$</th>
<th>$d_{bus\ 2}$</th>
<th>ISO</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_{bus\ 2}$ → ISO</td>
<td>$35.12$</td>
<td>·</td>
<td>·</td>
<td>$-35.12$</td>
</tr>
<tr>
<td>$d_{bus\ 3}$ → ISO</td>
<td>$4,886.60$</td>
<td>·</td>
<td>·</td>
<td>$-4,886.60$</td>
</tr>
<tr>
<td>ISO → $g_{bus\ 1}$</td>
<td>$1,219.93$</td>
<td>$1,219.93$</td>
<td>·</td>
<td>·</td>
</tr>
<tr>
<td>ISO → $g_{bus\ 2}$</td>
<td>$192.13$</td>
<td>·</td>
<td>$192.13$</td>
<td>·</td>
</tr>
<tr>
<td>for $TCC_1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ISO → $g_{bus\ 2}$</td>
<td>$2,931.65$</td>
<td>·</td>
<td>$2,931.65$</td>
<td>·</td>
</tr>
<tr>
<td>for $TCC_2$ and $TCC_3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ISO → $g_{bus\ 1}$</td>
<td>$542.90$</td>
<td>$542.90$</td>
<td>·</td>
<td>·</td>
</tr>
<tr>
<td>$k = 4$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$d_{bus\ 2}$ → ISO</td>
<td>$36.50$</td>
<td>·</td>
<td>·</td>
<td>$-36.50$</td>
</tr>
<tr>
<td>$d_{bus\ 3}$ → ISO</td>
<td>$4,138.42$</td>
<td>·</td>
<td>·</td>
<td>$-4,138.42$</td>
</tr>
<tr>
<td>ISO → $g_{bus\ 1}$</td>
<td>$958.08$</td>
<td>$958.08$</td>
<td>·</td>
<td>·</td>
</tr>
<tr>
<td>ISO → $g_{bus\ 2}$</td>
<td>$208.43$</td>
<td>·</td>
<td>$208.43$</td>
<td>·</td>
</tr>
<tr>
<td>for $TCC_1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ISO → $g_{bus\ 2}$</td>
<td>$2,538.34$</td>
<td>·</td>
<td>$2,538.34$</td>
<td>·</td>
</tr>
<tr>
<td>for $TCC_2$ and $TCC_3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ISO → $g_{bus\ 1}$</td>
<td>$-231.90$</td>
<td>$-231.90$</td>
<td>·</td>
<td>·</td>
</tr>
<tr>
<td>for $CDF_1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$d_{bus\ 3}$ → $g_{bus\ 2}$</td>
<td>$385.32$</td>
<td>·</td>
<td>$385.32$</td>
<td>·</td>
</tr>
<tr>
<td>for $CDF_3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$g_{bus\ 1}$ → $d_{bus\ 3}$</td>
<td>$1.46$</td>
<td>$-1.46$</td>
<td>·</td>
<td>·</td>
</tr>
</tbody>
</table>

Table 6.2: Exchanges among the market participants and the ISO under the TCC scheme for $k = 3, 4$
expected, then the combination of the CFD’s and the TCC’s may not assure the necessary transmission capacity for carrying out the exchanges according to the contracts, as is the case at hour 2. As a matter of fact, at hour 2, the network capacity available to the suppliers at bus 2 is only 97.04MW instead of the 101.25MW offered by the ISO through the TCC’s. In this case, the ISO needs to buy back the TCC’s, due to insufficient capacity, from the market participants in order to satisfy the simultaneous feasibility criterion, thus satisfying the revenue neutrality of the ISO [31]. The difficulty is in allocating adequate funds to the ISO for carrying out such tasks as buying back the TCC’s because of the non-profit nature of the entity.

Since in the average sense, the amount of network capacity available throughout the entire year, i.e., \( k = 1, 2, 3, \) and 4, matches what is offered through the TCC’s, the funds used for buying back the inadequate capacity may be recovered by offering more TCC’s in surplus hours, in this case at hour 4. However, as in this example, the demand for the TCC’s is directly related not only to how many exchanges actually take place but also to how many of the exchanges are committed to in the forward contracts, which implies that in hour 4, there may not be enough participants involved in the forward contracts to create the necessary demand for recovering the funds through the TCC’s. A similar problem arises with regards to the TCC’s for Transaction 3 in the example. Because the contract starts at hour 2 instead of at hour 1, the participants may be at odds with respect to purchasing the necessary TCC’s through the auction process which is only held once, at hour 1, for the entire year. Therefore, in order for the TCC scheme to perform at the desired level, the role of marketers involved in trading for the sake of trading (without any actual hedging against physical risks) becomes critical.

Under the Flowgate scheme, the ISO offers 80MW of flowgate rights on the transmission line from bus 2 to bus 3 to be auctioned off at the beginning of the year. For Transaction 1, Transaction 2 and Transaction 3 the PTDF’s with respect to the line between bus 2 and bus 3 are 0.6667, 0.3333 and -0.3333, respectively. Based on the forward contracts for each transaction, ideally the suppliers at bus 1 acquire 18.67MW of flowgate rights for hours 1, 2 and 3 (i.e., \( k = 1, 2, \) and 3) and -6.1667MW of flowgate rights for hours 2, 3, and 4 (i.e., \( k = 2, 3 \) and 4) while the suppliers at bus 2 acquire 67.5MW of flowgate rights for the entire year (i.e., \( k = 1, 2, 3, \) and 4). However, the actual bidding process for acquiring the flowgate rights may vary significantly depending on the exact entitlement of the rights and on the market setup. We make the following two assumptions (as was also done in the previous section): that the flowgate rights are physical transmission rights with a use-it-or-lose-it rule and that once the rights are auctioned off by the ISO, the market participants may utilize the secondary markets for trading these rights at each hour before the spot market is conducted. It is emphasized here again that these assumptions are consistent with the mechanism described in [17]. In [17], the mechanism is described while treating the power exchange as the only forward market, which is conducted every day before the spot markets are conducted over the same period. In addition, the ISO offers the flowgate rights to be auctioned off only once at the beginning of the year. The mechanism is extended here to include the forward contracts over longer periods of various length.

Suppose that since Transaction 3 is not in effect until hour 2, only the suppliers involved in Transaction
1 and Transaction 2 participate in the initial auction process conducted by the ISO. Because Transaction 1 and Transaction 2 are arranged with the loads at bus 3 paying 30.85 ($/MW) while the marginal costs at bus 1 and at bus 2 are 16.37 ($/MW) for generating 56.00MW and 1.90 ($/MW) for generating 101.25MW, the suppliers are willing to pay up to 43.43 ($/MW) for the flowgate rights at each hour, i.e.,

\[
43.43 = \frac{1}{0.3333}(30.85 - 16.37)
\]

(6.19)

for the suppliers at bus 2

\[
43.43 = \frac{1}{0.6667}(30.85 - 1.90)
\]

(6.20)

where 0.3333 and 0.6667 are the associated PTDF for Transaction 2 (supplier 1) and for Transaction 1 (supplier 2). We make the restriction that at the initial auction by the ISO the flowgate rights are only offered at a lump sum covering the entire year.\(^5\) Then, given that any flowgate rights purchased by the suppliers at bus 1 are lost at \(k = 4\) because of the use-it-or-lose-it rule, the maximum price the suppliers at bus 1 are willing to pay decreases to 32.57 ($/MW) \((= (3/4) \times 43.43)\). This leads to the allocation of 67.5MW \((= 0.6667 \cdot 101.25)\) flowgate rights to suppliers at bus 2 for a price of 43.43 ($/MW) while the rest of the flowgate rights, 12.5MW \((= 80 - 0.6667 \cdot 101.25)\), goes to the suppliers at bus 1 for a price of 32.57 ($/MW).

Following the initial auction of the flowgate rights by the ISO, the expected profit, \(\mathcal{E}\{\pi_i = \rho_i \cdot Q_i\}\) at each bus can be readily computed. For the suppliers at bus 1

\[
\mathcal{E}\{\pi_1[k]\} = 30.85 \cdot 56.00 + 16.37 \cdot 18.50 - \frac{1}{2}(16.37 \cdot 74.50) - 32.57 \cdot 12.50 - 43.43(0.3333 \cdot 56.00 - 12.50)
\]

(6.21)

where \(k = 1, 2,\) and 3. The first term and the second term in Eq. (6.21) are the expected revenue collected from realizing Transaction 2 of 56MW and from offering 18.50MW in the spot market, respectively. The third term is the expected cost of generating 74.50MW. The fourth term is the cost of acquiring the flowgate rights in order to eliminate the network related risks (both in price volatility and in dispatch scheduling) for 37.50MW of the 56.00MW in Transaction 2. The last term is the expected cost to be paid for the transmission charge on the rest of the 56.00MW in Transaction 2 not covered by the flowgate rights.

\[
\mathcal{E}\{\pi_1[4]\} = 16.37 \cdot 74.50 - \frac{1}{2}(16.37 \cdot 74.50) - 32.57 \cdot 12.50
\]

(6.22)

The first term and the second term in Eq. (6.22) are the expected revenue from offering 74.50MW in the spot market and the expected cost of generating 74.50MW. The last term is the payment for the forfeited flowgate rights in the last hour. The average expected profit for the suppliers at bus 1 is again $609.78 (compared to

---

\(^5\)This restriction is not inherent to the flowgate scheme but is added here to address the ambiguity of auctioning the flowgate rights only once at the beginning of the year while covering the entire year. At the time of this writing, the author is not aware of any existing literature, including [17], which spells out the mechanism necessary for debarring such ambiguity.
under the TCC scheme) of which $306.94 is assured in hours 1, 2 and 3 in the forward contracts for energy and matching flowgate rights. This is equivalent to saying that 18.50MW of generation committed to in the forward contracts for energy are completely exposed to network related risks.

For the suppliers at bus 2

$$E\{\pi_2[k]\} = 30.85 \cdot 101.25 - \frac{1}{2}(1.90 \cdot 101.25) - 43.43 \cdot 67.50$$

(6.23)

where again $k = 1, 2, 3$ and 4. The first term in Eq. (6.23) is the expected revenue collected from realizing Transaction 1 of 101.25MW. The second term is the expected cost of generating 101.25MW. The last term is the cost of acquiring the flowgate rights to eliminate all network related risks for 101.25MW in Transaction 1. It is recognized that the profit of the suppliers at bus 2 includes no uncertainties and is assured to be $96.07 throughout the year due to the forward contracts for energy and matching flowgate rights.

Since the transmission rights under the flowgate scheme are physical, at each hour $k$ the functions of only the residual demand and supply are submitted to the ISO, which then conducts the spot market. Here we assume that any bilateral transactions arising from the forward contracts for energy without the matching flowgate rights are also bid into the spot market as a paired demand and supply [3].

For $k = 1$ the result of clearing the spot market yields the prices of 16.37 ($/MW), 1.90 ($/MW), and 30.85 ($/MW) at bus 1, bus 2 and bus 3, respectively. The suppliers at bus 1 are then scheduled for 56.00MW under the forward contracts of which 37.50MW is already assured before clearing the spot market through the flowgate rights. In addition 18.50MW is also scheduled for the suppliers at bus 1 through the spot market. The suppliers at bus 2 are scheduled for 101.25MW while no dispatch is scheduled at bus 3. At the end of hour 1, the ISO collects from the loads at bus 2 $35.12 (= 1.90 ($/MW) \times 18.50 (MW)) and pays the suppliers at bus 1 302.94 (= 16.37 ($/MW) \times 18.50 (MW)).

Before the spot market is conducted for the next hour, the secondary market for flowgate rights is conducted. With the presence of Transaction 3 between the suppliers at bus 1 and the loads at bus 2, a counterflow of 6.17MW may be created on the transmission line from bus 2 to bus 3. This means that up to an additional 6.17MW of flowgate rights may be created depending on the demand for such transmission rights. Since the suppliers involved in Transaction 2 can use the additional flowgate to ensure the transfer, the entire 6.17MW of flowgate rights are created and assigned to the suppliers at bus 1 involved in Transaction 2. Subsequently the suppliers at bus 1 involved in Transaction 3 are paid $267.83 (= 0.3333 \cdot 43.43 \cdot 18.5).

For $k = 2$ Transaction 1, Transaction 2, and Transaction 3 are all covered by appropriate flowgate rights, and thus reserve the priority in using the network. Therefore, the result of clearing the spot market only with the residual demand and supply functions yields the prices of 17.69 ($/MW), -5.56 ($/MW), and 40.93 ($/MW) at bus 1, bus 2 and bus 3, respectively. The negative price at bus 2 is to lead the loads at that bus to consume more so that more transfer capacity can be created between bus 1 and bus 3. The suppliers are scheduled for 61.97MW (= 56.00 + 5.97), 101.25MW and 0MW at bus 1, bus 2 and bus 3, respectively.
At the end of hour 2, the ISO pays the loads at bus 2 $16.60 (=-5.56 ($/MW) \times 2.98 (MW)) and collects from the loads at bus 3 $122.13 (= 40.93 ($/MW) \times 2.98 (MW)) for a total of $105.54. The ISO then pays the suppliers at bus 1 $105.54 (= 40.93 ($/MW) \times 5.97 (MW)).

Again before the spot market is conducted for the next hour, the secondary market for flowgate rights is conducted. The arrangement made at this time is the same as before with Transaction 1, Transaction 2, and Transaction 3 all having the appropriate flowgate rights according to the forward contracts.

For \( k = 3 \) the result of clearing the spot market for the residual demand and supply functions yields the same prices and same scheduling as given for \( k = 1 \). Since all the transactions are done through the forward contracts no financial exchanges take place through the ISO.

With Transaction 2 taken out from the secondary market for transmission rights, there is no longer the demand for additional flowgate rights created by Transaction 3. Thus, the only flowgate rights to be exercised belong to the suppliers involved in Transaction 1.

For \( k = 4 \) the result of clearing the spot market for residual demand and supply functions yields the prices of 14.51 ($/MW), 1.98 ($/MW), and 27.05 ($/MW) at bus 1, bus 2 and bus 3, respectively. The suppliers at bus 1 are then scheduled for 18.50MW under the forward contracts and for an additional 47.52MW in the spot market. The suppliers at bus 2 are scheduled for 101.25MW under the forward contracts as well as 4.24MW in the spot market. At the end of hour 4, the ISO collects from the loads at bus 3 $1,400.00 (= 27.05 ($/MW) \times 51.76 (MW)) and pays to the suppliers at bus 1 $689.62 (= 14.51 ($/MW) \times 47.52 (MW)) and to the suppliers at bus 2 $8.38 (= 1.98 ($/MW) \times 4.24 (MW)) for a total of $698.00. The difference between the revenue collected and the cost paid, $53.13, is the transmission revenue collected by the ISO for conducting the spot market.

Table 6.3 summarizes the financial as well as physical exchanges among the market participants and the ISO. As is evident from the example, physical exchange among the participants is assured to take place according to the forward contracts if and only if the contracts are covered by the appropriate flowgate rights. For instance, at hours 2 and 3, all of the transactions are initially committed to in the forward contracts on energy and on transmission portion of the electric services. Thus, the transactions take place exactly as committed to. As in the case at hour 2, when the spot market demand creates a shortage of transmission capacity, the ISO is forced to bring more capacity into the market by paying the loads at bus 3 to consume more.\(^6\) This is due to the binding physical commitment made through the flowgate rights. The difficulty is in extending this result to the longer term forward contracts rather than just to the power exchange type forward contract. As shown in hours 1 and 4, if the forward contracts are not synchronized and do not cover all of the transactions (in the expected sense), then there may be an inadequacy in the flowgate rights to be traded among the participants. This is due to the counterflows brought about by various contracts creating more than the mere absolute amount of rights initially offered by the ISO. There

\(^6\)The task is equivalent to utilizing adjustment bids by the suppliers if there are any.
<table>
<thead>
<tr>
<th>for flowgate rights:</th>
<th>$g_{bus 1}$</th>
<th>$g_{bus 2}$</th>
<th>$d_{bus 2}$</th>
<th>$d_{bus 2}$</th>
<th>ISO</th>
</tr>
</thead>
<tbody>
<tr>
<td>$g_{bus 2} \rightarrow ISO$</td>
<td>$11,726.58$</td>
<td>-$11,726.58$</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>$g_{bus 1} \rightarrow ISO$</td>
<td>$1,628.69$</td>
<td>.</td>
<td>-$1,628.69$</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>$k = 1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$d_{bus 2} \rightarrow ISO$</td>
<td>$35.12$</td>
<td>.</td>
<td>.</td>
<td>-$35.12$</td>
<td>.</td>
</tr>
<tr>
<td>ISO $\rightarrow g_{bus 1}$</td>
<td>$302.94$</td>
<td>$302.94$</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>for flowgate rights:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$g_{bus 1}(TR2) \rightarrow g_{bus 1}(TR3)$</td>
<td>$267.83$</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>$k = 2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ISO $\rightarrow d_{bus 2}$</td>
<td>$16.60$</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>$d_{bus 3} \rightarrow ISO$</td>
<td>$122.13$</td>
<td>.</td>
<td>.</td>
<td>-$122.13$</td>
<td>.</td>
</tr>
<tr>
<td>ISO $\rightarrow g_{bus 1}$</td>
<td>$105.54$</td>
<td>$105.54$</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>for flowgate rights:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$g_{bus 1}(TR2) \rightarrow g_{bus 1}(TR3)$</td>
<td>$267.83$</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>$k = 3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>for flowgate rights:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k = 4$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$d_{bus 3} \rightarrow ISO$</td>
<td>$1,400.00$</td>
<td>.</td>
<td>.</td>
<td>-$1,400.00$</td>
<td>.</td>
</tr>
<tr>
<td>ISO $\rightarrow g_{bus 1}$</td>
<td>$689.62$</td>
<td>$689.62$</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>ISO $\rightarrow g_{bus 2}$</td>
<td>$8.38$</td>
<td>.</td>
<td>$8.38$</td>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>

Table 6.3: Exchanges among the market participants and the ISO under the flowgate scheme for $k = 1, 2, 3, 4$
is another difficulty associated with the use-it-or-lose-it rule related to longer term applications. Unlike the power exchange cases, the participants may not be involved in forward contracts over the entire year for which the initial flowgate rights are auctioned. A clear compensation mechanism needs to be developed in such cases so that the charge mechanism may be viewed as fair by the spot market regardless of the several different ways of defining PTDF's.\(^7\)

Under the proposed scheme, the SO offers intermediate term transmission contracts on the transmission line between bus 2 and bus 3 at some price. Even though the intermediate term transmission contracts offered by the SO carry not only the congestion cost but also the *ex ante* flow tax as described in Chapter 5, for the purposes of comparison the flow tax portion of the network charge is assumed to be zero. Thus, the network charge on either the transmission line between bus 1 and bus 2 or the transmission line between bus 1 and bus 3 can be ignored since there is no congestion cost associated with the charge on these lines based on Eqs. (6.4) through (6.6) and Eqs. (6.7) and (6.8). Using these equations, the proper valuation of the intermediate term transmission contacts by the ITC leads to $43.43 in the direction from bus 2 to bus 3 and to -$43.43 in the opposite direction for the entire year, i.e. \(k = 1, 2, 3, \text{ and } 4\).\(^8\) We assume here that for Transaction 1, Transaction 2 and Transaction 3 the PTDF's with respect to the line between bus 2 and bus 3 are 0.6667, 0.3333 and -0.3333 respectively.

Based on the forward contracts for each transaction, the suppliers at bus 1 acquire 18.67MW of intermediate term transmission contracts from the SO for hours 1, 2 and 3 (i.e., \(k = 1, 2, \text{ and } 3\)) and -6.1667MW of flowgate rights for hours 2, 3, and 4 (i.e., \(k = 2, 3 \text{ and } 4\)) while the suppliers at bus 2 acquire 67.5MW of flowgate rights for the entire year (i.e., \(k = 1, 2, 3, \text{ and } 4\)). The price at which these contracts are offered is $43.43.

It is noted that by offering a total of 86.1667MW of contracts at hour 1, the SO (in turn the ITC) takes the network related risks, i.e., both price volatility and dispatch scheduling. What appears to be an over-commitment of the network capacity through the intermediate term transmission contracts is actually preferable from the perspective of the ITC. This is because the more capacity committed to through the intermediate term transmission contracts, while not violating the transfer limits based on expected market activities including the forward contracts for energy and the spot market, the more stable the revenue is for the ITC.\(^9\)

The expected profit, \(\mathcal{E}\{\pi_i = p_i \cdot Q_i\}\) at each bus can be readily computed based on the allocation of the

---

\(^7\)As discussed in Chapter 3, the PTDF's may be defined differently even under the same operating conditions, depending on the slack bus assignment.

\(^8\)Under the proposed scheme, the price of the intermediate term transmission contracts is determined by the ITC once the SO provides the (operational) transfer limits on each line and the PTDF's, even though the allocation of the contracts is conducted by the SO. The reason for this is because of the financial responsibility imposed on the SO as described in Chapter 5. The SO also assists the ITC with the projection of demand and supply functions.

\(^9\)A discussion of this preference is given in Chapter 5 in more detail.
transmission contracts. For the suppliers at bus 1

\[ E\{\pi_1[1]\} = 30.85 \cdot 56.00 + 16.37 \cdot 18.50 - \frac{1}{2} (16.37 \cdot 74.50) - 43.43 \cdot 18.67 \]  \hspace{1cm} (6.24)

\[ E\{\pi_1[2]\} = 30.85 \cdot 56.00 + 1.90 \cdot 18.50 - \frac{1}{2} (16.37 \cdot 74.50) - 43.43 \cdot 18.67 - (-43.43) \cdot 6.17 \]  \hspace{1cm} (6.25)

\[ E\{\pi_1[1]\} = 30.85 \cdot 56.00 + 1.90 \cdot 18.50 - \frac{1}{2} (16.37 \cdot 74.50) - 43.43 \cdot 18.67 - (-43.43) \cdot 6.17 \]  \hspace{1cm} (6.26)

\[ E\{\pi_1[1]\} = 1.90 \cdot 18.50 + 16.37 \cdot 56.00 - \frac{1}{2} (16.37 \cdot 74.50) - (-43.43) \cdot 6.17 \]  \hspace{1cm} (6.27)

The first term in Eqs. (6.24) through (6.26) is the expected revenue collected from realizing Transaction 2 of 56MW while the second term in Eqs. (6.25) and (6.26) and the first term in Eq. (6.27) is the expected revenue collected from realizing Transaction 3 of 18.5MW. The second term in Eqs. (6.24) and (6.27) is the expected revenue from the spot market. The third term in Eqs. (6.24) through (6.27) is the expected cost of generating 74.5MW. The last term in Eq. (6.24) and the fourth term in Eqs. (6.25) and (6.26) is the cost of acquiring the intermediate term transmission contracts in order to eliminate the network related risks for Transaction 2, while the last term in Eqs. (6.25) and (6.27) is the cost of the transmission contracts for Transaction 3.

For the suppliers at bus 2

\[ E\{\pi_2[k]\} = 30.85 \cdot 101.25 - \frac{1}{2} (1.90 \cdot 101.25) - 43.43 \cdot 67.50 \]  \hspace{1cm} (6.28)

where \( k = 1, 2, 3 \) and 4. The first term in Eq. (6.28) is the expected revenue collected from realizing Transaction 1 of 101.25MW. The second term is the expected cost of generating 101.25MW. The last term is the cost of acquiring the flowgate rights to eliminate all network related risks for 101.25MW in Transaction 1. It is recognized here again that the profit of the suppliers at bus 2 includes no uncertainties and is assured to be $96.07 throughout the year due to the forward contracts for energy and matching intermediate term transmission contracts.

For \( k = 1 \) the result of clearing the spot market yields the prices of 16.37 ($/MW), 1.90 ($/MW), and 30.85 ($/MW) at bus 1, bus 2 and bus 3, respectively. The suppliers at bus 1 are then scheduled for 74.50MW under the forward contracts and in the spot market. The amount committed to in the forward contract is actually scheduled before the spot market is conducted due to the matching intermediate term transmission contracts. The suppliers at bus 2 are scheduled for 101.25MW while no dispatch is scheduled at bus 3. At the end of hour 1, the SO collects from the loads at bus 2 $35.12 (= 1.90 ($/MW) × 18.50 (MW)) and pays the suppliers at bus 1 $302.94 (= 16.37 ($/MW) × 18.50 (MW)).

Before the spot market is conducted for the next hour, the secondary market for transmission rights is conducted in case any participants with intermediate term transmission contracts want to trade if the

145
operating conditions change. The price posted for intermediate term transmission contracts may vary as well at this time to reflect more accurate operating conditions according to the ITC's projection. We assume that the expected operating conditions for the market participants and for the ITC remain as given in Eqs. (6.4) through (6.6) and Eqs. (6.7) and (6.8).

For \( k = 2 \) Transaction 1, Transaction 2, and Transaction 3 are all covered by appropriate flowgate rights, and thus reserve priority in using the network. Therefore, the result of clearing the spot market only with the residual demand and supply functions yields the prices of 17.69 ($/MW), -5.56 ($/MW), and 40.93 ($/MW) at bus 1, bus 2 and bus 3, respectively. The negative price at bus 2 is to lead the loads at that bus to consume more so that more transfer capacity can be created between bus 1 and bus 3. The suppliers are scheduled for 61.97MW (= 56.00 + 5.97), 101.25MW and 0MW at bus 1, bus 2 and bus 3, respectively. At the end of hour 2, the ISO pays the loads at bus 2 $16.60 (= -5.56 ($/MW) \times 2.98 (MW)) and collects from the loads at bus 3 $122.13 (= 40.93 ($/MW) \times 2.98 (MW)) for a total of $105.54. The ISO then pays to the suppliers at bus 1 $105.54 (= 40.93 ($/MW) \times 5.97 (MW)). This result is identical to the result under the flowgate scheme.

Similarly, for \( k = 3 \), the result of clearing the spot market for the residual demand and supply functions yields the same prices and same scheduling as given for \( k = 1 \). Since all the transactions are done through the forward contracts no financial exchanges take place through the ISO. Plus, for \( k = 4 \) the result of clearing the spot market for residual demand and supply functions yields the prices of 14.51 ($/MW), 1.98 ($/MW), and 27.05 ($/MW) at bus 1, bus 2 and bus 3, respectively, again identical to the result under the flowgate scheme.

Table 6.4 summarizes the financial as well as physical exchanges among the market participants and the ITC through the SO under the proposed scheme. As evident from the example, physical exchange among the participants is assured to take place by the appropriate intermediate term transmission contracts. Unlike under the flowgate scheme the market participants may not rely on other participants to create transmission rights because the ITC takes the financial risks of over-committing through the transmission contracts. Thus, longer term hedging against network related risks is possible at any time. The change in market conditions (although not shown in the example) may result in price variation for the transmission rights, and thus more accurately reflects the system status. Finally, the intermediate term transmission contracts may be converted to financial rights entitling a transmission charge in case the participant decides not to commit physically as specified by the contracts.

### 6.2.4 Implementation of transmission rights and locational market power

It turns out that the ability to take on the financial risks given to the TP under the proposed scheme is also very important when implementing the longer term transmission rights in the presence of an apparent locational market power.
<table>
<thead>
<tr>
<th>Exchange</th>
<th>g_{bus1}</th>
<th>g_{bus2}</th>
<th>d_{bus1}</th>
<th>d_{bus2}</th>
<th>ISO</th>
</tr>
</thead>
<tbody>
<tr>
<td>\text{g_{bus2} \to ISO}</td>
<td>$11,726.58</td>
<td>-$11,726.58</td>
<td>-</td>
<td>-</td>
<td>$11,726.58</td>
</tr>
<tr>
<td>\text{g_{bus1} \to ISO}</td>
<td>$2,432.18</td>
<td>-</td>
<td>-$2,432.18</td>
<td>-</td>
<td>$2,432.18</td>
</tr>
<tr>
<td>\text{ISO \to g_{bus1}}</td>
<td>$803.49</td>
<td>$803.49</td>
<td>-</td>
<td>-</td>
<td>-803.49</td>
</tr>
<tr>
<td>\text{d_{bus2} \to ISO}</td>
<td>$35.12</td>
<td>-</td>
<td>-</td>
<td>-35.12</td>
<td>$35.12</td>
</tr>
<tr>
<td>\text{ISO \to g_{bus1}}</td>
<td>$302.94</td>
<td>$302.94</td>
<td>-</td>
<td>-</td>
<td>-302.94</td>
</tr>
<tr>
<td>\text{ISO \to d_{bus2}}</td>
<td>$16.60</td>
<td>-</td>
<td>-</td>
<td>$16.60</td>
<td>-16.60</td>
</tr>
<tr>
<td>\text{d_{bus3} \to ISO}</td>
<td>$122.13</td>
<td>-</td>
<td>-</td>
<td>-$122.13</td>
<td>$122.13</td>
</tr>
<tr>
<td>\text{ISO \to g_{bus1}}</td>
<td>$105.54</td>
<td>$105.54</td>
<td>-</td>
<td>-</td>
<td>-$105.54</td>
</tr>
<tr>
<td>\text{k = 3}</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>\text{d_{bus3} \to ISO}</td>
<td>$751.13</td>
<td>-</td>
<td>-</td>
<td>-$751.13</td>
<td>$751.13</td>
</tr>
<tr>
<td>\text{ISO \to g_{bus1}}</td>
<td>$689.62</td>
<td>$689.62</td>
<td>-</td>
<td>-</td>
<td>-$689.62</td>
</tr>
<tr>
<td>\text{ISO \to g_{bus2}}</td>
<td>$8.38</td>
<td>-</td>
<td>$8.38</td>
<td>-</td>
<td>-$8.38</td>
</tr>
</tbody>
</table>

Table 6.4: Exchanges among the market participants and the SO under the proposed scheme for \(k = 1, 2, 3, 4\)

It is well documented in [48] that the entities with apparent locational market power may purchase longer term transmission contracts (both physical and financial) and increase their profits further by withholding their capacity beyond the optimal level without the transmission contracts. This is because once the entities with the locational market power purchase longer term transmission contracts, the increase in price affects not only the profit from offering energy but also that from the financial entitlement specified in the transmission contracts. For example, suppose a supplier at bus \(g_i\) is situated in the import region and enjoys the locational market power, and the supplier’s profit increases above the allowed level under perfect competition with an increase in price until the decrease in energy sold has a greater effect on the profit. Let \(\hat{Q}_{gi}\) and \(\hat{p}_{gi}\) denote the optimal level of generation and the corresponding price for this supplier. By purchasing transmission contracts, the supplier’s profit may increase further for generation \(Q_{gi}\) less than \(\hat{Q}_{gi}\) at the price \(p_{gi}\) higher than \(\hat{p}_{gi}\), since the profit of the supplier now depends on not only the offered electricity but also on the financial compensation from the longer term transmission contracts. The compensation from the transmission contracts usually increases with an increase in price in the import region because the congestion charge increases. For the types of transmission contracts studied in [48], this increase in the profit of the supplier results in an increase in cost to the consumers while the income of TP (typically ISO) is unaffected.

Under the proposed PCR scheme, the incentive structure is created in such a way that the congestion charge is not desired to stay within the ceiling prices agreed to in advance between the regulator and the TP. This is because the profit of the TP decreases if the congestion charge exceeds the ceiling prices. Suppose the supplier with locational market power and intermediate term transmission contracts withholds its capacity. Then, the profit of the supplier increases in the same manner described in [48]. However, this
increase in profit now results in an increase in cost to the consumers as well as in a decrease in income of the TP. It is conjectured here without proof that in this case the TP reacts to the decrease in income by increasing the price of the intermediate term transmission contracts in the short term and by increasing the transmission capacity in the long term. The increase in price results in reduced incentives for the supplier with locational market power to purchase intermediate term transmission contracts, and thus in reduced incentives for the supplier to withhold its capacity in the short term. The increase in transmission capacity results in the reduced locational market power of the supplier and again in reduced incentives to withhold its capacity. Therefore, the proposed market structure and PCR schemes in this thesis are speculated to be highly desirable, especially in the presence of locational market power.

6.3 Role of open access same time information systems (OASIS)

Fluent information exchange among market participants is a critical component of any well functioning market. The importance of fluent information exchange is even greater for the electric power markets because the reliability of the overall system heavily depends on the efficient operation and planning of the network, which is induced by the market activities. This need is well recognized by the interested parties, including the regulators, and as a consequence, Order 889 issued by FERC has provided the basis for establishing an information infrastructure called open access same time information systems (OASIS) in the US. Currently OASIS supports the operation of regional electric power markets by posting many useful data on the worldwide web so that the market participants can make informed decisions. Here we examine the minimum data\textsuperscript{10} and an efficient way of communicating them through OASIS under the proposed mechanisms for transmission provision.

6.3.1 Minimum data requirement under the proposed market mechanisms for transmission provision

From the perspective of the TP, the management of OASIS starts with a posting of the data pertaining to the steady state network status. These include the operating capacity limits, $F_l^{\text{max}}$ and the PTDF's, $H_{l(i)}$ for each transmission line $l$ within the network. This information represents steady state values due to the requirement that the TP is responsible for operating the regional network while the maximum flow limits and the PTDF's seen by the network users stay invariant throughout the year (or the season) as described in Chapter 5.

Once $F_l^{\text{max}}$ and $H_{l(i)}$ for each line $l$ are posted for the entire system, the SO publishes the prices for

\textsuperscript{10}A tremendous amount of data is required to accurately describe the status of an electric power network on a minute by minute basis. However, in this thesis we only focus on the data elements pertaining to the markets for electric services conducted in time scales longer than or equal to an hour. The markets other than energy and transmission markets, such as the ancillary services markets, are also beyond the scope of this thesis.
acquiring the intermediate term transmission contracts, $\rho_l[k]$ throughout the year, i.e. $k = 1, 2, \cdots, 8760$. The computation for pricing the contracts is based on the methods described in Chapter 5.

Suppose there are a supplier at bus $g_l$ and a load at bus $d_j$ negotiating a long term bilateral contract of $Q_{g_l \rightarrow d_j}$ for the duration of $T_s$ to $T_e$. Then, the expected value of the transmission charge to be levied on the transaction is given by

$$\mathcal{E}\{TC_{g_l \rightarrow d_j}(Q_{g_l \rightarrow d_j}, T_s, T_e)\} = \sum_{k=T_s}^{T_e} \sum_{l} \rho_l[k](H_{g_l} - H_{ld_j})Q_{g_l \rightarrow d_j}$$

where $\rho_l[k]$ is the price published for the intermediate term transmission contract on line $l$ for hour $k$. The supplier at bus $g_l$ and the load at bus $d_j$ may purchase the intermediate term transmission contract from the SO according to the published price.

At the end of each hour $k$, the SO replaces the published price for the intermediate transmission contract, $\rho_l[k]$ for line $l$, with the actual value of the transmission capacity, $\hat{\rho}_l[k]$, determined as a result of conducting the spot market. Any market participants holding intermediate term transmission contracts for hour $k$ can use the contracts to pay their network charge incurred by implementing bilateral transactions or collect the difference between $\hat{\rho}_l[k]$ and $\rho_l[k]$. When the market participant requests the transmission capacity to implement bilateral contracts with the transmission contract at the beginning of the hour and returns the contract to SO in lieu of payment at the end of the hour, the contract is a physical right. Otherwise, the intermediate term transmission contract is a financial right and represents a financial obligation between the TP and the holder. In addition to replacing the contract price with the spot market price, based on the market activities observed, the SO posts the newly updated prices for intermediate term transmission contracts, $\rho_l[h]$ for $h = k + 1, \cdots, 8760$ reflecting the changes in network status. Table 5.4 shows an example of data posted by the SO at the beginning of each year and repeated here as Table 6.5 for completeness. Suppose the actual load in the system is 0.5% less than that initially projected. Then, after the short term

<table>
<thead>
<tr>
<th>Year</th>
<th>Season</th>
<th>Day</th>
<th>Hour:</th>
<th>line (1)</th>
<th>line (2)</th>
<th>line (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1:</td>
<td>$6.92$</td>
<td>$6.92$</td>
<td>$50.35$</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2:</td>
<td>$6.92$</td>
<td>$6.92$</td>
<td>$50.35$</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1:</td>
<td>$6.92$</td>
<td>$6.92$</td>
<td>$7.32$</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2:</td>
<td>$6.92$</td>
<td>$6.92$</td>
<td>$7.32$</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>3</td>
<td>1:</td>
<td>$6.92$</td>
<td>$6.92$</td>
<td>$6.92$</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>3</td>
<td>2:</td>
<td>$6.92$</td>
<td>$6.92$</td>
<td>$6.92$</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1:</td>
<td>$6.92$</td>
<td>$6.92$</td>
<td>$6.92$</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2:</td>
<td>$6.92$</td>
<td>$6.92$</td>
<td>$6.92$</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1:</td>
<td>$6.92$</td>
<td>$6.92$</td>
<td>$6.92$</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2:</td>
<td>$6.92$</td>
<td>$6.92$</td>
<td>$6.92$</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>3</td>
<td>1:</td>
<td>$6.92$</td>
<td>$6.92$</td>
<td>$6.92$</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2:</td>
<td>$6.92$</td>
<td>$6.92$</td>
<td>$6.92$</td>
</tr>
</tbody>
</table>

Table 6.5: Prices of intermediate term transmission contracts
spot market is conducted at hour 1 in year 2, the new prices posted by SO might look like Table 6.6. In

<table>
<thead>
<tr>
<th>Year</th>
<th>Season</th>
<th>Day</th>
<th>Hour:</th>
<th>line (1)</th>
<th>line (2)</th>
<th>line (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1:</td>
<td>($6.92)</td>
<td>($6.92)</td>
<td>($49.27)</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2:</td>
<td>$6.92</td>
<td>$6.92</td>
<td>$49.27</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1:</td>
<td>$6.92</td>
<td>$6.92</td>
<td>$6.92</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2:</td>
<td>$6.92</td>
<td>$6.92</td>
<td>$6.92</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>3</td>
<td>1:</td>
<td>$6.92</td>
<td>$6.92</td>
<td>$6.92</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>3</td>
<td>2:</td>
<td>$6.92</td>
<td>$6.92</td>
<td>$6.92</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1:</td>
<td>$6.92</td>
<td>$6.92</td>
<td>$6.92</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2:</td>
<td>$6.92</td>
<td>$6.92</td>
<td>$6.92</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1:</td>
<td>$6.92</td>
<td>$6.92</td>
<td>$6.92</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2:</td>
<td>$6.92</td>
<td>$6.92</td>
<td>$6.92</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>3</td>
<td>1:</td>
<td>$6.92</td>
<td>$6.92</td>
<td>$6.92</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2:</td>
<td>$6.92</td>
<td>$6.92</td>
<td>$6.92</td>
</tr>
</tbody>
</table>

Table 6.6: Updated prices of intermediate term transmission contracts

Table 6.6 the numbers in parentheses denote the prices for network services, determined after conducting the spot market, while the numbers in bold face denote the revised prices for the intermediate term transmission contracts after the 0.5% decrease in load is incorporated in the projection following hour 1.

Given that there are over 2000 lines present in modeling any real life network, the size of a table such as Tables 6.5 and 6.6 may be quite large. In analyzing the value of a particular generation supply with respect to the transmission charge, thus may be computationally quite intensive. However, as can be seen in Tables 6.5 and 6.6, the price of intermediate term contracts is different from that of the ex ante flow tax only when the operating limits on that particular line are binding. Therefore, it may be desirable for SO to publish only the relevant transmission data and provide a simplified method for market participants to assess their usage of the network and relevant transmission charges. One such method is approximate charging mechanisms accomplished by the aggregation of buses into clusters depending on their impact on the network operation [74]. We describe the extension of the bus aggregation into clusters over a longer period of time.

### 6.3.2 Aggregation method for minimum date requirement under the proposed market mechanisms

Since the main focus of the transmission charging mechanisms in the aggregation method is related to the congestion charge only, we assume that the ex ante flow tax element is equal to zero for simplicity without the loss of generality.

**Optimal power flow problem revisited**

With the ex ante flow tax set to be equal to zero, the transmission charge is directly related to the shadow cost of operating constraints in solving the OPF problem reviewed in Chapter 5. We repeat the formulation
here for completeness.

\[ Q_{G_i}[k] = \arg \min_{Q_{G_i}[k]} \sum_{g_i} (a_{g_i} Q_{g_i}^2[k] + b_{g_i} Q_{g_i}[k]) \]  

(6.30)

\[ \sum_{g_i} Q_{g_i}[k] = \sum_{d_j} Q_{d_j}[k] : \lambda[k] \]  

(6.31)

\[ Q_{g_i}^{\min}[k] \leq Q_{g_i}[k] \leq Q_{g_i}^{\max}[k] : \eta_{g_i}[k] \]  

(6.32)

\[ \sum_{g_i} H_{l_i} Q_{g_i}[k] - \sum_{d_j} H_{l_d} Q_{d_j}[k] \leq F_{l_i}^{\max}[n] : \mu_l[k] \]  

(6.33)

where \( Q_{d_j}[k] \) is assumed to be inelastic and thus given ahead of time. The optimization problem in Eq. (6.30) can be thought of as the spot market clearing process, and the shadow cost denoted as \( \mu_l[k] \) can be thought of as the transmission charge on the transmission line \( l \). We again use the DC load flow simplifying assumption as described in Appendix B.

As done in [62], we can solve the optimization problem in (6.30) by constructing a Lagrangian function of the form [13]

\[ L = \sum_{g_i} a_{g_i} Q_{g_i}^2[k] + b_{g_i} Q_{g_i}[k] + \lambda[k] \left( \sum_{d_j} Q_{d_j}[k] - \sum_{g_i} Q_{g_i}[k] \right) 
+ \sum_l \mu_l[k] \left( \sum_{g_i} H_{l_i} Q_{g_i}[k] + \sum_{d_j} H_{l_d} Q_{d_j}[k] - F_{l_i}^{\max} \right) 
+ \sum_{g_i} \eta_{g_i}[k] \left( Q_{g_i}[k] - Q_{g_i}^{\max} \right) \]  

(6.34)

where \( \mu_l[k] \neq 0 \) if and only if \( F_l[k] = F_l^{\max} \). Taking the first derivative of \( L \) with respect to \( Q_{g_i}[k] \) and setting it equal to zero yields

\[ 2a_{g_i} Q_{g_i}[k] + b_{g_i} + \eta_{g_i}[k] = \lambda[k] + \sum_l \mu_l[k] H_{l_i} \]  

(6.35)

Then, we can define the so-called nodal price, \( \rho_{g_i}[k] \), and compute \( Q_{g_i}[k] \) in Eq. (6.35) as

\[ \rho_{g_i}[k] = \lambda[k] + \sum_l \mu_l[k] H_{l_i} \]  

(6.36)

and

\[ Q_{g_i} = \begin{cases} 
Q_{g_i}^{\max} & \rho_{g_i}[k] \geq p_{g_i}^{\max} \\
\frac{\rho_{g_i}[k]}{2a_{g_i}} & 0 \leq \rho_{g_i}[k] \leq p_{g_i}^{\max} \\
0 & \text{otherwise}
\end{cases} \]  

(6.37)

where \( p_{g_i}^{\max} = 2a_{g_i} Q_{g_i}^{\max} \). Here we assume \( b_{g_i}[k] = 0 \) for simplicity. This completes the description of the bus-based congestion management system (CMS) commonly known as the nodal pricing method. In the nodal pricing method, each bus in the system network receives a particular nodal price based on the supplier’s willingness to produce so that the quantity produced is limited by this price.
For example, suppose we denote the supply bid by generation unit $G_i$ as $B_{G_i}$ at some hour $k$. Then, the solutions given in Eqs. (6.36) and (6.37) have the graphical interpretation in Figure 6-5. For the generator $G_i$, shown in the figure, the dispatched generation amount is equal to $Q_{G_i}^{\text{max}}$ since $p_{G_i} \geq p_{G_i}^{\text{max}}$.

**Congestion clustering as an aggregation method**

In contrast to the bus-based CMS, there is the cluster-based CMS in which the network is pre-divided into several clusters and the buses belonging to the same cluster are assigned a single cluster-wide price following the market clearing process.

The cluster-based CMS is very useful as an aggregation method for posting prices for network services since, once the clusters are defined, the only transmission lines to consider are inter-cluster lines. Any transmission lines inside the cluster boundaries can be ignored for all practical purposes since we assume no congestion is likely to appear on these lines, and thus the prices to be charged on these lines are equal to the ex ante flow tax. In addition, the cluster-based CMS may be more desirable in many markets than CMS since it implements bilateral transactions much more easily by providing transparent information on the status of transmission (system) congestion. The uniform prices within clusters are also to the advantage of the cluster-based CMS since they considerably simplify the computation of the financial risks in bilateral transactions arising from the limitation on generation in the presence of transmission congestion. However, there are some disadvantages to the cluster-based CMS. The disadvantages are related to the unfavorable increase in the cost of dispatched generators in the short term. The short term dispatch is suboptimal due to two factors: (1) the cost from the cluster-wide prices in inter-cluster pricing and (2) the cost from the uplift charges in intra-cluster pricing.

The congestion cluster pricing method is quite suitable as a viable CMS since it reduces the effect of disadvantages while preserving the effect of the advantages of the implementation of the cluster-based CMS.
The key to the method is the novel approach proposed in [73] used to compute the sensitivity measures of injection.

The implementation of the congestion cluster pricing method consists of two steps: (1) the aggregation of individual nodes into clusters and (2) the computation of cluster-wide prices. The resulting clusters and prices determine an operating condition that is at an optimum with respect to some pre-defined objective function while keeping the power transfer across cluster interfaces within the acceptable limits. More details of the congestion cluster pricing method can be found in [71] and [72].

The number of clusters and the duration of the fixed cluster boundaries are required to be specified ahead of time with respect to some heuristic measure of long term efficiency, and according to the needs of the market and its participants. Typically, the desired number of clusters is limited to at most 30, and the duration is limited to at least a season. The TP is then assigned the task of cluster design, i.e. defining the cluster boundaries in order to aggregate the individual nodes into clusters.

The minimum desired criteria for the congestion cluster pricing method can be summarized as

1. the transactions between any buses within the same cluster have little impact on the flows through the congested transmission lines

2. the energy cost, computed after relieving inter-cluster congestion, is relatively small

3. the additional energy cost necessary for relieving intra-cluster congestion is relatively small

The first criterion is related to accommodating the bilateral transactions by providing transparent information on the status of transmission (system) congestion. The cluster affiliation of each node affords enough transparent information to market participants on how to structure bilateral transactions so that the congestion charge remains within the acceptable bounds. The second and third criteria are related to reducing the short-term cost of dispatched generators arising from the costs in inter-cluster pricing and the uplift charge in intra-cluster pricing.

**Formulation of cluster design problem**

First, we present the formulation of the aggregation step in the implementation of the congestion cluster pricing method as a stochastic optimization problem given by

\[
\theta^* = \arg \min_{\theta \in \Theta} \int_{0}^{T} J(\theta, t) dt \equiv E \left[ \int_{0}^{T} L(\theta, \xi(t), t) dt \right] \tag{6.38}
\]

where \( \Theta \) represents the search space from which various cluster design alternatives can be selected for aggregating individual nodes into clusters. In Eq. (6.38) the performance measure denoted by \( J(\theta, t) \) is the expected value of the sample performance, and \( L(\theta, \xi(t), t) \) is a function of the cluster design, \( \theta \), and the uncertainty, \( \xi(t) \), in the system. Given that it is desirable to keep the same cluster boundaries for a certain
period of time, i.e. a season, \( T \) represents the duration of the fixed boundaries. A slightly modified form of Eq. (6.38) is a little more useful, since the minimum time scale at which the operation of the system takes place is typically one hour. Thus, the optimization problem of the interest is given by

\[
\theta^* = \arg \min_{\theta \in \Theta} \sum_{k=0}^{T} J(\theta, k) \equiv \mathbb{E} \left[ \sum_{k=0}^{T} L(\theta, \xi[k], k) \right]
\]  

(6.39)

**Modeling uncertainties in a given network** In Eq. (6.39) the uncertainty, \( \xi(t) \) in the system can actually be broken into three parts: the uncertainty in load \( \xi_{Q_d_j}(t) \), the uncertainty in generation bid, \( \xi_{S_{g_i}}(t) \), and the uncertainty in the status of the equipment, i.e. the generator or transmission line, \( \xi_{Q_{g_i}} \) or \( \xi_{F_i} \), respectively. As described in Chapter 2 the uncertainty in load is related to the inability of the SO to forecast the demand perfectly. This is due to the fact that the many variables affecting the demand, such as the ambient temperature, etc., are rather unpredictable. The uncertainty in generation bid is related to the inability of the SO to predict the bidding behavior of the individual supplier within the system. The reason for this is because the variables influencing the bidding behavior, such as fuel cost, unit commitment strategy, etc., are for the most part unknown except to the supplier. The uncertainty in the equipment is related to the inability of the SO to determine the status of either the generators or transmission lines in advance since indeterminate variables, such as an overgrown tree near the transmission lines, are the main cause of equipment outages.

There are many ways of accounting for the discussed uncertainties. For our purposes, we need a time series representation of each uncertainty.

1. **Modeling load uncertainty**

The time series model of load can be represented in a general version of a discrete time random walk by

\[
Q_{d_j}[k + 1] = f_{d_j}(Q_{d_j}[k + 1], Q_{d_j}[k]) + e_{Q_{d_j}}[k + 1]
\]  

(6.40)

where \( e_{Q_{d_j}}[k] \) is normally distributed with zero mean and variance \( \sigma^2_{Q_{d_j}} \), and is independent of \( e_{Q_{d_j}}[l] \) for any \( k \neq l \). The expected value of the demand at \( k \) is denoted by \( \bar{Q}_{d_j}[k] \) while the projected demand at \( k \) computed through Eq. (6.40) is indicated by \( Q_{d_j}[k] \). Typically \( f(\cdot) \) is assumed to take on either linear or exponential form, and the parameter estimation is performed to complete this regression model [1].

2. **Modeling generation bid uncertainty**

The time series model of generation bid can be represented in a similar way by

\[
S_{g_i}[k + 1] = 2a_{g_i}[k + 1]Q_{g_i} + b_{g_i}[k + 1]
\]  

(6.41)

where typically the slope, \( a_{g_i} \), is assumed to be fixed, i.e. \( a_{g_i}[k] = a_{g_i} \), and the intercept follows another
linear regression model given by

\[ b_{gl}[k + 1] = b_{gl}[k] + e_{S_{gl}}[k] \]  

(6.42)

where \( e_{S_{gl}}[k] \) is again normally distributed with zero mean and variance \( \sigma^2_{S_{gl}} \) [1].

3. Modeling equipment status uncertainty

The time series model of equipment status can be represented by using a conventional Markovian chain consisting of two states as shown in Figure 6-6. The parameters for the failure rate and the repair rates are denoted by \( \lambda \) and \( \mu \) respectively for each component in the figure. Using these parameters, the steady state probability for states 0 (unit down state), \( \pi_0 \), and 1 (unit up state), \( \pi_1 \), are given by

\[ \pi_0 = \frac{\lambda}{\lambda + \mu} \]  

(6.43)

\[ \pi_1 = \frac{\mu}{\lambda + \mu} \]  

(6.44)

respectively [61].

**Function for sample performance** In Eq. (6.39) the function describing sample performance, \( L(\theta, \xi[k], k) \), determines how the superior designs are compared to the inferior ones; i.e. if

\[ \mathcal{E} [L(\theta_i, \xi[k], k)] < \mathcal{E} [L(\theta_j, \xi[k], k)], \]

then \( \theta_i \) is a better cluster design than \( \theta_j \). Thus, the function is directly related to the various criteria for a good congestion cluster pricing method. The minimum desired criteria for the method are already discussed earlier and are listed here again for completeness:

1. the transactions between any buses within the same cluster have little impact on the flows through the congested transmission lines, \( L_{D(i,j)}(\theta, \xi[k], k) \)

2. the energy cost, computed after relieving inter-cluster congestion, is small, \( L_{Q_{gl}}(\theta, \xi[k], k) \)

3. the additional energy cost necessary for relieving intra-cluster congestion is small, \( L_{\Delta Q_{gl}}(\theta, \xi[k], k) \)
Limiting the sample performance to reflect only the measures of the above three criteria, we consider the overall sample performance function to be given as

\[ L(\cdot) = \alpha_{D(i,j)} L_{D(i,j)}(\cdot) + \alpha_{Q_{se}} L_{Q_{se}}(\cdot) + \alpha_{\Delta Q_{se}} L_{\Delta Q_{se}}(\cdot) \]  

(6.45)

where \( \alpha \)'s denote the relative importance factors of each criterion. Typically, the factors are selected such that \( \alpha_{D(i,j)} L_{D(i,j)}(\cdot) \geq \alpha_{\Delta Q_{se}} L_{\Delta Q_{se}}(\cdot) \geq \alpha_{Q_{se}} L_{Q_{se}}(\cdot) \).

The congestion distribution factors (CDFs) proposed in [73] give good measure of the impact of transactions between buses to the congested lines. CDFs are derived from distribution factors. First, distribution factors in the usual sense are computed twice with respect to two different slack bus locations within the same system for the transmission line of interest, i.e. \( \{D_{m}^{(i,j)}\} \) and \( \{D_{n}^{(i,j)}\} \) where bus \( n \) is used as the slack bus for the first computation, and bus \( m \) is for the second. Then, the difference between these two sets of distribution factors, \( \beta_{m,n}^{(i,j)} \), is the result of having two slack buses in different locations. Defining the difference as

\[ \beta_{m,n}^{(i,j)} \{1\} = \{D_{m}^{(i,j)}\} - \{D_{n}^{(i,j)}\} \]  

(6.46)

where \( \{1\} \) is the vector of all ones, \( \beta_{m,n}^{(i,j)} \), can be expressed as [73]

\[ \beta_{m,n}^{(i,j)} = D_{m}^{(i,j)}(n) = -D_{n}^{(i,j)}(m) \]  

(6.47)

where \( D_{m}^{(i,j)}(n) \) denotes the \( n \)th element of the vector \( \{D_{m}^{(i,j)}\} \).

Define the shift vector, \( \phi \) as

\[ \phi^{i,j} = -\frac{D_{m}^{(i,j)}(i) + D_{m}^{(i,j)}(j)}{2} \]  

(6.48)

for given distribution factors, \( \{D_{m}^{(i,j)}\} \) with respect to the slack bus, \( m \). Then, we can subtract out the locational effect of the slack bus from the distribution factors by adding the sum of the shift vector elements to the given distribution factors. The resulting vectors are what is defined as CDF, \( \{D^{(i,j)}\} \):

\[ \{D^{(i,j)}\} = \{D_{m}^{(i,j)}\} + \phi^{(i,j)} \{1\} \]  

(6.49)

The magnitude of the resulting CDF defines the sensitivity of the flow in the transmission line of interest on a transaction; this formulation ensures that the sensitivity of the flow on the line of interest with respect to a bus injection decreases monotonically as the electrical distance between the line and the bus increases. The sign denotes if the transaction will increase or relieve the congestion.

The energy cost after relieving inter-cluster congestion is closely related to the computation of cluster-wide prices step in the implementation of the congestion cluster pricing method. As a matter of fact, the equations used for computing the energy cost and the cluster-wide prices are the same. Suppose the nodes
$g_i, g_{i+1}, \ldots, g_{i+k}$ are in the cluster $z_j$. Then, at some $t$ the new generation cost associated with the cluster $z_j$ is given by

$$C_{z_j}(Q_{z_j}) = f_{z_j}(Q_{g_i}, Q_{g_{i+1}}, \ldots, Q_{g_{i+k}})$$  \hspace{1cm} (6.50)$$

where $f_{z_j}$ is the monotonically increasing nonlinear function representing the least cost combination of $Q_{g_i}$'s in $z_j$ for producing $Q_{z_j}$. The marginal cost of zone $z_j$, $MC_{z_j}$, can be used in order to compute $f_{z_j}(\cdot)$ where

$$MC_{z_j} = \begin{cases} 
\left( \frac{1}{\alpha_{12}} + \frac{1}{\alpha_{23}} + \cdots + \frac{1}{\alpha_{i+1}} \right)^{-1} Q_{z_j} & Q_{z_j} \in R_{I_1} \\
\left( \frac{1}{\alpha_{m2}} + \frac{1}{\alpha_{m+1}} + \cdots + \frac{1}{\alpha_{n+1}} \right)^{-1} Q_{z_j} & Q_{z_j} \in R_{I_2} \\
\vdots & \\
\left( \frac{1}{\alpha_{m+n}} + \frac{1}{\alpha_{m+n+1}} + \cdots + \frac{1}{\alpha_{n+n}} \right)^{-1} Q_{z_j} & Q_{z_j} \in R_{I_n}
\end{cases}$$  \hspace{1cm} (6.51)$$

where $R_{I_i}$'s define the region of operating conditions in cluster $j$ with $q$ number of generators are still below the generation limits. $\alpha_i$'s represent the coefficient of the associated marginal cost of those generators below their generation limits.

With $C_{z_j}(Q_{z_j})$, the generation costs (and/or cluster-wide prices) are computed by solving the optimization problem given as

$$Q^{*}_{z_j} = \arg\min_{Q_{z_j}} \sum_{z_j} C_{z_j}(Q_{z_j})$$  \hspace{1cm} (6.52)$$

subject to the load flow constraint, i.e., total generation is equal to the system load,

$$\sum_{z_j} Q_{z_j} = \sum_{d_j} Q_{d_j} : \lambda$$  \hspace{1cm} (6.53)$$

the congestion interface flow limit constraints, i.e., the power flow on any line $l$ along only the congestion interfaces is within the maximum rating of the line,

$$|F_l| = \left| \sum_{z_i} H_{lz_i} Q_{z_i} - \sum_{d_j} H_{ld_j} Q_{d_j} \right| \leq F_{l}^{\text{max}} : \mu_l$$  \hspace{1cm} (6.54)$$

and the generation limit constraints, i.e., the dispatch amount in cluster $z_j$ is within the sum of the maximum rating of the corresponding generators within the cluster

$$0 \leq Q_{z_j} \leq \sum_{g_i \in z_j} Q_{g_i}^{\text{max}} : \eta_{z_j}$$  \hspace{1cm} (6.55)$$
The computation of $H_{1zi}$ yields

$$H_{1zi} = \frac{dF_i}{dQ_{g_i}} \frac{\partial Q_{gi}}{\partial Q_{zj}} + \frac{dF_i}{dQ_{g_{i+1}}} \frac{\partial Q_{g_{i+1}}}{\partial Q_{zj}} + \cdots + \frac{dF_i}{dQ_{g_{i+k}}} \frac{\partial Q_{g_{i+k}}}{\partial Q_{zj}}$$ (6.56)

with

$$\frac{dF_i}{dQ_{g_i}} = H_{1gi}$$ (6.57)

and with

$$Q_{gi} = \frac{1}{2a_i} \left( \frac{1}{2a_{i+1}} + \frac{1}{2a_{i+2}} + \cdots + \frac{1}{2a_{i+k}} \right)^{-1} Q_{zj}$$ (6.58)

if $Q_{gi} \in R_i$.

The solution to the optimization problem (6.52) then given by

$$\rho_{zi} = \lambda + \sum_i \mu_i H_{1zi}$$ (6.59)

where $\mu_i \neq 0$ if and only if $|F_i| = F_i^{\text{max}}$ and

$$Q_{gi} = \begin{cases} 
Q_{gi}^{\text{max}} & \rho_{zi,g_i} \in z_i \geq p_{gi}^{\text{max}} \\
\frac{\rho_{zi}}{2a_{gi}} & 0 \leq \rho_{zi,g_i} \in z_i \leq p_{gi}^{\text{max}} \\
0 & \text{otherwise}
\end{cases}$$ (6.60)

where $p_{gi}^{\text{max}} = 2a_g Q_{gi}^{\text{max}}$. Graphically, the above derivation has the following interpretation. Without loss of generality we consider a zone consisting of only two generators. Given the supply bids, $B_{G_i}$ and $B_{G_j}$ at nodes $G_i$ and $G_j$ respectively, the aggregate supply bid, $B_{z_k}$ for zone $z_k$ can be constructed as shown in Figure 6-7. For Region I

![Figure 6-7: Aggregation of Marginal Supply Bids in Zone k](image)

158
\[
\frac{dF_i}{dQ_{z_k}} = \frac{1}{2a_{ij}} \left( \frac{1}{2a_{ii}} + \frac{1}{2a_{ij}} \right)^{-1} H_{iG_i} + \frac{1}{2a_{ij}} \left( \frac{1}{2a_{ii}} + \frac{1}{2a_{ij}} \right)^{-1} H_{iG_j}
\]

(6.61)

and for Region II

\[
\frac{dF_i}{dQ_{z_k}} = H_{iG_j}
\]

(6.62)

The total energy cost after relieving inter-cluster congestion is then given by

\[
TC_{Q_\alpha} = \sum_{z_i} \rho_{z_i} Q_{z_i}
\]

(6.63)

The computation of the energy cost after relieving intra-cluster congestion is similar to that after inter-cluster congestion. The optimization problem to be solved in order to determine the location marginal prices is given by

\[
\Delta Q_{g_i} = \arg \min_{\Delta Q_{g_i} \in Z} \sum_{g_i} C_{g_i} (\Delta Q_{g_i})
\]

(6.64)

where

- \(\Delta Q_{g_i}\) : the adjusted generation amount at node \(g_i\)
- \(Z\) : the subset of clusters experiencing intra-cluster congestion
- \(C_{g_i}\) : the cost paid to the generator \(g_i\) based on the bid \(S_{g_i}\)

subject to the load flow constraint

\[
\sum_{g_i \in Z} \Delta Q_{g_i} = 0
\]

(6.65)

the transmission line flow limit constraints, i.e., the power flow on any line \(l\) in the entire system is within the maximum rating of the line,

\[
|F_l + \Delta F_l| = \left| H_{lG_i} (Q_{g_i} + \Delta Q_{g_i}) + H_{lD_j} Q_{d_j} \right| \leq F_l^{\max}
\]

(6.66)

and the generation limit constraints, i.e., the dispatch amount at node \(g_i \in Z\) is within the maximum rating of the corresponding generator

\[
0 \leq Q_{g_i} + \Delta Q_{g_i} \leq Q_{g_i}^{\max}
\]

(6.67)

The additional energy cost necessary for relieving intra-cluster congestion is then given by

\[
TC_{\Delta Q_\alpha} = \sum_{g_i \in Z} \alpha_{g_i} \Delta Q_{g_i} (Q_{g_i} + \Delta Q_{g_i})
\]

(6.68)

**Practical solution to the cluster design problem**

After the formulation we discover quickly that the so-called real value based methods are unlikely to yield a good result for solving this particular optimization problem. Real value based methods refer to analytical
approaches to finding the optimal solution which require a sequential improvement by examining the gradients of smooth trajectories in the system with respect to search space. The reason for the difficulty in applying real value based methods to the problem lies in the lack of a nice structure of the search space, Θ, such as continuity, differentiability, etc., which are essential for finding smooth trajectories and computing gradients. This leads to the belief that search based methods are more suitable for the optimization problem in Eq. (6.39). Search based methods refer to simulation supported approaches to finding the optimal solution which requires a ranking of all possible designs after a thorough evaluation of the performance of each design alternative.

In order to apply search based methods to the problem in Eq. (6.39) we first examine the search space, Θ. Suppose that the system is composed of $N_{TR}$ transmission lines and $N_B$ buses; $N_G$ generators and $N_D$ loads, and that the maximum number of clusters allowed is limited to $N_z$. Since once the maximum number of clusters are fixed, it is always possible to devise a cluster design to perform better than or at least equal to any existing design by allowing one more cluster in terms of $L(\theta, \xi[k], k)$ defined earlier [72], we start with a search space of size $|\Theta|$, given by

$$|\Theta| = N_z^{(N_B-N_z)}$$  \hspace{1cm} (6.69)

A typical electric power system consists of hundreds or thousands of buses, so conservatively let $N_B = 100$. Even if the number of clusters allowed is less than 10, assume $N_z = 5$, the number of designs to be considered using search based methods is given by

$$|\Theta| = 5^{(100-5)} \approx 2.5 \times 10^{66}$$  \hspace{1cm} (6.70)

which is typical by combinatorial standards.

Even though a further reduction in the size of Θ may be possible depending on the topology of the system, it is clear from examining the size of the search space that a brute force application of search based methods is not likely to be a good approach for any reasonable simulation time. Therefore, it is necessary to exploit any structural characteristics of the search space linked to the sample performance function.

One such characteristic is the first cut cluster design based on CDFs. Even though no analytical justification of effective measures is available, there are a few empirical results which suggest that the size of the search space can be reduced significantly by designing clusters based on CDFs, with little concern for carelessly excluding good designs from the remaining search space [71]. This is especially true if the sample performance function,

$$L(\cdot) = \alpha_{D(i,j)} L_D(i,j)(\cdot) + \alpha_{Q_{s_i}} L_{Q_{s_i}}(\cdot) + \alpha_{\Delta Q_{s_i}} L_{\Delta Q_{s_i}}(\cdot)$$  \hspace{1cm} (6.71)

is such that [72]

$$\alpha_{D(i,j)} L_D(i,j)(\cdot) \gg \alpha_{\Delta Q_{s_i}} L_{\Delta Q_{s_i}}(\cdot) \gg \alpha_{Q_{s_i}} L_{Q_{s_i}}(\cdot)$$  \hspace{1cm} (6.72)
A practical approach to the clustering design thus starts with the SO identifying the potentially critical lines, some of which may be congested at the same time or at different times. Typically, the number of critical lines, $N_{TR}$, is less than five, so again conservatively let $N_{TR} = 3$. For each of the three transmissions, the corresponding CDFs are computed. Then, based on the relative values of the CDFs the system is divided into clusters as described in [73]. Since there are multiple critical lines, the clusters defined for each line must be superposed on top of each other, and the intersections of the clusters constitutes the first cut design. The empirical results show that for a system of $N_B = 100$, three critical lines result in around 20 clusters. Given that the desired number of clusters is five, the search space of the problem is reduced from $2.52 \times 10^{66}$ to $3.05 \times 10^{10}$.

Although the size of $\Theta$ is reduced by orders of magnitude, the problem is still not manageable from an optimization point of view. Suppose 10,000 samples are selected randomly from $\Theta$ and serve as a sample set, $\Theta'$, for applying the search based method. The probability of the optimum solution from $\Theta$ being contained in this sample space is given by

$$\text{Prob}(\theta^* \in \Theta') = 1 - \left(1 - \frac{1}{3.05 \times 10^{10}}\right)^{10,000} = 3.28 \times 10^{-7}$$

which is less than unlikely.

Still the sample size must be further reduced to a manageable size before applying any search based method to Eq. 6.39. Fortunately much of the $3.05 \times 10^{10}$ designs in the search space is infeasible since geographically distant clusters after the first cut design cannot be combined to be included in the sample set. Some more topological characteristics allow for a further reduction of the size of the sample set. Even though generalizing the techniques of exploiting the topological characteristics of the system may be possible based on the recent developments in various graph partitioning methods, we employ a more heuristic approach to reducing the sample set. For instance, there are some rules of thumb, such as not allowing clustering near the critical lines, that significantly limit the possible designs to be included in the sample set. We claim without any analytical proof that the heuristic approach by an experienced system operator allows for a sample set containing around 1,000 designs from which at least 50 designs belong in the top 100 designs of the original search space for $N_B \approx 100$, $N_{TR} \approx 200$ and $N_z \approx 10$. Thus, by and large the complexity of finding an optimal solution to the problem in Eq. (6.39) is reduced from a search space of $|\Theta| \approx 2.5 \times 10^{66}$ to a sample space of $|\Theta'| \approx 1,000$.

**Application of ordinal optimization method**

Here we examine the optimization problem in Eq. (6.39) from the perspective of the ordinal optimization (OO) method. In dealing with search based methods applied to optimization problems, the (OO) method has been proven very effective [27]. The strength of the OO method is in its considerable savings of computational time when dealing with optimization problems with large search spaces and high uncertainty.
The basic idea of the OO method is the softening of the objective from finding an optimal solution to finding any design belonging to the “good enough” subset. For example, the good enough subset can be defined as the top-n% of the design space. The softening of the objective allows for working only in the much reduced selected subset with the confident expectation of a reasonable number of designs belonging to the good enough set. If the performance of each design is measured without any noise, the original optimization problem is transformed into the problem of selecting the design with the smallest evaluated performance belonging to the selected subset [27]. When the performance estimate is noisy, it becomes necessary to include more than one design in order to secure with higher confidence a certain degree of matching, or alignment, between the selected subset and the good enough subset [52].

Suppose that the size of the search space containing all possible designs is on the order of $10^{10}$, as is the case with our problem. By the goal softening principle we limit our goals to picking any of the top 5% designs. Consider a set consisting of 1,000 random samples from the search space. Then, the probability of retaining at least one of the top 5% designs in this sample space is given by

$$
\text{Prob}(G \cap \Theta' \neq \emptyset) = 1 - (1 - 0.05)^{1,000} \approx 1
$$

(6.74)

where $G$ and $\Theta'$ denote the set of the top 5% designs and the sample space respectively.

Similar to the idea of taking an exit poll from a limited number of voters in an election, if the designs in the sample space are chosen completely randomly, then we may assume that $\Theta'$ of the size 1,000 will more or less include 50 designs that belong to $G$. We can thus reduce the problem from finding any designs that belong to the set of the top 5% designs from a search space of size $10^{10}$ to finding any designs that belong to the top 50 designs from the sample space of size 1,000. The reduction of complexity is, indeed, quite considerable.

Let $G'$ denote the set consisting of the top 50 designs contained in the sample space $\Theta'$. Now consider the selected subset consisting of $s$ designs chosen randomly from $\Theta'$. We are interested in a necessary $s$ such that the alignment probability is defined as $\text{Prob}(|G \cap S| \geq k) \geq \mathcal{P}_A$ where $k$ and $\mathcal{P}_A$ are defined depending on the purpose. For example, let $k = 3$ and $\mathcal{P}_A = 90\%$. For the parameters given the equation for computing, the alignment probability is given as [27]

$$
\text{Prob}(|G \cap S| \geq 3) = \sum_{i=3}^{50} \left[ \frac{50}{i} \right] \left[ \frac{1,000 - 50}{|S| - i} \right] \geq 0.90
$$

(6.75)

Using Eq. (6.75) we deduce that the selected subset is required to have at least 102 designs in order to have at least 3 of them belong to the top 50 designs of the sample space.

This translates into a tremendous savings in computational time since a fairly accurate comparison of
approximately 100 designs will result in picking a design that is one of the top 50 designs of the sample space or one of the top 5% of the entire search space.

To summarize, the application of the OO method allows for a considerable savings of computational time in the obtaining of an acceptable solution to the optimization problem through search based methods while the method itself involves only the following simple steps [52]

1. selecting a sample set of size $N$, $|\Theta'| = N$

2. defining the goals: # of good designs, $g$, # of good design alignments in the selected subset, $k$, and the probability of alignment, $P_A$

3. determining the subset size, $s$, and selection rules that meet the goals

4. constructing the selected subset, $S$

5. comparing the designs in the selected subset

Before describing the method for accurately comparing the designs, we point out that the goal stated at the beginning is not a very impressive one since, given that the size of the search space is on the order of $10^{10}$, the top 5% designs include designs that are as far as $5 \times 10^8$ away from the true optimum.

For the optimization problem at hand, however, the top 50 designs in the sample space consisting of 1,000 are much better representatives than the top 5% of the entire search space. As discussed earlier, this is because the designs in the sample space are not picked randomly but through a rigorous testing of performance based on the first criterion for good cluster design. It is stated earlier that if the clusters are defined based on the CDFs, and if the importance of each criterion is defined such that first criterion is weighed on orders of magnitude higher than the other two, then the designs based on the CDFs are ranked much closer to the true optimum than the rest of the possible designs. It may not be possible to quantify accurately how much better the top 50 designs in the sample space are than the top 5% of the entire search space. However, it would not be surprising to find that the sample space contains at least 50 of the top 100 cluster designs from the search space if the clusters are defined based on the CDFs, with respect to the critical transmission lines identified by an experienced operator relying on many heuristic tools.

**Fairly accurate comparison of designs in the selected subset**

The ranking of each design alternative requires an evaluation of the sample performance. According to the three criteria for good cluster design, $L(\theta, \xi[k], k)$ is defined as a function consisting of a linear combination of three parts, namely $L_D^{(i,j)}(\cdot)$, $L_Q_{\xi_1}(\cdot)$, and $L_{\Delta Q_{\xi_1}}(\cdot)$ as shown in Eq. (6.45). Assume that the relative weights, $\alpha$, are chosen so that

$$\alpha_L^{(i,j)}L_D^{(i,j)}(\cdot) \gg \alpha_{\Delta Q_{\xi_1}}L_{\Delta Q_{\xi_1}}(\cdot) \gg \alpha_{Q_{\xi_1}}L_{Q_{\xi_1}}(\cdot)$$

(6.76)
Then we claim without proof that only \( L_{Q_{g_i}}(\cdot) \) and \( L_{\Delta Q_{g_i}}(\cdot) \) are relevant for evaluating the designs in the selected subset. The reason for this is because when the designs are chosen to be included in the selected subset, \( L_{D^{(i,j)}}(\cdot) \) is already used for comparison purposes. The designs in the selected set are assumed to have about the same \( \Delta Q_{g_i} \) compared to the others in the same set; otherwise the selected subset can be further reduced due to Ineq. (6.76).

Consider the modified sample performance, \( L'(\theta, \xi[k], k) \). We write \( \sum_{k=0}^{T} L'(\cdot) \) as

\[
\sum_{k=0}^{T} L'(\theta, \xi[k], k) = \sum_{k=0}^{T} \left[ \min_{Q_{z_j}[k]} \sum_{z_j} C_{z_j}(Q_{z_j}[k], k) \right.

+ \min_{\Delta Q_{g_i}, g_i \in \xi[k]} \sum_{g_i} C_{g_i}(\Delta Q_{g_i}[k], k) \left.
\right]
\]

subject to the load flow constraints at each hour

\[
\sum_{z_j} Q_{z_j}[k] = \sum_{d_j} Q_{d_j}[k] \tag{6.78}
\]

\[
\sum_{g_i \in \xi[k]} \Delta Q_{g_i}[k] = 0 \tag{6.79}
\]

the transmission line flow limit constraints\(^{11}\)

\[
|F_l[k]| = \left| \sum_{x_{i}} H_{l'x_{i},z_{i}}[k] Q_{x_{i}}[k] - \sum_{d_j} H_{l'd_j, z_{j}}[k] Q_{d_j}[k] \right| \leq F_{l{\max}}[k] \tag{6.80}
\]

\[
|F_l[k] + \Delta F_l[k]| = \left| H_{l{g}_i, z_{i}}[k] (Q_{g_i}[k] + \Delta Q_{g_i}[k]) + H_{l'd_j, z_{j}}[k] Q_{d_j}[k] \right| 
\leq F_{l{\max}}[k] \tag{6.81}
\]

and the generation limit constraints

\[
0 \leq Q_{z_j}[k] \leq \sum_{g_i \in \xi} Q_{g_i{\max}}[k] \tag{6.82}
\]

\[
0 \leq Q_{g_i}[k] + \Delta Q_{g_i}[k] \leq Q_{g_i{\max}}[k] \tag{6.83}
\]

In the formulation presented above, the uncertainty in the system is incorporated by considering

1. Load uncertainty
substitute Eq. (6.40) into \( Q_{D_{i}} \) in Eqs. (6.78), (6.80) and (6.81)

2. Generation bid uncertainty
substitute Eqs. (6.41) and (6.42) into \( \frac{dc_{g_i}}{dq_{g_i}} \) in Eq. (6.77)

\(^{11}\)prime denotes the lines only on the congestion cluster interfaces.
3. Equipment status uncertainty

substitute 0 for \( F_l^{\text{max}}[k] \) (or \( Q_{g_i}^{\text{max}}[k] \)) if the transmission line \( l \) (or the generator \( g_i \)) is in the “down” state.

With Eqs. (6.77) through (6.83) we can rewrite the cluster design problem as a stochastic optimization problem given by

\[
\theta^* = \arg \min_{\theta \in \Theta} \sum_{k=0}^{T} J'(\theta, k) \equiv \mathcal{E} \left[ \sum_{k=0}^{T} L'(\theta, \xi[k], k) \right] \tag{6.84}
\]

The expectation in Eq. (6.84) can be evaluated using the search based method (the Monte Carlo method) by

\[
\mathcal{E} \left[ \sum_{k=0}^{T} L'(\theta, \xi[k], k) \right] = \frac{1}{N_{\text{iter}}} \sum_{i=1}^{N_{\text{iter}}} \sum_{k=0}^{T} L'(\theta, \xi_i[k], k) \tag{6.85}
\]

where \( \xi_i \) represents the \( i \)th sample of the uncertainty [27].

It is recognized that, because the uncertainty is modeled using either a general version of a discrete time random walk or the transient Markovian chain, the number of probabilistic states that need to be evaluated grows exponentially with time \( k \) in order to compute Eq. (6.77). This is quite limiting when applying the search based method. Therefore, some modifications are necessary in order to simplify the optimization and make it manageable. One such modification is to work with steady state probability rather than transient probability.

**Steady state approximation of uncertainty**

To represent the uncertainty in the load and the uncertainty in the generation bid through steady state probability, the models described in [74] are useful. The formulation is identical to the approximate method used for pricing intermediate term transmission contracts, described in Chapter 5. We repeat it here for completeness.

First, to model the load we identify several basic load patterns: the typical peak load pattern, the normal load pattern and the off-peak pattern as shown in Figure 6-8, and the range of system load levels given in discretized steps of \( h \) MW starting from \( Q_D^0 \) MW, i.e., \( Q_D^{\text{tot}}(k) = Q_D^0 = kh \) as shown in Figure 6-9. Then

![Figure 6-8: Membership Functions for Individual Load Pattern](image)

165
in the model, if the total system load is larger than $Q_{[4]}$, the load distribution follows that of the peak; if the system load falls between $Q_{[2]}$ and $Q_{[3]}$, it follows a normal load distribution; and if the system load is less than $Q_{[1]}$, the off-peak load pattern is used to depict the load distribution. If the system load is either between $Q_{[1]}$ and $Q_{[2]}$ or $Q_{[3]}$ and $Q_{[4]}$ the appropriate patterns are meshed to create a typical individual load pattern. This process can be, employing fuzzy logic, summarized as

$$Q_D^{[k]} = \left( \eta^{[N]} \frac{Q_D^{[N]}}{1^T Q_D^{[N]}} + \eta^{[OP]} \frac{Q_D^{[OP]}}{1^T Q_D^{[OP]}} + \eta^{[PK]} \frac{Q_D^{[PK]}}{1^T Q_D^{[PK]}} \right)$$

(6.86)

where $\eta$ denotes the membership function. A similar approach is taken in order to model the generation bid. The details of the modeling the generation bid, using steady state probability, is referred to in [74].

To model uncertainty in the status of the equipment, the model presented through Eqs. (6.43) and (6.44) is used directly by considering the same probabilities as $k \to \infty$, i.e., the steady state approximation. The resulting probability is given by

$$\pi_0[\infty] = \frac{\lambda}{\lambda + \mu}$$

(6.87)

$$\pi_1[\infty] = \frac{\mu}{\lambda + \mu}$$

(6.88)

Using the probabilities given in Eqs. (6.87) and (6.88), the probability of a different system status can be derived. For example, the probability corresponding to having three transmission line failures is given as

$$\text{Prob(3 line outage)} = \left[ \begin{array}{c} N_{TR} \\ 3 \end{array} \right] \pi_0^{3[\infty]} \pi_1^{N_{TR} - 3[\infty]}$$

(6.89)

**Illustrative examples**

We illustrate the approach described in the paper using a simple test case shown in Figure 6-10. The system consists of 118 buses: 54 generators and 64 loads, and 186 transmission lines interconnecting the entire system; i.e., $N_B = 118$ ($N_G = 54$ and $N_D = 64$), and $N_{TR} = 186$. 

166
Figure 6-10: One line diagram of 118 bus (power flow test case) system
The congestion cluster pricing method is to be implemented on the 118 bus system for the cluster boundaries defined at \( k = 0 \) for a season consisting of 90 days (\( T = 2160 \) (hours)). The maximum number of clusters allowed is limited to, \( N_z = 15 \). Thus, the maximum size of the search space is \( 1.37 \times 10^{121} \) computed by

\[
|\Theta| = 15^{(118-15)}
= 1.37 \times 10^{121}
\]  

(6.90)

which is an astronomical figure.

For each load in the system, three types of load patterns are assigned: peak, off-peak and normal. Instead of introducing the uncertainty in the load as described in Eq. (6.86) each pattern is arranged to increase by a constant step at the beginning of the month \( (k = 720, 1440) \). An example of a load is shown in Figure 6-11.

The generator marginal supply bid is assumed to be a linear function of \( Q_{G,i} \), i.e. \( MC(Q_{G,i}) = 2a_G Q_{G,i} \),

![Figure 6-11: Demand pattern for load i in the system](image)

Figure 6-11: Demand pattern for load i in the system

with no uncertainty. Further, each generator in the system is assumed to be operational without outages throughout the season; thus there is no uncertainty in the status of the generators. The only uncertainty considered in the system is related to the status of the transmission line. The transmission lines in the system may experience outages with a failure rate of \( \lambda = 5 \times 10^{-4} \) and a repair rate of \( \mu = 0.5 \). For example, the probability associated with no transmission line failure is given by

\[
\text{Prob(no line outage)} = \left[ \begin{array}{c} N_{TR} \\ 0 \end{array} \right] \pi^0(\infty) \pi^{N_{TR}}(\infty)
= 83\%
\]  

(6.91)

Based on the system parameters it is determined that there are four critical lines in the system, namely the transmission lines between buses 30 and 38, between buses 59 and 63, between buses 70 and 71, and between buses 94 and 100. The critical lines are associated with the lines likely to be congested by a reaching of the transfer limits. Some of these lines may be congested at the same time or at different times, reflecting the stress being applied to the system in more than one possible way at different times throughout the season.
The first cut cluster design is performed for each of these critical lines based on the CDFs. For example, Figure 6-12 shows the cluster boundaries defined based on the CDF computed for the transmission line between buses 30 and 38. Once the first cut designs are determined, the clusters are superimposed on top of each other to create the clusters over the entire season. The resulting number of clusters after the superposition is found to be 18. Therefore, the maximum size of the sample space is reduced to a measly 3,375, computed by

\[
|\Theta'| = 15^{(18-15)} \approx 3,375 \quad (6.92)
\]

The size of the actual sample space is even smaller once the clearly inferior cluster designs (or infeasible cluster designs) are eliminated from the initial sample space. This results in \(|\Theta'| \approx 300\). From this sample space, 30 cluster designs are picked randomly to form a selected subset. The alignment probability for at least 3 matches in the selected subset of 30 designs for the top 50 designs is, approximately 91%, computed by

\[
\text{Prob}(|G \cap S| \geq 3) = \sum_{i=3}^{30} \left[ \begin{array}{c} 50 \\ i \\ 30-3 \\
\end{array} \right] \left[ \begin{array}{c} 300-50 \\ 30-i \\ 30 \\
\end{array} \right] = 90.91\% \quad (6.93)
\]

Finally, the performance function is estimated for each of these 30 designs in the selected subset, \(S\). The uncertainty in the status of the transmission line is not considered in this estimation step. Table 6.7 summarizes the estimated sample performance. As shown in the table, the three cluster designs with the smallest evaluated performances are \(\theta_9\), \(\theta_{13}\) and \(\theta_{32}\). Tables 6.8, 6.9 and 6.10 describe how the clusters are defined for each of these three designs.

For \(\theta_{13}\) we incorporate the uncertainty in the status of the transmission lines into an estimation of the
<table>
<thead>
<tr>
<th>Design</th>
<th>$\theta_1$</th>
<th>$\theta_2$</th>
<th>$\theta_3$</th>
<th>$\theta_4$</th>
<th>$\theta_5$</th>
<th>$\theta_6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L'()$</td>
<td>189.283</td>
<td>185.457</td>
<td>188.195</td>
<td>189.283</td>
<td>185.678</td>
<td>187.644</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Design</th>
<th>$\theta_7$</th>
<th>$\theta_8$</th>
<th>$\theta_9$</th>
<th>$\theta_{10}$</th>
<th>$\theta_{11}$</th>
<th>$\theta_{12}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L'()$</td>
<td>185.841</td>
<td>187.121</td>
<td>184.424</td>
<td>185.407</td>
<td>185.436</td>
<td>185.709</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Design</th>
<th>$\theta_{13}$</th>
<th>$\theta_{14}$</th>
<th>$\theta_{15}$</th>
<th>$\theta_{16}$</th>
<th>$\theta_{17}$</th>
<th>$\theta_{18}$</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Design</th>
<th>$\theta_{19}$</th>
<th>$\theta_{20}$</th>
<th>$\theta_{21}$</th>
<th>$\theta_{22}$</th>
<th>$\theta_{23}$</th>
<th>$\theta_{24}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L'()$</td>
<td>185.736</td>
<td>185.687</td>
<td>188.195</td>
<td>184.481</td>
<td>184.434</td>
<td>184.478</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Design</th>
<th>$\theta_{25}$</th>
<th>$\theta_{26}$</th>
<th>$\theta_{27}$</th>
<th>$\theta_{28}$</th>
<th>$\theta_{29}$</th>
<th>$\theta_{30}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L'()$</td>
<td>184.440</td>
<td>184.470</td>
<td>187.277</td>
<td>184.727</td>
<td>185.687</td>
<td>184.440</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Design</th>
<th>$\theta_{31}$</th>
<th>$\theta_{32}$</th>
<th>$\theta_{33}$</th>
<th>$\theta_{34}$</th>
<th>$\theta_{35}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L'()$</td>
<td>187.978</td>
<td>184.243</td>
<td>186.929</td>
<td>185.457</td>
<td>188.195</td>
</tr>
</tbody>
</table>

Table 6.7: Estimated sample performance for $\theta_i \in S$

<table>
<thead>
<tr>
<th>Cluster #</th>
<th>Bus #</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,117</td>
</tr>
<tr>
<td>2</td>
<td>25,26,27,28,29,31,32,114,115</td>
</tr>
<tr>
<td>3</td>
<td>16,17,18,19,30,113</td>
</tr>
<tr>
<td>4</td>
<td>20,21,22,23,24</td>
</tr>
<tr>
<td>5</td>
<td>37,38,39,40</td>
</tr>
<tr>
<td>6</td>
<td>33,34,35,36</td>
</tr>
<tr>
<td>7</td>
<td>79,80,89,99,100,101,102,103,104,105,106,107,108,109,110,111,112</td>
</tr>
<tr>
<td>8</td>
<td>43,44,45,46,47,48,49</td>
</tr>
<tr>
<td>9</td>
<td>41,42</td>
</tr>
<tr>
<td>10</td>
<td>50,51,52,53,54,55,56,57,58</td>
</tr>
<tr>
<td>11</td>
<td>59,60,61,62,66,67</td>
</tr>
<tr>
<td>12</td>
<td>63,64,65</td>
</tr>
<tr>
<td>13</td>
<td>77,78,82,83,84,85,86,87,88,89,90,91,92,93,94,95,96,97</td>
</tr>
<tr>
<td>14</td>
<td>68,69,70,74,75,76,81,116,118</td>
</tr>
<tr>
<td>15</td>
<td>71,72,73</td>
</tr>
</tbody>
</table>

Table 6.8: Individual bus cluster affiliation for $\theta_9$
<table>
<thead>
<tr>
<th>Cluster #</th>
<th>Bus #</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,117</td>
</tr>
<tr>
<td>2</td>
<td>27,28,29,31,32,114,115</td>
</tr>
<tr>
<td>3</td>
<td>16,17,18,19,25,26,30,113</td>
</tr>
<tr>
<td>4</td>
<td>20,21,22,23,24</td>
</tr>
<tr>
<td>5</td>
<td>37,38,39,40,41,42</td>
</tr>
<tr>
<td>6</td>
<td>33,34,35,36,43,44</td>
</tr>
<tr>
<td>7</td>
<td>103,104,105,106,107,108,109,110,111,112</td>
</tr>
<tr>
<td>8</td>
<td>79,80,98,99,100,101,102</td>
</tr>
<tr>
<td>9</td>
<td>77,78,82,83,84,85,86,87,88,89,90,91,92,93</td>
</tr>
<tr>
<td>10</td>
<td>94,95,96,97</td>
</tr>
<tr>
<td>11</td>
<td>50,51,52,53,54,55,56,57,58</td>
</tr>
<tr>
<td>12</td>
<td>59,60,61,62,66,67</td>
</tr>
<tr>
<td>13</td>
<td>63,64,65</td>
</tr>
<tr>
<td>14</td>
<td>45,46,47,48,49</td>
</tr>
<tr>
<td>15</td>
<td>68,69,70,74,75,76,81,116,118</td>
</tr>
<tr>
<td>15</td>
<td>71,72,73</td>
</tr>
</tbody>
</table>

Table 6.9: Individual bus cluster affiliation for $\theta_{13}$

<table>
<thead>
<tr>
<th>Cluster #</th>
<th>Bus #</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,117</td>
</tr>
<tr>
<td>2</td>
<td>25,26,27,28,29,31,32,114,115</td>
</tr>
<tr>
<td>3</td>
<td>16,17,18,19,30,113</td>
</tr>
<tr>
<td>4</td>
<td>20,21,22,23,24</td>
</tr>
<tr>
<td>5</td>
<td>103,104,105,106,107,108,109,110,111,112</td>
</tr>
<tr>
<td>6</td>
<td>33,34,35,36,37,38,39,40</td>
</tr>
<tr>
<td>7</td>
<td>79,80,98,99,100,101,102</td>
</tr>
<tr>
<td>8</td>
<td>41,42,43,44,45,46,47,48,49</td>
</tr>
<tr>
<td>9</td>
<td>83,84,85,86,87,88,89,90,91,92</td>
</tr>
<tr>
<td>10</td>
<td>50,51,52,53,54,55,56,57,58</td>
</tr>
<tr>
<td>11</td>
<td>59,60,61,62,66,67</td>
</tr>
<tr>
<td>12</td>
<td>63,64,65</td>
</tr>
<tr>
<td>13</td>
<td>77,78,82,93,94,95,96,97</td>
</tr>
<tr>
<td>14</td>
<td>68,69,70,74,75,76,81,116,118</td>
</tr>
<tr>
<td>15</td>
<td>71,72,73</td>
</tr>
</tbody>
</table>

Table 6.10: Individual bus cluster affiliation for $\theta_{32}$
sample performance. It turns out that the probability associated with multiple line outages is very small; i.e., less than 1.6%. Thus, we consider only single line outages. The newly estimated sample performance is given as

$$\mathcal{E} \left[ \sum_{k=0}^{T} L'(\theta_9, \xi[k], k) \right] = 183.012 \quad (6.94)$$

As expected, some slight correction is made to the earlier estimation of the sample performance.\(^{12}\)

\(^{12}\)This system exhibits a somewhat degenerate feature of the reduced system-wide generation cost with some of the lines taken out. This implies that the system operator may reduce the system congestion by cleverly controlling the existing resources.
Chapter 7

Transmission provision and pricing by multiple transmission providers (TP’s)

Figure 7-1 shows the three transmission systems serving the entire U.S., part of Canada and part of Mexico: (1) the Eastern Interconnected System, covering the eastern United States and some of the Canadian Provinces; (2) the Western Interconnected System, consisting of the western United States and the northern portion of Mexico; and (3) the Texas Interconnected System. The North American Electric Reliability Council (NERC) regions in Figure 7-1 refer to ten administrative areas established across North America in order to promote the reliability of the electricity supply following the systemwide blackouts on November 9, 1965 [85]. From the perspective of transmission network development it is important to note that the boundaries of the NERC regions are defined by aggregating 152 regional control areas into appropriate electrical geographic sizes rather than by being limited to the administrative utility boundaries. A control area is an entity that is electrically bounded through tie-line metering and telemetry, and it is responsible for maintaining its interchange schedule with other control areas and participating in the frequency regulation of the interconnection through scheduling, dispatching and controlling generation within its area. The Eastern Interconnection is comprised of 109 control areas, the Western of 33, and the Texas of 10.

In Chapters 2, 3, 5 and 6, we focus de facto on the role of the transmission provider (TP) in a single regional control area isolated from other control areas. The TP is assumed to have the sole operational authority of a control area and to alone be responsible for short-term reliability. Related to this operational authority, the TP conducts numerous off-line reliability studies so that the probability of network failure is below the acceptable limit. Based on these reliability studies the TP decides on the adequate level of interconnected operations services (IOS) required by the regional network. The IOS are the essential
HERC Regional Councils

ECAR = East Central Area Reliability Coordination Agreement
ERCOT = Electric Reliability Council of Texas
FRCC = Florida Reliability Coordinating Council
MAAC = Mid-Atlantic Area Council
MAIN = Mid-America Interconnected Network, Inc.
MAPP = Mid-Continent Area Power Pool
NPCC = Northeast Power Coordinating Council
SERC = Southeastern Electric Reliability Council
SPP = Southwest Power Pool
WSCC = Western Systems Coordinating Council

Source: Energy Information Administration, Office of Coal, Nuclear, Electric and Alternate Fuels.

Figure 7-1: North American Transmission Systems and NERC Reliability Council Regions
functions needed for the continuous balancing of generation and demand, transmission system security, and emergency preparedness under uncertainties [82]. On one hand, if the TP fails to acquire the adequate level of IOS by underestimating the uncertainties, then the reliability of the network operation is jeopardized. On the other hand, if the TP attains an excessive level of IOS by overestimating the uncertainties, then the efficiency of the network operation suffers. Thus, the tasks of determining the adequate level of IOS and subsequently assessing accurately the area-wide uncertainties are quite arduous and, at the same time, very important for reliability as well as for efficiency. These already difficult tasks become even harder to deal with when there are interconnections among neighboring control areas and transactions taking place across several market boundaries encompassing multiple control areas. This chapter describes the market mechanisms necessary for implementing inter-regional transactions while maintaining a high level of reliability and efficiency.

We first describe the advantages and disadvantages of having interconnections with neighboring control areas. Next, we described the newly proposed market mechanisms (and transmission provision) for implementing inter-regional transactions. The proposed mechanisms are then contrasted to the methods under the vertically integrated utility scheme and under the present restructuring process. Finally, the mechanisms are compared to other methods recently proposed in the industry.

7.1 Objective of interconnections with neighboring control areas; advantages and disadvantages

Consider the 5-bus electric power network as shown in Figure 7-2. The network is composed of two regions with each region having enough generation to meet its own loads. The network lines between the regions are called tie-lines, i.e., lines 4 and 5.

For illustration purposes, assume that at some hour $k$ the generators in the network are dispatched by the respective TP’s to meet the load as described in Table 7.1. By continually matching the supply and

<table>
<thead>
<tr>
<th>Generation at bus #</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Output (MW)</td>
<td>77.25</td>
<td>100</td>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Demand at bus #</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demand (MW)</td>
<td>0</td>
<td>157.25</td>
<td>0</td>
<td>120</td>
</tr>
</tbody>
</table>

Table 7.1: Nominal dispatch schedule for the 5-bus electric power network at hour $k$

demand, the network runs within the normal operating limits including the acceptable range of voltage and the typical frequency (of 60Hz in the US) [45].

Suppose the demand at bus 3 suddenly increases from 157.25MW to 207.25MW a few minutes after the beginning of hour $k$, deviating from the dispatch anticipated when scheduling. If region I is isolated from region II, then the area-wide frequency in region I initially drops by $2 \times 10^{-3}$Hz following the sudden load
increase. The area-wide frequency continues to drop until this drop in frequency is detected by the generators participating in the IOS and these generators react by increasing their generation to bring the frequency back to the level it was before the load increase. This particular service is often referred to as the regulation service, which provides generation response capability, under automatic generation control (AGC), in order to continually balance the supply with minute-to-minute load variations within the control area [82]. Figure 7-3 shows the change in the area-wide frequency in region I following an unanticipated load increase if region I is isolated from region II.

If region I is interconnected with region II, then the same unanticipated load increase in region I affects the network-wide frequency, instead of only the area-wide frequency, in a similar way. That is to say, following the load increase, the network-wide frequency initially drops by $1.8 \times 10^{-3}\text{Hz}$ and continues to drop until the deviation in frequency is detected by the generators participating in the IOS, and these generators react by increasing their generation to bring the frequency back to the level it was before the load increase. Figure 7-4 shows the change in the network-wide frequency in region I following the unanticipated load increase when region I is interconnected to region II.

However, it is evident from comparing Figures 7-3 and 7-4 that the temporary deviation in frequency is smaller (by about $2 \times 10^{-4}\text{Hz}$, or about 11%) when regions I and II are interconnected. This is due to the higher inertia carried within the interconnected network than within the isolated system. Plus, the recovery of the frequency is also much easier in the interconnected network than in the isolated system because the
Figure 7-3: The change in the area-wide frequency deviation in region I measured at bus 2 following an unanticipated load increase if region I is isolated from region II

Figure 7-4: The change in the network-wide frequency deviation in region I measured at bus 2 following an unanticipated load increase when region I is interconnected to region II
responsibility for recovering the frequency is shared among more generators. This is one of the biggest advantages of having an interconnected network.

In the case of more severe deviations in nominal operating conditions such as equipment outages, the sharing of IOS among many control areas becomes even more significant. For instance, suppose that each generator within the network has a 10% probability of failure while being dispatched equally at 50MW. Further suppose that in region I there are five generation units while in region II there are two generation units. If the reliability criterion specifies that no loss of load should occur for at least 85% of any operating conditions during an outage of at most one generation unit, then the isolated regions I and II are each required to have an additional 50MW of stand-by generation (a total of 100MW network-wide). With 50MW stand-by generation each, region I is fully operational at 91.85%, i.e.,

\[ 0.9185 = 0.9^5 + \left( \frac{5}{1} \right) 0.1 \cdot 0.9^4 \tag{7.1} \]

and region II is fully operational at 99.00%, i.e.,

\[ 0.9900 = 0.9^2 + \left( \frac{2}{1} \right) 0.1 \cdot 0.9 \tag{7.2} \]

If region I does not have the stand-by generation of 50MW, then the region is only operational at 59.05%, i.e., 0.5905 = 0.9^5. Similarly, if region II does not attain the stand-by of 50MW, then the region is only operational at 81.00%, i.e., 0.8100 = 0.9^2. For the interconnected network of regions I and II, however, a total of 50MW stand-by between the regions is necessary in order to meet this particular reliability criterion, i.e.,

\[ 0.8503 = 0.9^7 + \left( \frac{7}{1} \right) 0.1 \cdot 0.9^6 \tag{7.3} \]

The savings from requiring only 50MW of stand-by generation, rather than 100MW, may be tremendous. Incidentally, the stand-by generation service that makes additional capacity available to serve customer demand immediately should a contingency occur is often called reserves [82].

Beside the savings from sharing the IOS through the interconnected network, additional savings may be possible if the control areas linked through the tie-lines are significantly different in terms of the cost of the available generation resources. For example, suppose that the supply functions at the various buses are, as shown in Figure 7-5, based on the individual marginal costs of the generation units. At hour \( k \) let the demand of the loads at different buses be inelastic and given as \( Q_{d_2}[k] =, Q_{d_3}[k] =, Q_{d_4}[k] = \) and \( Q_{d_5}[k] = \). Table 7.2 summarizes the dispatch schedule determined through the market mechanism if regions I and II are isolated from each other. The difference in prices in region I is due to the binding network constraints of 80MW limit on transmission line 3. In comparison, Table 7.3 represents the dispatch schedule if the
Figure 7-5: The supply functions at buses 1, 2, 3 and 4 based on the individual marginal costs of the generation units

<table>
<thead>
<tr>
<th>Generation at bus #</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>For region I</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Output (MW)</td>
<td>74.50</td>
<td>82.75</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Price ($)</td>
<td>16.38</td>
<td>1.55</td>
<td>31.20</td>
<td></td>
</tr>
<tr>
<td>For region II</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Output (MW)</td>
<td></td>
<td></td>
<td></td>
<td>160.00</td>
</tr>
<tr>
<td>Price ($)</td>
<td></td>
<td></td>
<td></td>
<td>64.00</td>
</tr>
<tr>
<td>Total cost of generation ($)</td>
<td>15,588.57 = 1,348.57 (region I) +10,240.00 (region II)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7.2: Generation dispatch schedule if regions I and II are isolated
regions belong to the same market within a single control area. The result in Table 7.3 assumes that the

<table>
<thead>
<tr>
<th>Generation at bus #</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Output (MW)</td>
<td>156.39</td>
<td>78.84</td>
<td>0</td>
<td>82.02</td>
</tr>
<tr>
<td>Price ($)</td>
<td>34.37</td>
<td>1.47</td>
<td>67.27</td>
<td>32.81</td>
</tr>
<tr>
<td>Total cost of generation ($)</td>
<td>8,182.10 = 5,491.02 (region I) + 2,691.08 (region II)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7.3: Generation dispatch schedule if the regions I and II belong to the same energy market of a single control area

transmission charge levied on the market participants is only the congestion price without any additional costs such as the *ex ante* flow tax described in Chapter 3. Given that there typically exist transmission charges other than the congestion price and that the market mechanisms vary from one region to another due to regional characteristics, however, the result in Table 7.3 is neither likely feasible nor necessarily optimal. Nevertheless, comparing the total costs of generation in Tables 7.2 and 7.3, it is plausible to expect some savings if the control areas linked through the tie-lines are significantly different in terms of the cost of the available generation resources.

Therefore, an argument may be made that the major advantages of interconnection are improved reliability and efficiency through the sharing of IOS and further increased efficiency through cost savings in generation.

Suppose in order to take advantage of inexpensive generation cost, the loads at bus 5 enter into an energy contract with the supplier at bus 2 for 50MW. Then, using the DC load flow approximation (as described in Appendix B), the electric power flows through different network lines can be computed as shown in Figure 7-6. It is interesting to note that a significant amount of electric power (more than half of the entire transaction amount) flows through the tie-line 5 despite the availability of a closer tie-line, line 4, that could handle the entire transaction. From this example it may be deduced that if there are more than two control areas, then even if the proposed transaction takes place between two adjacent control areas, the rest of the interconnected network is also affected by the transaction. The so-called loop flow refers to the effect of electricity flowing not according to the possibly contracted transmission path (based on the corresponding energy contract) but rather according to physical law [81]. We consider this as the first of two types of loop flows; it is related to the inability of the market participants to control the transmission path.

The second type of loop flow is related to the inability of each individual TP to control the transmission path. For example, when the dispatch schedule is made by individual TP’s as given in Table 7.2 the markets at regions I and II are conducted completely separate from each other and no transactions between the two regions are committed, as shown in Figure 7-7. However, due to the presence of the tie-lines between the regions, the actual electric power flows through the network are realized as shown in Figure 7-8. We note that the actual flow through line 3 exceeds the operational limit on power transfer. This difference is quite significant since the network is being operated in a hazardous regime where the reliability of the system is
Figure 7-6: Electric power flows through network lines caused by 50MW transactions between the suppliers at bus 2 and the loads at bus 5.

Figure 7-7: Electric power flows based on the supply and demand determined through the market mechanisms separately between regions I and II.
Figure 7-8: Actual electric power flows through the network determined based on the supply and demand of regions I and II

no longer assured.

Since no single TP has complete control over the flows throughout the interconnected network as demonstrated by the loop flow of the second type, systemwide coordination becomes necessary in an interconnected network of many control areas in order to avoid a serious breach in the reliability [44]. Moreover, if we revisit the earlier example of increased reliability through the sharing of IOS, the need for systemwide coordination and strict tie-line flow control becomes even clearer.

Figure 7-9 shows the change in the frequency measured in region II following an unanticipated load increase when region I is interconnected to region II. Even though the deviation in frequency in region II is not as severe as in region I, the effect of the load increase at bus 3 is still felt there according to Figure 7-9.

As mentioned in Chapter 2 there are a number of network related controllers within the system, other than generators, also reacting to the deviations in frequency. These controllers are typically tuned around some anticipated operating conditions. Suppose that following the disturbance in frequency shown in Figure 7-9 several network related controllers, as well as the generators in region I, are activated in order to restore the frequency to its pre-disturbance level. If the result of the generators reacting to the deviation in frequency significantly alters the operating conditions in region I, the operating conditions in region II are also modified due to the loop flow of the second type. This is reflected in a change in flows caused by the different electric power flows entering or leaving region II through the tie-lines. The change is due exclusively to the effect of different operating conditions created outside region II since here we assume no generator in region II reacts to the deviation, and once the frequency is restored, the network related controllers are deactivated. Thus, from the perspective of region II, no change is made other than the electric power flow through the tie-line, and consequently the area-wide flows according to the loop flow of the second type. The next time region I
Figure 7-9: The change in the network-wide frequency deviation in region II measured at bus 4 following an unanticipated load increase when region I is interconnected to region II.

undergoes a similar kind of disturbance, the network related controllers in region II might not work properly because the controllers are initially tuned for certain operating conditions which may be quite different from the post-disturbance operating conditions.

To make the matters worse, it is not easy to tune the controller for the new operating conditions since the change in system conditions is entirely external. The TP in region II might not be exactly aware of the effect of the new operating conditions without full knowledge of the operating conditions in region I. The only way to ensure the proper functioning of the network related controllers in region II, therefore, is to restore the tie-line flows back to the pre-disturbance level so that the effect from the loop flow of the second type is minimized and the only change in operating conditions in region I is observed by the TP in region I only. The TP in region I can restore its own control area to a state of readiness for other contingencies since a full knowledge of the operating conditions in region I is assumed to be bestowed by the same entity. Incidentally, because of the difficulties in defining the controller settings based on numerous off-line reliability studies with respect to outside its own region, it is often implied that tie-line flows may change only once or twice within a day.

If the disturbance described above occurs, and the generators in region II react to the deviation in frequency by increasing their generation, instead of the network related controllers doing so, then at a time of electricity scarcity, there are also significant economic consequences in terms of “stealing electric power” as explained in [33].

Therefore, an argument may be made that the major disadvantage of interconnection is reduced reliability through the loop flows of the first and second types.

Given the advantages and the disadvantages of the interconnected network described above, the market
mechanisms necessary for implementing inter-regional transactions must have the following characteristics:

- They should maximize the improvement in reliability and efficiency realized through the sharing of IOS
- They should maximize the increase in efficiency realized through the cost savings in generation
- They must include the regional characteristics providing transmission when determining the optimal transactions
- They should minimize the effect of the loop flows of the first and second types through systemwide coordination and strict tie-line flow control

In the following section, we briefly describe the newly proposed market mechanisms in [40] for implementing inter-regional transactions.

7.2 Market mechanisms for implementing the inter-regional transactions as proposed in [40]

The overall market mechanisms for implementing the inter-regional transactions as proposed in [40] are composed of two parts: the auction mechanisms and the control mechanisms. Roughly speaking, the auction mechanisms are designed such that the apparent inter-regional transactions, as reflected in the tie-line flows, maximize the improvements in reliability and efficiency through the sharing of IOS and, at the same time, maximize the benefit achieved through the cost savings in generation while reflecting the appropriate regional characteristics of transmission provision of each control area. The control mechanisms allow the effect of the loop flows of the first and second types to be minimized. Here we give a brief description and illustrate the proposed market mechanisms through a simple example. Refer to [40] for a detailed description of the algorithm.

The main driver of the auction process is the so-called inter-regional transmission organization (IRTO) [33]. Under the proposed market mechanism in [40] the IRTO is a for-profit entity created solely to support the inter-regional transactions.

In the Northeast market, for instance, the IRTO will be on a scale large enough to embrace the Mid-Atlantic States and the Northeast Power Coordination Council (NPCC). Through an iterative auction, the IRTO clears the market for inter-regional transactions based on bids from RTOs - including Transmission Providers, Control Areas and Independent System Operators (ISO) - and marketers interested in inter-regional transactions. The IRTO coordinates the activities of the market participants as they maintain strict control of the tie-line flows for the duration of the transaction. This design of the IRTO makes the framework essentially independent of the type of market and the transmission tariffs in the regions within its boundaries.
The IRTO accepts bids from RTOs and marketers interested in inter-regional transactions. Through an iterative auction the IRTO clears the market for inter-regional transactions and coordinates the activities of the market participants as they maintain strict control of the tie-line flows for the duration of the transaction. In terms of hierarchical power system controls [39], the IRTO operates at the tertiary level. Based on the preferences of marketers, it establishes the optimal tie-line flow for a given period. Other market participants operate at the primary and secondary levels - not much different from the present operation of the power system. These marketers implement their transactions while ensuring that tie-line flows remain at the levels determined by the IRTO.

Consider the 5-bus electric power network example presented in Figure 7-2 at the beginning of this chapter. Due to the reasons explained earlier, the tie-line flow schedules are assumed to be adjusted no more than once a day. For simplicity without loss of generality assume that a day is composed of 2 hours and that the load demands in regions I and II consist of elastic and inelastic portions. On a typical day \( n \), the inelastic portion of the demand is given as summarized in Table 7.4. The elastic portion of the demand is created by the loads at bus 5 only. Given that there is a significant price differential between region I and region II, as shown in Table 7.2, it is suggested that this elastic portion of the demand is satisfied through the inter-regional transactions from the suppliers at bus 2.

<table>
<thead>
<tr>
<th>Demand at bus #</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
</table>
| Day \( n \), hour 1
| Demand (MW)     | 0 | 157.25 | 0 | 68.4 |
| Day \( n \), hour 2
| Demand (MW)     | 0 | 107.25 | 0 | 68.4 |

Table 7.4: Inelastic portion of demands in regions I and II

7.2.1 Auction mechanisms

At the beginning of the day, the TP’s in regions I and II first submit bids for utilizing the tie-lines, for reliability purposes, to the IRTO. Suppose that the two TP’s in the network are created from the respective vertically integrated utilities through the functional unbundling process, as is usually the case in the US. If the vertically integrated utilities in regions I and II had a limited exchange between them, ranging between 13MW and 23MW through tie-line 4, and between 16MW and 22MW through tie-line 5, then it is reasonable to infer that the existing network has evolved to perform at the highest reliability level when the exchange is within that range. For example, the network in region I is built to support operating conditions where the exchange between the regions is 15MW through tie-line 4 and 20MW through tie-line 5. Similarly, the network in region II is constructed to support operating conditions where the exchange between the regions is 21 MW through tie-line 4 and 18MW through tie-line 5. Hence, the reliability level of the entire network comprising regions I and II is first-class with minimal IOS if the exchange between the regions is within the
ranges typical under the vertically integrated utility structure. As the exchange deviates from these ranges, the TP’s, in order to maintain a similar level of reliability in the short term, may have to acquire more of the IOS or, in the long term, they may have to force the network to support operating conditions with new exchange schedules. Thus, the cost associated with the exchange from the perspective of the TP’s, in terms of reliability, may be as shown in Figure 7-10 for tie-line 4 and Figure 7-11 for tie-line 5. The negative costs in Figures 7-10 and 7-11 indicate the benefit to the TP’s, in terms of improved reliability, if they have an interconnected network rather than two isolated systems. Based on the combined costs the network-wide reliability level is highest, with minimal IOS, if the exchange is scheduled at 17MW through tie-line 4 and at 20MW through tie-line 5. Thus, if there are no economically motivated transactions scheduled by market participants, then the IRTO may schedule an exchange between regions I and II at 17MW and 20MW for the entire day n. It is interesting to note that the level of exchange here is much lower than what the systemwide optimal is expected to be, without considering the transmission network as given in Table 7.3 where the exchange is around 37MW through tie-line 4 and 41MW through tie-line 5.¹ The main reason for this difference is the lack of network support, inherited from the vertically integrated utility era.

Similar to the bids submitted by the TP’s, the network users also express the intent to use the tie-lines for inter-regional transactions in the form of bids to the IRTO at the beginning of day n. The bid is based on the benefit associated with cost savings due to purchasing from less expensive generation sources.

Suppose that the demand of the load at bus 5 is elastic. Given the higher cost of generation in region II

¹The comparison is not entirely accurate since the result given in Table 7.3 not only assumes the inelastic demand of loads at bus 5 but also considers only a one hour snapshot whereas here the exchange schedule is over a day composed of multiple hours. Nevertheless, a few key concepts may be conveyed by comparing the examples.
shown in Figure 7-5, the loads at bus 5 may want to satisfy some of its demand by making a purchase from the suppliers at bus 2. The overall benefit from realizing the transaction between bus 2 and bus 5 may, then, be as shown in Figure 7-12. The benefit function given in Figure 7-12 is typical, and the demand function for the desired transaction can be constructed by taking the first derivative of the benefit function.

When the actual tie-line schedule is determined, some parts of the flows are due to the TP’s utilizing tie-lines for reliability purposes while the rest are because of network users carrying out economically beneficial transactions. Thus, the difference between the flows due to the TP’s and those due to network users needs to be accounted for, and appropriate charging mechanisms need to be developed. The charging mechanisms are due to two factors. On one hand, the difference in flows results, from the perspective of the TP’s, in the deterioration of the reliability level if no further action is taken; in order to maintain the same level of reliability as before, the TP’s may have to incur additional costs in reinforcing the network and/or in purchasing more of the IOS. On the other hand, the difference in flows reflects the usage of the individual networks in regions I and II by network users involved in inter-regional transactions. Under the open access principle, the market participants and network users must be subject to equivalent transmission charges when employing the transmission system to satisfy energy needs by using the resources within the region and through the inter-regional transactions, respectively. By differentiating the usage of the tie-line by the TP’s from the usage by the network users, the TP’s can correctly impose network-related charges to the proper participants. Under the ex ante flow tax and congestion pricing scheme discussed in Chapter 3, the transmission costs levied on network users involved in inter-regional transactions may resemble those in Figures 7-13 and 7-14. The transmission costs shown in Figures 7-13 and 7-14 are used to compute the supply bids to be submitted to the IRTOS by the TP’s, so that the transmission charges reflecting the regional
Figure 7-12: Benefit associated with the transaction between the suppliers at bus 2 and the loads at bus 5 in terms of cost savings

Figure 7-13: Transmission cost to be levied on the network users involved in inter-regional transactions using tie-line 4
Figure 7-14: Transmission cost to be levied on the network users involved in inter-regional transactions using tie-line 5

characteristics of providing transmission are included in the auction mechanisms. It is interesting to note that, in the case when the \textit{ex ante} access fee and congestion pricing scheme or the \textit{ex ante} injection tax and congestion pricing scheme described in Chapter 3, are used instead, then the transmission charges levied on the network users involved in inter-regional transactions result in the so-called “pancaking” [84]. Pancaking refers to multiple transmission rates levied on transactions spanning several regional markets. This is due to inaccurately charging for transmission not based on flows but based on membership (in the case of the access fee scheme) or on injection (in the case of the injection tax scheme).

Once the bids are submitted, the IRTO can determine the tie-line schedules by minimizing the transmission cost and the reliability cost while maximizing the benefit associated with the cost savings due to purchasing from less expensive generation sources. For the 5-bus electric power network example above, the cleared bids result in the scheduled flows of 45.1MW through tie-line 4 and 46.5MW through tie-line 5 and an inter-regional transaction, between the suppliers at bus 2 and the loads at bus 5, of 91.6MW for both hours 1 and 2 on day \( n \).

7.2.2 Control mechanisms

Since the TP in region I is affected by the change in operating conditions in region II (and vice versa), if and only if the tie-line flows into or out of region I (or region II) deviate from the tie-line schedule, the ability of the individual TP in each region to operate its own network more or less independently from the other region depends greatly on how well the tie-line flows can be maintained at the scheduled level.

At the beginning of hour 1 on day \( n \) the TP’s in regions I and II conduct the respective regional markets in order to schedule generation dispatches to balance the supply and demand. Since a net of 91.6MW is
scheduled to be delivered from region I and region II, the generation dispatch following the overall market activities in region I produces 91.6MW of surplus in generation. Similarly, the generation dispatch results in 91.6MW shortage in generation. The surplus and the shortage are due to the inter-regional transaction between the suppliers at bus 2 and the loads at bus 5. Suppose the overall market activities produce the dispatch schedule shown in Table 7.5. Then, because of the loop flow of the second type, the flows are

<table>
<thead>
<tr>
<th>For hour 1 on the day n</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Generation at bus #</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Output (MW)</td>
<td>167.51</td>
<td>91.6</td>
<td>0</td>
<td>68.4</td>
</tr>
<tr>
<td>Demand at bus #</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Demand (MW)</td>
<td>10.26</td>
<td>157.25</td>
<td>0</td>
<td>160</td>
</tr>
</tbody>
</table>

Table 7.5: Dispatch schedule for regions I and II at hour 1 on day n

44.2MW through tie-line 4 and 47.4MW through tie-line 5, which are different from the scheduled flows of 45.1MW through tie-line 4 and 46.5MW through tie-line 5. Thus, in order to ensure the reliable operation of the interconnected network, there is a clear need for systemwide coordination and strict tie-line flow control so that the actual flows through the tie-lines match the scheduled flows.

This can be accomplished by implementing tertiary level control along with secondary level control and primary level control [39] [24]. Primary control refers to fast stabilization at the individual generator level with respect to the disturbance of fast dynamics nature. Secondary control refers to the automatic generation scheduling of frequency regulation at the control area level. Tertiary level control refers to the compensation for inadvertent flows between control areas by momentarily offsetting generator frequencies [24]. With the network assistance provided by the TP’s at the regional level, the IRTO can utilize various controllers, both generator-related and network-related (flexible AC transmission systems (FACTs) in particular), participating in inter-regional transaction support.

At the beginning of hour 2 on the same day, the TP in region I is required to conduct the regional market for the second time in the day because of the significant change in the load demands at bus 3. In contrast, the TP in region II has no need for any further market activities since the load demands at bus 5 remain unchanged from those of the previous hour. The dispatch schedule following the market activities at hour 2 is summarized in Table 7.6. It can be seen from Table 7.6 that the net generation is a 91.6MW surplus

<table>
<thead>
<tr>
<th>For hour 2 on the day n</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Generation at bus #</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Output (MW)</td>
<td>67.51</td>
<td>131.34</td>
<td>0</td>
<td>68.40</td>
</tr>
<tr>
<td>Demand at bus #</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Demand (MW)</td>
<td>0</td>
<td>107.25</td>
<td>0</td>
<td>160</td>
</tr>
</tbody>
</table>

Table 7.6: Dispatch schedule for regions I and II at hour 2 on day n
in region I and a 91.6MW shortage in region II. As before, due to the loop flow of the second type, tertiary level control is needed to match the actual flows through the tie-lines to the scheduled flows.

Figure 7-15 shows the result of employing tertiary level control to reinforce the flows through the tie-lines between regions I and II throughout day \( n \). Although it is not shown here, tertiary level control is used throughout the day, and not only when the some regional markets are conducted, in order to maintain the flows through the tie-lines. So, in the case of a sudden occurrence of plausible contingency\(^2\) in a particular region within the interconnected network, tertiary level control ensures that the other regions may operate without being affected, except for the few minutes immediately following the contingency.

Therefore, with the market mechanisms composed of auction mechanisms and control mechanisms, as proposed in [40], the inter-regional transactions may be implemented while maximizing the advantages and minimizing the disadvantages of the interconnected network. In the following section we describe the implementation of the inter-regional transactions both under the vertically integrated utility structure and under the current development, for comparison purposes.

\(^2\)The occurrence of contingencies, such as a natural disaster causing the loss of 75% of the transmission lines, leading to a systemwide blackout, needs to be excluded from the discussion.
7.3 Implementation of inter-regional transactions under the vertically integrated utility structure and under the current development

Under the vertically integrated utility structure the implementation of the inter-regional transactions is limited in scale, and the tie-lines are not designed to handle the import and export of large amounts of electricity over long distances that marketers would like to see in a deregulated electricity market. As described in the previous section, the principal reasons for having an interconnected network through tie-lines are more reliability-related than economics-driven. Thus, the amount of the transactions does not fully reflect the possible cost savings of importing electric power from inexpensive regions to expensive regions.

The electric utilities are established as (regulated) vertically integrated natural monopolies serving captive markets in a cost-plus business. Utilities function more as colleagues than competitors, since the very nature of the business prevents competition. With the assurance that costs can be passed on to the rate-payers, the utilities build extensive and fairly reliable systems with the main aim of moving power from the generating plants to the consumers with an appreciable level of reliability. With no need for competition, utilities trade power primarily to help meet acceptable levels of reliability, which limits the scale of inter-regional transactions.

The goal of maintaining reliability also means that vertically integrated utilities do not have to trade power over long distances. It is sufficient to import (or export) enough power from (or to) adjacent regions for reliability purposes. Inter-regional transactions are therefore limited in scope.

Additionally, utilities do not strictly control tie-line flows, but agree on, and monitor, the net inter-change between regions. Since it is difficult to assess the change in tie-line flows, from the agreed schedule between two adjacent system operators, to the systemwide reliability, typically the tie-line flows are agreed upon over a relatively longer period of time. The minimum time scale possible at the moment is considered to be one day. The flows on the lines are then somewhat loosely regulated and at the end of the period (a day, month or season), the deviation from the flows is paid back in kind as net. This means that a utility that experiences a net export for a number of hours within a particular operating period (peak or off-peak, say) would have to export to the other utility for the same number of hours during a similar operating period, and vice versa.

With the introduction of competition the characteristics of inter-regional transactions have changed as market participants attempt to take advantage of cheaper power in distant locations by transporting power over longer distances and across several regions and market structures. This has led to an increase in both the scope and scale of inter-regional transactions, making the management of transactions through voluntary cooperation insufficient.

There are several identifiable reasons for the change in management systems. First, the tie-line inter-
change can no longer be agreed upon by two adjacent system operators because they (1) do not have any incentive to do so and (2) the people who actually have the incentive to drive the inter-regional transfers often request the transfers that take place over multiple regions. This means that the participants in the transaction will have to deal with one or more intermediary regions in addition to the source and sink regions. Therefore a more structured approach to managing the transaction will be required than the selling and purchasing regions simply agreeing to a net scheduled interchange.

Second, the amount of energy involved in the transactions desired by marketers for economic reasons may exceed the amount assessed by the utility as necessary for optimal reliability. As mentioned earlier, the utilities design the tie-lines with a view to accommodating interchange quantities close to that assessed for optimum reliability. To accommodate the higher level transaction will then mean incurring additional cost to purchase IOS or to reinforce the transmission network. This will require a balancing of these costs and the economic benefits of implementing the transactions.

Third, the tie-line flows can no longer be regulated loosely since there are already examples of riding on neighbors to acquire power at high price hours and to return in-kind payment at low price hours; this is stealing since the price at each hour is different. Rather, there is the need for strict tie-line flow control. This will not only help to minimize the effects on regions not on the contract path (which are affected due to loop flows of the second type), but will facilitate the assignment of the costs involved to the appropriate agents involved in the transaction. It is important to note that implementing a control mechanism according to the proposed market mechanisms in [40] is not very different from what the industry practices under the vertically integrated utility. The only addition is the tertiary level control for strict tie-line flow, which does not require any additional equipment to be installed in the interconnected network.

Under the current development the inter-regional transactions are managed by an entity called the security coordinator (SC), which is independent of any merchant functions [81]. The SC is responsible for the safe and reliable operation of the interconnected network, including the several control areas managed by the respective TP’s.

First, the network users enter into various energy contracts for trading electricity across multiple regional boundaries. Of these contracts, the users involved in physical transactions determine the shortest transmission path possible between the injection point and the withdrawal point of each transaction. This transmission path is then used for measuring the usage of the network for carrying out the trade as specified by the contract. Because the transmission path decided on by the users is only for contractual purposes and is not related to the actual usage of the transmission system, it is called the contract path [81]. The users can reserve the transmission capacity necessary for the transaction along the contract path with the respective TP’s. There are 7 priority levels of transmission capacity reservation defined by NERC at the time of this writing. Once the necessary transmission capacity reservations is made over the specified period of time, according to preference the suppliers (and loads) involved in the transactions may inject (and take out) the amount of power specified by the contract into (and from) the interconnected network.
Then, while the various inter-regional transactions take place as specified by the respective contracts, any TP may call for so-called transmission loading relief (TLR) procedures to be implemented by the SC in the case of any violations of the operating security limits, typically network-related limits such as the transfer limits on flowgate\(^3\), believed to be caused by inter-regional transactions. It is assumed that if any of the operating security limits defined by the individual TP in each region are violated, the reliability of the entire interconnected network is in danger of being lost. The TLR procedure is a method for mitigating potential or actual operating security limit violations [81]. When particular operating security limits are violated (requiring the implementation of the TLR procedures), the SC identifies the likely inter-regional transactions causing the violations using a simple computation called the interchange distribution calculator. The identified transactions are then curtailed on the order of the lowest level of transmission capacity reservation until the system conditions are again within the operating security limits. The details of the TLR procedures may be found in [81].

There are several inefficiencies associated with the inter-regional transactions managed by the SC due to the improper placement of incentives and responsibilities. We discuss a few of the rather major inefficiency issues here.

First, one of the major problems under the SC scheme is that the TP in each region has no strong incentive to define clearly the operating security limits related to inter-regional transactions. Suppose some TP’s define several flowgates in the interconnected network without carefully considering the projected inter-regional transactions to take place. If this is true, operating within the security limits of these flowgates may not ensure the reliability of the system because of the effect of the inter-regional transactions. Similarly, a violation of the security limits may not mean a degradation in reliability. Given that the TP’s are only responsible for the safe and reliable operations of their respective networks, the security limits defined for flowgates are likely to be highly conservative without thoughtful concern given to the economic aspects of the inter-regional transactions. In comparison, the proposed market mechanisms in [40] instigate the TP’s to consider carefully the effect of the inter-regional transactions through the bids associated with the reliability and transmission costs.

The other major problem is related to the passive nature of the SC. Upon TP’s request for the implementing of TLR procedures, the SC identifies the likely inter-regional transactions causing the violations and then curtails those transactions on the order of the lowest level of transmission capacity reservation until the system conditions are again within the operating security limits. Before the implementation of the TLR procedure is requested, however, the SC is not in any way involved in the inter-regional transactions. Given that the SC may be most familiar with the operations of the interconnected network, the SC can support the network users by identifying the truly economical inter-regional transactions which result in savings not only in generation costs but also in systemwide IOS costs, etc. by avoiding the transactions which may cause the

\(^3\)Flowgate refers to the transmission link associated with likely network congestion.
implementation of the TLR procedures. Under the proposed market mechanisms in [40] the IRTO, which effectively carries out the functions of the SC, participates proactively in the market process of realizing the most efficient inter-regional transactions by clearing the bids before the reliability is threatened, not reactively by implementing TLR procedures after the reliability-related problems are identified.

Finally, there is another major problem linked with the restoration of the interconnected network back to within the operating security limits. With a number of curtailments implemented by the SC following the TLR procedures, the operating conditions may no longer violate the security limits on the flowgates. However, this type of rigid process of restoring the network can hardly be optimal given the constantly changing nature of network conditions. In some cases it may be more reliable not to implement TLR procedures immediately following the violation of the security limits because the system condition may soon be changed so that carrying all of the inter-regional transactions supports the overall network better than curtailing some of the transactions. In comparison, the restoration of the interconnected network under the proposed market mechanisms in [40] is based on fundamentally sound technical criteria and mostly utilizes the existing controllers to adjust constantly around evolving system conditions.

Therefore, with the market mechanisms proposed in [40], many issues related to the current SC scheme are resolved because the implementation is based on technically sound fundamentals while incorporating proper economical incentives. Plus, it is not very difficult to implement the proposed mechanism since the underlying structure is already in place. That is to say, the only necessary improvements are replacing the SC with the for-profit IRTO and substituting for the reactionary TLR procedure with the proactive bidding process. In the following section we describe the implementation of the inter-regional transactions under other proposed market mechanisms, for further comparison purposes.

### 7.4 Other proposed market mechanisms for implementing the inter-regional transactions

At the time of this writing, there are currently two main proposals on how to replace the SC scheme for implementing inter-regional transactions. For convenience we refer to them as (1) the coordinated optimal power flow method and (2) the flowgate rights allocation method across multiple regions.

#### 7.4.1 Coordinated optimal power flow method across multiple regions

The coordinated optimal power flow (OPF) method is mainly based on the analyses given in [16], [10] and [50]. The method is based on the nodal pricing paradigm and seeks to attain a system-wide cost-based OPF using a coordinated, distributed method. The price of transmission is calculated from the differences in prices of energy at the various nodes. Only a brief description of the method is given here for discussion purposes; a detailed explanation of the method is found in [16], [10] and [50].
In this approach each control area performs a system-wide economic dispatch. However, the operator in each area considers as binding only the constraints in his area. Constraints on lines outside of his control area are considered as added costs in his objective function.

In an iterative process each operator reports the net loads and locational congestion costs arising from constraints in his region that would apply to adjustments in the net loads at any location in the grid. Each control area operator then adjusts the energy prices and schedules and recalculates the new transmission prices in his area based on the adjusted nodal prices.

To illustrate with the 5-bus system in Figure 7-2, first each of the operators will balance his respective markets and arrive at the desired operating levels. It is assumed that tie-lines belong to one of the 2 regions; for instance line 4 may belong to region I and line 5 to region II. If the schedules are feasible for the interconnected system, no redispatch will be required. However, redispatch will be required if the schedule in one region causes a violation in another when implemented simultaneously. For example the simultaneous dispatch may result in a line flow on line 3 that is in excess of the 80MW limit.

If a redispatch is necessary, then regions I and II will exchange information on their net loads and adjustment bids for generators. Region I will then perform a system-wide economic dispatch using the net loads from region II, and the adjustment bids to price generation in region II. In this case the operator explicitly models the actual limits on lines 1, 2 3 and 4 only, and assumes that lines 5 and 6 are limitless. Region II includes limits on lines 5 and 6 only, meaning that it ignores the 80MW limit on line 3.

Based on the resulting solution, each operator can determine the locational congestion costs arising from the constraints that would apply to adjustments in the net loads at any location in the grid. The operators exchange information on net loads and adjustment bids and congestion costs update estimates of net loads, reformulate their economic dispatch problems to include the adjustment bids and the associated congestion costs from the other regions, and perform a new redispatch.

This process continues until there is no significant change in the dispatch of either region.

There are several inadequacies associated with the coordinated OPF method across multiple regions because of the impracticality in implementation. We discuss a few of the major impracticality issues here.

First, one of the major problems under the coordinated OPF method is that there are inherent difficulties in defining security limits for the network. As described earlier, the security limits are defined as a result of numerous off-line reliability studies. This entails, at the minimum, establishing a few system operating conditions around which the regional network is usually being managed. These operating conditions are often referred to as nominal conditions. Although it is not trivial, the establishing of nominal conditions is a doable task for a individual TP so long as the uncertainties to be considered are contained within its own region. Thus, often times, the uncertainties associated with the interactions to outside the region are responded to by modeling several possible exchanges through tie-lines. If the exchanges through tie-lines are expected to vary extensively, then the security limits may need to be time-varying as well or, at the least, may mean quite different levels of reliability. If there is a minimum level of reliability to be achieved, then this requires
procuring different amounts of IOS. Since under the coordinated OPF method the security limits need to be defined by the TP in each control area without considering the amount of the IOS necessary for keeping the acceptable level of reliability with respect to possible contingencies in the other control areas, the security limits are either very conservative or time-varying as the system conditions change. Given that the TP’s are only responsible for the safe and reliable operation of their respective networks, the security limits are likely to be defined as highly conservative rather than as time-varying, and consequently a significant efficiency loss is expected. In comparison, based on the proposed market mechanisms in [40], the process of defining actual security limits is internalized by the individual TP in each region while the change in reliability level (or the different amounts of IOS to be procured) is allowed to be directly communicated to the network users through the bids, so that a higher efficiency is achieved.

Another major problem is related to the inability to convey the regional characteristics of individual control areas in deciding the transfer across multiple regions. Some control areas may have more expensive transmission networks due to many peculiarities in the region, such as higher property costs and so on. By treating each line in the entire interconnected network in the same way, the regional tariff structures developed to be best suited for the respective regions by the market participants are completely ignored in implementing the coordinated OPF method. Under the proposed market mechanisms in [40] these regional characteristics are respected by allowing the TP’s to submit separate bids accounting for the usage of their respective transmission networks.

Finally, there is a problem linked with the restoration of the interconnected network as operating conditions change. If any one of the regions goes through a significant change in operations, then the operating conditions for the rest of the interconnected network need to be modified in order to accommodate this change. A good example was discussed in Table 7.4: when the energy market in region I is conducted to meet a significant change in the load demands at bus 3, the energy market in region II also needs to be conducted again to make certain that no security limits are violated in region II due to the change in region I. If the continuously evolving operating conditions are considered due to the plausible contingencies, as in the case in the electric power network, this implies that the various energy markets in the entire interconnected network need to be synchronized so that any change in operating conditions in one region does not result in a violation of the security limits in other regions. In comparison, under the proposed market mechanisms in [40], the effect from any changes in operating conditions in one region is contained within the region once the tertiary level control mechanism restores the interconnected network following any plausible contingencies.

Thus, in order to properly implement the coordinated OPF method for managing inter-regional transactions, a significant number of modifications needs to be made to the network, the least of which being the synchronization of the market activities throughout the entire interconnected network. This is quite the contrary to the proposed market mechanism which requires only minor modifications to the network.

197
7.4.2 Flowgate rights allocation method across multiple regions

The flowgate rights approach to inter-regional transaction management is representative of the link-based approach [17]. Although the term flowgate may refer to any transmission line in the system, in general it refers to the links in the network that are likely to be congested.

The flowgate method is a system of flow-based transmission rights that, unlike the contract path approach, attempts to match scheduled transactions with the actual power flow by using the power transfer distribution factors (PTDF) derived from Kirchhoff's laws to translate the physical effects of each energy transaction into requirements of transmission rights.

The underlying market structure assumed for flowgate rights is rate of return regulation imposed on the transmission owners and operational authority assigned to a non-profit independent system operator (ISO). Under this market structure, market participants submit bids to purchase flowgate rights once at the beginning of the year (or season). The ISO then determines the price and amount of flowgates to be made available, and allocates the network capacity corresponding to the flowgate rights based on the bids. Each flowgate right issued to a participant specifies, at least, the designated flowgate (i.e. the line that is likely to be congested) and the capacity offered on that flowgate.

The flowgate rights grant the holder a capacity reservation or scheduling priority for using specific transmission links. If the holder fails to use the right by scheduling power transactions, the scheduling priority expires and the right reverts to the system operator. The holder can therefore not use it to prevent others from accessing unused transmission capacity.

Once the allocation of flowgate rights is concluded, two separate markets, the forward and spot markets, are conducted sequentially. First, the participants in the forward market arrange for transactions and acquire the flowgate rights necessary to implement the transactions from the current holders. If a participant arranges a transaction (backed by a flowgate) that reduces congestion at another flowgate, then he becomes the holder of the newly created flowgate rights in the amount by which the congestion is reduced. The process continues until all the transactions arranged are covered by flowgate rights. The network capacity of unused flowgate rights are returned to the ISO which then conducts a spot market.

Marketers who choose to participate in the spot market can submit bids to that market. The ISO clears the spot market by solving the OPF problem subject to the network capacity limits which are redefined to incorporate the unused flowgate rights. Again, as a result of the market clearing process, the combined price of the energy and transmission portions of the electric services are determined by the nodal prices at each bus. The ISO collects and distributes revenue that is determined by the product of the injection into the bus and the corresponding nodal price. Part of the revenue is used to compensate holders of unused flowgate rights.

One of the major problems under the flowgate rights allocation scheme is that the TP in each region has to define ahead of time the amount of rights available exclusively for inter-regional transactions. For
reasons related to the maximum possible number of tie-lines schedulings being only once or twice per day as discussed earlier, it is often implied that inter-regional transactions need to be handled separately from the energy markets for trades within the region. Then, when a TP offers the flowgate rights for inter-regional transactions only, either the TP needs to estimate the flowgate rights needed for the trades within the region, or the TP needs to conduct the auction process once for the market participants within the region and once for the inter-regional transactions together.

In that case the TP needs to estimate the available flowgate rights, a strong incentives is required to project well the usage of flowgate rights in the energy market within the region. Suppose it is found after conducting the spot market that some TP’s overestimated the usage of the available capacity through the flowgates. Then, the flowgate rights offered to the network users involved in inter-regional transactions may not ensure the scheduling priority desired.

In the case the TP needs to conduct the auction process once for the market participants within the region and once for the inter-regional transactions together, the markets for the entire interconnected network need to be conducted in a synchronized fashion with the majority of transactions being taken care of through this market process, leaving only the unanticipated balancing in the spot market at each region. However, as is pointed out in [30], many of the transactions in the current electricity markets still rely heavily on the spot market process. So long as this is the case, the markets under the flowgate scheme may not achieve high efficiency.

On top of the problem mentioned above, the problem still exists about how to accurately assess the total amount of the flowgate rights to be offered by the individual TP’s with very limited knowledge about the operations in the other regions. In comparison, based on the proposed market mechanisms in [40], the network users involved in the inter-regional transactions are handled completely separately from the market participants due to the proactive participation of the TP’s. Plus, instead of defining a rigid amount of flowgate rights available, the individual TP in each region may reach a higher efficiency by communicating to the network users the change in reliability level through the bids.

In addition, there is a problem linked with the change in the amount of flowgate rights available in a region due to the evolving operating conditions in the rest of the interconnected network. For example, when the operating conditions in one region change, some operating conditions believed to be secure in some other regions may no longer be so. Then, the operating security limits for certain links in those regions need to be adjusted. If the link on which the flowgate rights are issued happens to undergo the adjustment, then the amount of the flowgate rights available on that link also changes [30]. Because of this problem the amount of flowgate rights offered may be highly speculative and may require continual adjustment depending on the evolving operating conditions of the entire interconnected network. Under the proposed market mechanisms in [40], this problem is resolved by minimizing the effect of any disturbances which might be propagating throughout the interconnected network by IRTO performing systemwide coordination and strict tie-line flow control.
Therefore, with the market mechanisms proposed in [40], many issues related to the flowgate rights allocation method are resolved because of IRTO’s presence. By having an entity solely responsible for handling inter-regional transactions, the regional energy markets can be conducted separately from these transactions and may co-exist while having very different characteristics from one another. This is important since in order to achieve higher efficiency, the well functioning markets need to reflect the unique features of the respective regions. Plus, the systemwide coordination and strict tie-line flow control allow for the further independence of each regional market.
Chapter 8

Concluding remarks

The operation and planning of the 2-bus electric power network shown in Figure 8-1 are examined in order to recapitulate the contribution of the thesis.

![One-line diagram of 2-bus electric power network](image)

Figure 8-1: One-line diagram of 2-bus electric power network

Before discussing the example in detail, it is recognized that there are a few general assumptions and assertions made about the network throughout the thesis. They are re-stated here for completeness.

(A1) First, it is asserted that the characteristics of the electric power network are such that the supply and the demand need to balance continuously and that the transfer of electricity from a supplier to a load affects the entire network. The former is due to the lack of practical means for storing electricity, and the latter is due to the lack of practical means for directing electric power flows through an arbitrarily designated path in a network.

Starting from this assertion, the scope of the thesis is limited to studying only steady state problems and does not include the dynamic problems in the operation and planning of an electric power network. As a consequence of this restriction in scope, all analyses are performed considering only very simple components such as generators, loads, and transmission lines.

The scope of the study is further limited by considering real power related problems only; there is no consideration of reactive power related ones. This allows all functions described in the thesis to be defined without using reactive power related variables.

(A2) Then, it is asserted that the following uncertainties need to be considered in any electric power network: (1) the uncertainty (random fluctuation) in the utility functions of various loads and (2) the uncertainty
(random outages) in the status of equipment, either the generator or transmission lines.

(A3) Among these uncertainties a simplifying assumption is made that the uncertainty in the utility function of a load can be separated into two different time scales: hourly around some mean and continuous around a zero mean while treating the hourly utility functions of each load as being independent from hour to hour, i.e. the demand at hour k does not depend on the demand at hour l for \( k \neq l \). For convenience, different variables \([k]\) and \((t)\) are used for indexing the functions evolving in the hourly time scale and the functions evolving in the continuous time scale, respectively.

With this simplifying assumption, the problem of balancing supply and demand for electricity can be dealt with in two separate time scales as well. This thesis considers the supply and demand evolving hourly only. As a consequence, the cost functions for generators, the utility functions for loads, and the cost functions for transmission lines are defined using only discrete time variables.

(A4) It is assumed that all equipment (both generators and transmission lines) is subject to hard rating constraints. Hard rating constraints refer to any violation in these constraints resulting in severe equipment damage. As a consequence, the analyses performed in the thesis are limited to the operating conditions which do not violate any of the rating constraints even under various uncertainties.

(A5) It is asserted that there are only three factors affecting the ratings of the generator or transmission lines, namely reinforcement through investment, (physical) depreciation, and outages. Through reinforcement, the ratings of equipment can be increased at a given investment cost. Once the investment is made, the physical capability of equipment deteriorates gradually thus reducing the ratings over time. At any given moment, the equipment may be temporarily taken out of service, reducing its ratings either partially or entirely due to the outages.

It is evident from the assertion made in A2 and the assumption made in A4 that a number of mechanisms is needed for dealing with uncertainties so that none of the rating constraints are violated at any time. Typically, so-called ancillary services (including reserve services) are defined in the operation of the electric power network in order to deal with the uncertainties. For example, a certain percentage of generation capacity is put aside as operating reserves so that in case of an equipment outage the systemwide generation may be re-dispatched in order to restore the operation of the network to within the rating constraints following the outage. Therefore, depending on the availability of the ancillary services, the network is not allowed to operate at certain operating conditions even though they satisfy all of the network constraints before an outage if it is not possible to restore the network to within the constraints following the outage. One way to avoid certain operating conditions is by defining the transmission line flow limits not as respective ratings but rather to include the effect of the uncertainties. This remains an open problem.

(A6) Consequently, based on (A5), it is assumed in the thesis that the relationship between the transmission line flow limits and the state of a given electric power network (including the effect of outages) is given, so that the uncertainties in equipment status need not be considered. The implications of the assumption of the effectiveness of market designs and reliability-related risk management remain an open problem.
Finally, it is assumed that there is no practical substitute for electricity.

In the 2-bus network example, the operation and planning of the network over a four-year period are considered. For simplicity without loss of generality, here it is assumed that a year consists of 4 hours, i.e., $k = 1, 2, \cdots, 16$. The network is composed of 2 generation substations, 2 load centers and two identical parallel lines connecting two buses.

Suppose the utility function for bus 1 is given as the following:

For years 1 and 2, i.e., $k = 1, 2, \cdots, 8$

$$U_{d_1}(Q_{d_1}[k], k) = -10^7 \cdot (Q_{d_1}[k] - \beta_{d_1}[k])^2 + 342.25 \times 10^7$$  \hspace{1cm} \text{(8.1)}

where

$$\text{Prob}(\beta_{d_1}[k] = 18.50\, MW) = 100\%$$  \hspace{1cm} \text{(8.2)}

This way of modeling the utility function is based on the linear demand function assumption and assumes no uncertainties. For years 3 and 4, i.e., $k = 9, 10, \cdots, 16$

$$U_{d_1}(Q_{d_1}[k], k) = -10^7 \cdot (Q_{d_1}[k] - \beta_{d_1}[k])^2 + 377.52 \times 10^7$$  \hspace{1cm} \text{(8.3)}

where

$$\text{Prob}(\beta_{d_1}[k] = 19.43\, MW) = 100\%$$  \hspace{1cm} \text{(8.4)}

Similarly, the utility function for bus 2 is given as the following:

For years 1 and 2, i.e., $k = 1, 2, \cdots, 8$

$$U_{d_2}(Q_{d_1}[k], k) = -10^7 \cdot (Q_{d_1}[k] - \beta_{d_1}[k])^2 + 24479.73 \times 10^7$$  \hspace{1cm} \text{(8.5)}

where

$$\text{Prob}(\beta_{d_1}[k] = 156.46\, MW) = 80\%$$  \hspace{1cm} \text{(8.6)}

$$\text{Prob}(\beta_{d_2}[k] = 161.15\, MW) = 10\%$$  \hspace{1cm} \text{(8.7)}

$$\text{Prob}(\beta_{d_2}[k] = 151.77\, MW) = 10\%$$  \hspace{1cm} \text{(8.8)}

Here some uncertainties are modeled in the utility function but assume no intertemporal effect. It should be noted that if there is no intertemporal aspect to the uncertainties affecting the evolution of the system states, the dynamic programming (DP) formulation is not necessary. For years 3 and 4, i.e., $k = 9, 10, \cdots, 16$

$$U_{d_2}(Q_{d_1}[k], k) = -10^7 \cdot (Q_{d_1}[k] - \beta_{d_1}[k])^2 + 26987.92 \times 10^7$$  \hspace{1cm} \text{(8.9)}
where

$$\text{Prob}(\beta_{d_2}[k] = 164.28\,\text{MW}) = 80\% \quad (8.10)$$

$$\text{Prob}(\beta_{d_2}[k] = 169.21\,\text{MW}) = 10\% \quad (8.11)$$

$$\text{Prob}(\beta_{d_2}[k] = 159.35\,\text{MW}) = 10\% \quad (8.12)$$

Suppose at \( k = 0 \) the generation substation at bus 1 consists of \( N_h \) hydro units only while the generation substation at bus 2 consists of \( N_n \) number of nuclear units and \( N_t \) number of thermal units. Denoting the maximum generation capacity and the minimal combination production cost of output \( Q_{g_i} \) at bus \( g_i \) at hour \( k \) by \( K_{g_i}[k] \) and \( c_{g_i}(Q_{g_i}[k]) \), the cost of generation at \( k = 0 \) for each generation substation can be expressed as the following:

$$c_{g_{s1}}(Q_{g_{s1}}[0]) = 0.0869Q_{g_{s1}}^2[0] \quad (8.13)$$

$$c_{g_{s2}}(Q_{g_{s2}}[0]) = 0.1811Q_{g_{s2}}^2[0] \quad (8.14)$$

where \( 0 \leq Q_{g_{s1}}[0] \leq K_{g_{s1}}[0] = 79.43\,\text{MW} \) and \( 0 \leq Q_{g_{s2}}[0] \leq K_{g_{s2}}[0] = 600\,\text{MW} \) respectively. Of the 600MW initial \( (k = 0) \) capacity at bus 2, 300MW belong to the nuclear units and the rest belong to the thermal units.

A reinforcement may be made to each generation substation increasing the maximum generation ratings and reducing the minimal combination production cost of output at the respective bus for a given investment cost. For the 2-bus network, only a set of gas-turbine units of 300.00MW capacity can be built at bus 1 at an investment cost of $270.00 for the first set and $10,000 per set for any additional sets. The cost of generation using a set of gas-turbine units, once built, is given by

$$c(Q) = 0.0869(Q + 79.43)^2 \quad (8.15)$$

for \( 0 \leq Q \leq 300 \).

For simplicity it is assumed that no physical depreciation and outage occurs for any of the installed generating units. Then, any investment in generation simply adds to the existing generation ratings at the respective bus and remains at that level until the next new investment. As mentioned above, the cost of the generation is reduced at the bus where the investment is made. For example, if a set of gas-turbine units is built at bus 1, the new cost of generation at some hour \( k \) can be computed using

$$c_{g_{s1}}(Q_{g_{s1}}[k]) = 0.0869Q_{g_{s1}}^2[k] \quad (8.16)$$

where \( 0 \leq Q_{g_{s1}}[k] \leq 379.43\,\text{MW} \). For the generation output between 79.43MW and 379.43MW the cost of generation is reduced from infinite before the investment to \( 0.0869Q_{g_{s1}}^2[k] \) after the investment.

The transmission lines connecting buses 1 and 2 are identical and are each given 30MW ratings at \( k = 0 \).
Similar to the generation capacity it is assumed that no physical depreciation occurs in any of the installed transmission lines. However, the status of the lines is subject to random outages.

Suppose there is a 10% probability of the transmission lines being down at any given moment. For the 2-bus network example, the operation of the transmission system is considered such that, with good maintenance, the period of time for which the line is out is longer than the time required to melt the line but shorter than the time required to sustain a simultaneous outage of both lines. This is equivalent to stating that the probability of having both lines unavailable for service and the probability of having only one line unavailable for service at any given time are 1% and 18%, respectively, if the lines are poorly maintained, while the same probabilities are 0% and 20% if the lines are well-maintained. Further, it is restricted that the maintenance decision of the network is binary for the entire year, i.e., either poor \((e_m[k] = 0)\) or good \((e_m[k] = 1)\) where \(k = 4(n-1) + 1, 4(n-1) + 2, 4(n-1) + 3\) and \(n\) is an index for a particular year. The cost associated with the maintenance decisions is given by

\[
\upsilon_m(e_m[k] = 0) = 0.00($) \tag{8.17}
\]

\[
\upsilon_m(e_m[k] = 1) = 10.00($) \tag{8.18}
\]

Thus, the decision to maintain the network well in year 3, for example, results in the expenditure of $40.00.

A reinforcement may be made to each transmission line, increasing the maximum ratings at a given investment cost. For the 2-bus network example, only a set of 10.00MW capacity up to 40.00MW can be built on lines 1 and 2 simultaneously at the investment costs of $1,147, $1,588, $2,323, $3,352, and $4,675 for one, two, three and four sets of investing in additional 10.00MW capacities.

An additional 5MW may be added to the ratings of the transmission lines through the use of flexible technologies (mostly software related). It is restricted that the operation decisions to use the flexible technologies is binary for the entire year, i.e., either negative \((e_{tech}[k] = 0)\) or positive \((e_{tech}[k] = 1)\) where \(k = 4(n-1) + 1, 4(n-1) + 2, 4(n-1) + 3\) and \(n\) is an index for a particular year. The cost associated with the operation decisions is given by

\[
\upsilon_{tech}(e_{tech}[k] = 0) = 0.00($) \tag{8.19}
\]

\[
\upsilon_{tech}(e_{tech}[k] = 1) = 45.00($) \tag{8.20}
\]

Thus, the decision to use flexible technology in year 2, for example, results in the expenditure of $180.00.

According to A6, the transmission line flow limits need to be defined to include the effect of equipment outages, specifically the transmission line outages for the 2-bus system example. Here a qualitative argument is given for defining the flow limits.

It is evident from the utility functions of the load centers for the 2-bus system example that any loss
of load is quite costly. For example, a 1MW loss from the 18.5MW demand in bus 1 at \( k = 1 \) results in a $10,000,000.00 loss in utility for the respective load center. Thus, it is clear that no loss of load is highly desirable. Suppose a maximum of 5MW generation capacity can be made available for immediate production at each generation substation in the case of line outages. Two cases are considered: (1) \( e_m[k] = 0 \) and (2) \( e_m[k] = 1 \) where \( k = 0 \).

If \( e_m[k] = 0 \), then the most limiting case is that there is a 1\% probability that no transfer between buses 1 and 2 is possible at times. As described before, any loss of load is highly undesirable. Therefore, the flow limits between the buses need to be defined such that even when no transfer between buses 1 and 2 is possible, no load is lost, i.e. \( F_i^{\text{aux}} = 2.50 \text{MW} \). Based on the same line of reasoning, it can be deduced that if 10MW generation capacity is available for immediate production, then the limit can be defined at \( F_i^{\text{aux}} = 5.00 \text{MW} \) instead of the more restrictive 2.50MW limits.

If \( e_m[k] = 1 \), then the most limiting case is that there is a 10\% probability that one of the lines is not available for transfer between buses 1 and 2. Thus, the flow limits between the buses need to be defined such that when one of the lines is not available, using additional generation capacity (reserve) can sustain the loss of a single line without causing a loss of load. Such a limit is 17.50MW. Suppose each line is carrying up to 17.50MW each at some given moment for a total of a 35.00MW transfer from the exporting bus to the importing bus. As soon as one of the lines is lost due to a random outage, the importing bus can increase the generation at the respective bus by 5MW and subsequently the exporting bus can decrease its generation by 5MW resulting in a 30.00MW transfer between the buses. The entire 30.00MW transfer between the buses is now supported by the remaining transmission line without violating its maximum ratings.

The generation capacity available for immediate production is the operating reserves and, just as described earlier, these play an important role in defining the maximum flow limits on transmission lines in this example.

It is important to remember that the generation cost also plays an important role in defining the maximum flow limits since higher flow limits usually result in lower generation costs and in higher utility when the supply and demand balance. For this particular example, however, the generation costs do not need to be considered since the utilities of the load centers at buses 1 and 2 are several orders of magnitude greater than the costs; this may be an implied assumption of the electric power network since the demand in many cases is considered to be perfectly inelastic. Thus, only the loss of load is important is this particular example.

For the rest of the 2-bus network example, it is assumed that no generation capacity is available for immediate production in the case of outages. In addition, the maximum maintenance effort is assumed to be made at each hour. Using this assumption, the maximum flow limits on the transmission lines can simply be defined as half of the respective ratings without having to consider how the reserve is provided at different operating conditions for this particular example. Generalizing the definition of the maximum flow limit on transmission lines is becoming a very important topic and is to be explored in future research.

For the sake of argument, consider the following scenario for the 2-bus system example. At the beginning of year 2, all of the nuclear units at bus 2 are scheduled to be taken out of the network for maintenance.
purposes. With probabilities of 10% and 100%, these units are available for generation again at the beginning of year 3 and at the beginning of year 4, respectively. Unlike the uncertainties imposed on the utility functions of loads, it is noted that the uncertainties of nuclear plants actually affect the evolution of the system states, in this case the generation capacity available at bus 2. Thus, the DP formulation is justifiably well suited.

Now define the systemwide social welfare at hour $k$ as

$$ SW_{\text{system}}[k] = \sum_{d_j} U_{d_j} - \sum_{g_i} c_{g_i} - \sum_{g_i} C^G_{g_i} - \sum_{i} C^T_i - v_{\text{tech}} $$

(8.21)

For the 2-bus network example, the reinforcement into generation and transmission through investment is assumed to be made only at the beginning of each year in order to reduce the computational complexity.

Thus, the problem of optimizing the systemwide social welfare becomes

$$ \max_{q_{g_i}, q_{d_j}, \epsilon, \bar{e}_{\text{tech}}, \bar{e}_{\text{tech}}} \mathcal{E}_{\text{load}} \{ SW_{\text{system}}[k] \} $$

(8.22)

where

$$ K^G_{g_i}[k + 1] = K^G_{g_i}[k] + I^G_{g_i}[k] $$

(8.23)

$$ K^T_i[k + 1] = K^T_i[k] + I_i[k] $$

(8.24)

subject to

$$ \sum_{g_i} Q_{g_i}[k] = \sum_{d_j} Q_{d_j}[k] $$

(8.25)

$$ 0 \leq Q_{g_i}[k] \leq K^G_{g_i}[k] $$

(8.26)

$$ F_i[k] \leq F_i^{\text{max}}[k] = 0.5(K^T_i[k] + 5e_{\text{tech}}[k]) $$

(8.27)

This problem of optimizing the social welfare function in (8.22) can be described using the decisions tree shown in Figure 8-2 with respect to the consumption, the production, the investments in generation and transmission, and the operation effort.

Figure 8-3 shows the sum of the expected generation cost and the expected transmission cost incurred systemwide in years 2, 3 and 4 based on the various decisions made at the beginning of each year. The costs in year 1 are not shown because they are not relevant to the decision making process. It can be seen from Figure 8-3 that the costs in year 4 do not play any role in the decision making process either since the relevant costs are equal to one another.

Working backwards from year 3 then, the optimal decision can be computed by pruning all the branches except for the minimum cost branch. This process is shown in Figure 8-4. As seen from Figure 8-4 the optimal operating and planning decisions for maximizing the expected value of the social welfare function
Figure 8-2: Decision tree for the social welfare optimization problem in the 2-bus network example
Figure 8-3: The sum of the expected generation cost and the expected transmission cost for years 2, 3, and 4.
Figure 8-4: Computation of the optimal decision for the systemwide social welfare maximization problem
result in the investing of 40MW of transmission capacity into the network line and 300MW of gas-turbine units at bus 1 at the beginning of year 2. Plus, an operation effort to use flexible technology is required at the beginning of year 2 and, if the nuclear plants remain out of service, at the beginning of year 3 as well. This results in a minimum combined cost of $17,518 for generation and transmission that meet the load.

In Chapter 2 the formulation of the operation and planning problems of the network seen by the benign social planner is reviewed as a systemwide social welfare maximization problem by generalizing the type of study performed for the 2-bus network example to any electric power network. The dynamic programming technique provides a simple way of representing the decision tree pruning and is, thus, used extensively in the formulation. Following the review of the optimization problem, some discussions of operation and planning seen in the vertically integrated utility are given, and compared to the solution of the welfare maximization problem. Chapter 2 closes with the implications of functional unbundling through the deregulation process.

Suppose the operation and planning problems of the 2-bus network example is re-visited, as seen by a transmission provider (TP) and a regulator following the functional unbundling with an introduction of market mechanisms. First, the cost functions for generators and the utility functions for loads need to be replaced with respective supply and demand functions.

(A8) It is assumed throughout the thesis that there is no market power on the part of either suppliers or loads. As a consequence of this assumption, the energy portion of the electric services is provided through perfect competition, and the supply and demand functions have the simple interpretation of being the first derivative of production cost functions and utility functions, respectively.

Starting from the utility function, the demand function for the loads at bus 1 is given as the following:

For years 1 and 2, i.e., $k = 1, 2, \cdots 8$

$$D_{d_1}(Q_{d_1}[k], k) = -2 \times 10^7 \cdot (Q_{d_1}[k] - \beta_{d_1}[k])$$ \hspace{1cm} (8.28)

where

$$\text{Prob}(\beta_{d_1}[k] = 18.50MW) = 100\%$$ \hspace{1cm} (8.29)

For years 3 and 4, i.e., $k = 9, 10, \cdots 16$

$$D_{d_1}(Q_{d_1}[k], k) = -2 \times 10^7 \cdot (Q_{d_1}[k] - \beta_{d_1}[k])$$ \hspace{1cm} (8.30)

where

$$\text{Prob}(\beta_{d_1}[k] = 19.43MW) = 100\%$$ \hspace{1cm} (8.31)

Similarly, the demand function for bus 2 is given as the following:

For years 1 and 2, i.e., $k = 1, 2, \cdots 8$

$$D_{d_2}(Q_{d_1}[k], k) = -2 \times 10^7 \cdot (Q_{d_1}[k] - \beta_{d_1}[k])$$ \hspace{1cm} (8.32)
where
\[
Prob(\beta_d_2[k]) = 156.46\text{MW} = 80\% \tag{8.33}
\]
\[
Prob(\beta_d_2[k]) = 161.15\text{MW} = 10\% \tag{8.34}
\]
\[
Prob(\beta_d_2[k]) = 151.77\text{MW} = 10\% \tag{8.35}
\]

For years 1 and 2, i.e., \( k = 9, 10, \cdots 16 \)
\[
D_{d_2}(Q_{d_1}[k], k) = -2 \times 10^7 \cdot (Q_{d_1}[k] - \beta_d_1[k]) \tag{8.36}
\]

where
\[
Prob(\beta_d_2[k]) = 164.28\text{MW} = 80\% \tag{8.37}
\]
\[
Prob(\beta_d_2[k]) = 169.21\text{MW} = 10\% \tag{8.38}
\]
\[
Prob(\beta_d_2[k]) = 159.35\text{MW} = 10\% \tag{8.39}
\]

The aggregate supply function for the suppliers at bus 1 and for the suppliers at bus 2 are given respectively, by
\[
S_{g_1}(Q_{g_2}[k]) = 0.1738Q_{g_1}[k] \tag{8.40}
\]
\[
S_{g_2}(Q_{g_2}[k]) = 0.3622Q_{g_2}[k] \tag{8.41}
\]
where \( 0 \leq Q_{g_1}[k] \leq 79.43 \) or \( 0 \leq Q_{g_1}[k] \leq 379.43 \) depending on the investment decision about generation, and \( 0 \leq Q_{g_2}[0] \leq K_{g_2}[0] = 600 \).

Due to A8, the process of balancing the supply and demand at the equilibrium leads to a maximum consumer and supplier surplus at each hour in years 1, 2, 3, and 4. Mathematically, this can be represented as
\[
\max_{Q_{d_1}, Q_{g_1}} \sum_{d_j} D_{d_j}[k] - \sum_{g_i} Q_{g_i}[k] \tag{8.42}
\]
subject to
\[
\sum_{g_i} Q_{g_i}[k] = \sum_{d_j} Q_{d_j}[k] \tag{8.43}
\]
\[
0 \leq Q_{g_i}[k] \leq K_{g_i}^C[k] \tag{8.44}
\]
\[
F_i[k] \leq F_i^{max}[k] = 0.5(K_i^T[k] + 5e_{tech}[k]) \tag{8.45}
\]

It should be noted that the maximum ratings for generation and for transmission lines vary depending on the investment decisions made at the beginning of each year up to and including the current year. Thus, two different equilibria for the same year but under different system conditions may be significantly different.
from each other due to differing investment decisions. The effect of investment decisions leading to different equilibria can be seen by examining the expected price at which the energy is traded between the suppliers and the loads.

Figure 8-5 shows the expected prices at which the electricity is traded between the suppliers and the loads. It can be seen in Figure 8-5 that the same investment decisions about generation lead to very different market equilibria depending on the investment decisions about transmission, and vice versa.

Once the prices are determined, the expected revenue of the energy suppliers as well as of the TP can be readily computed. Of particular interest is the expected revenue (and subsequently the expected profit) derived from the new investments in generation and in transmission.

Figure 8-6 shows the expected revenue from the new gas-turbine generators at bus 1 based on the expected price shown in Figure 8-5. The dotted line represents the decision by the suppliers not to invest in gas-turbine generators, and the dashed lines represent the operation and planning decisions by the TP. It can be seen in the Figure 8-6 that the profit of the gas-turbine generators depends very much on the transmission investment as well as the availability of nuclear plants in year 3. Given that the cost of installing the gas-turbine generation units is given by $270.00, the investment yields a positive return only if the nuclear plants remain unavailable in year 3, and if the TP expands its transmission capacity by 40MW, or higher, through investment in year 2 and an additional 5MW or higher through flexible technology in year 3.

Figure 8-7 shows the expected revenue collected from the marginal pricing of increased transmission capacity reduced by the transmission cost. This time the dashed lines represent the investment decisions by the energy suppliers. Again, the revenue depends very much on the investment decisions of the suppliers as well as on the availability of nuclear plants in year 3. Another interesting feature in Figure 8-7 is that the expected revenue collected from the marginal pricing of increased transmission capacity reduced by the transmission cost is highest when no investment is made in year 2, and when an additional 5MW to the existing transmission capacity is achieved through flexible technology in year 3, under the conditions that the nuclear plants remain unavailable and that the investment in new gas-turbine units is made in year 2. This is because of the high fixed cost for investing in new transmission lines, approximately $1,000 for this particular example.

(A9) It is assumed throughout the thesis that there exists a high degree of economies of scale and economies of scope in the transmission network.

Therefore, when comparing the possible solution to the investment problem of the TP, based on the incentives shown in Figure 8-7, and the result of the optimal decisions, shown in Figure 8-4, it is clear that there needs to be an additional revenue requirement to the marginal pricing of transmission capacity for the TP in order for these two solutions to approach each other due to A9.

In Chapter 3 the formulation of the operation and planning problems of the network seen by the TP and the regulator is reviewed as an optimization problem by generalizing the type of study discussed for the 2-bus network example above to any electric power network. Taking into account the assumption in A9,
Figure 8-5: Expected prices at bus 1 and at bus 2 for different investment decisions
Figure 8-6: Expected revenue of a new gas-turbine generator at bus 1
Figure 8-7: Expected revenue from the marginal pricing of new capacity subtracted by the transmission cost

216
the TP is allowed to exist as a natural monopoly and is placed under strict regulation by the regulators. Two types of regulation schemes, cost-of-service (COS) regulation and price-cap regulation (PCR) in highly stylized versions, are discussed in the context of inducing optimal investment decisions from the perspective of the TP to approach the social welfare optimizing solutions described in Chapter 2.

It is shown that a considerable amount of knowledge by the regulator is required in order to induce such incentives from the TP in either regulation scheme although there is a distinct advantage in implementing the PCR over COS regulation because the final operation and planning decisions are placed with the TP only under the PCR whereas COS regulation places the final decisions with the regulators. For the 2-bus network example, this means that under COS regulation, at best the regulator decides on 60MW of investment in new transmission capacity and sets the appropriate rate in order to guarantee that the TP recovers its respective investment costs. No expense is spent for effort into flexible technology since no incentive can be created for expenditure by the TP other than capital costs. In contrast, under the PCR the regulator sets an appropriate rate beforehand and allows the TP to make the necessary decisions. When the rate is set correctly, this leads very close to the systemwide social welfare optimizing solutions. The setting of the rates under the PCR is to be explored in future research. For the thesis, it is assumed that the ceiling prices under the PCR are given.

Chapter 3 closes with a discussion of another advantage of PCR over the TP: the ability to actively participate in the market process.

In Chapter 4 a brief survey is given describing the current development in the restructuring process in some parts of the US, namely the California, Pennsylvania-New Jersey-Maryland (PJM) and Midwest regions. Following the survey some critical observations are discussed on the unique characteristics of the electric power system. These observations emphasize the need for well defined longer term transmission contracts in order to foster longer term energy trading.

The operation and planning problems of the 2-bus network example are again re-visited but this time as seen by the energy suppliers and the loads. It can be seen from Figure 8-5 that in year 3 the expected price at which the energy is traded between the suppliers and loads may vary widely depending on the investment decisions about generation and transmission. For example, the loads at bus 2 can pay a price for electricity, in year 3 as much as 25% higher than the systemwide optimal level, if the proper signal for new investments in generation and in transmission is not given. In the deregulated environment an active trading of long term contracts for energy as well as for transmission is essential for producing a strong signal for the right set of investments. Suppose the existing suppliers and potential suppliers offer their expected generation at a perfectly competitive level for years 2 and 3. Further suppose the loads are keen on purchasing enough electricity to cover their future demands at a stable price. Then, Figure 8-8 shows the expected prices at which the electricity is traded in years 2 and 3 through long term contracts. The computation in Figure 8-8 is performed by maximizing the surplus of the suppliers and loads while simultaneous maximizing the surplus of the TP, as shown in Figures 8-5 and 8-6, respectively.
Figure 8-8: Expected prices at which the electricity is traded in years 2 and 3 through long term contracts
Chapter 5 describes the designing principles for the for-profit Independent Transmission Company (ITC), an entity which is necessary in order to allow active participation by the TP in the market process. Specifically, it is pointed out that the active participation by the TP plays an important role in fostering the trade of electric services (both energy and transmission) over the longer term. The active trading of electricity over the longer term replaces the coordinated planning process of the vertically integrated utility since the market participants are motivated to make the necessary investments based on well founded informations on the supply and demand needs of the future. Unlike other commodities, this ability to replace the coordinated planning process with the market mechanism is critical for the electric power industry because it directly affects the adequacy of energy and transmission given the lack of a practical means of storing electricity. Thus, the concepts developed in Chapter 5 lay the foundation for underlying architectures for (long term) reliability. An explanation is given for the process through which the TP determines the prices at which the long term transmission contracts are initially offered. However, depending on the change in demand for transmission capacity by the suppliers and the loads, the pricing of long term transmission contracts needs to evolve. This requires a feedback design for the pricing, which is to be explored in future research. The chapter closes with a discussion of the critical relationship between the risk-managing by ITC and the proposed PCR scheme.

Suppose the suppliers build the new gas-turbine units without securing long term contracts for energy and transmission in the 2-bus network example described above. Then, the expected revenue of the new suppliers reduced by the respective production costs varies between $343.75 with a 90% chance and $126.50 with a 10% chance depending on the availability of nuclear plants in year 3. This implies that even if the expected revenue exceeds the combined cost of production and investment, there is a 10% chance the supplier loses $143.50 from the new investment. This changes dramatically if the suppliers enter into long term contracts for energy and transmission. In this case, the expected revenue of the new suppliers reduced by the respective production costs varies between $274.42 with a 90% chance and $749.05 with a 10% chance depending on the availability of nuclear plants in year 3. This implies that not only does the expected revenue exceed the combined cost of production and investment, but also the chance to incur a loss from the new investment is completely eliminated. A similar case can be made for the TP under the PCR with the ability to offer longer term transmission contracts as proposed in this thesis.

Chapter 6 describes the significance of longer term transmission contracts, along with bilateral energy contracts over a longer period, in encouraging new investments in generation and transmission based on proper signals in a market environment not limited to (financially) risk-neutral participants but rather including risk-averse participants, as has long been well understood by economists. There the liquidity of the contracts plays an important role in recognizing changes in system conditions and managing subsequent variations in the amount of (financial) risks to which the individual market participants are exposed. In Chapter 6 the new functions of the secondary markets for transmission and of the information infrastructure called the open access same-time information system (OASIS) are introduced with considerable detail about
the role of the TP as proposed in this thesis. Some qualitative description of the proposed longer term transmission rights are given, but a more formal treatment of these rights, especially under market imperfection and/or under irregularities such as generator ramping costs and "green" power compared to conventional financial rights and physical rights, is left for future research. The chapter closes with an extended treatment of the cluster-based congestion management system in the operation of a well designed OASIS.

A significant portion of the thesis deals with the operation and planning of an electric power network in the single control area market case where the TP is assumed to be able to define the maximum flow limit on the transmission lines in a meaningful way, considering many aspects affecting the short term reliability of the system. This is closely related to the TP’s ability to internalize various network uncertainties affecting short term reliability. As the market boundaries expand beyond a single control area case, however, this assumption is not valid since the factors in one region influencing the short term reliability of the entire interconnection are no longer well understood by the TP’s in the other regions.

Chapter 7 describes a new paradigm where the factors affecting the short term reliability of the entire interconnection are explicitly accounted by the network users and the TP’s of individual regions. In the paradigm a new entity, called the inter-regional transmission organization (IRTO), is proposed for assigning the benefit of economically driven transactions involving multiple regions to the change in the value of reliability over the entire interconnection, so that the market process, through an auction mechanism, takes the effort for maintaining reliability using a control mechanism. Only illustrative discussions are provided in Chapter 7 while a detailed description of the paradigm is referred to in the impending patent application.

As attempted in this thesis, we believe that any proposed designs for electricity market structures should be examined with a clear understanding of the implications on the overall industry performance, as well as with an understanding of the implications on the individual industry participants, such as the power suppliers, the provider of wires, and the consumers. Particular emphasis should be placed on understanding the long-term (in contrast to only short-term) effects of various changes on the adequacy of the supply and on the evolution of the grid necessary to support the long-term needs of the energy markets.

The problem the electric power industry has at hand is much more troublesome and complex, in terms of theoretical and practical challenges, than the very limited scope of this thesis. It is recognized that the attempts made here are only scratching the surface of the overall problem. It will take some deep thinking and patience to get the entire electric power industry to function properly following the restructuring process.
Appendix A

Restructuring process in the electric power industry

Often competition is the best way to protect the public interest and ensure the lowest price possible for the reliable supply of a commodity. In 1996 the Federal Energy Regulatory Commission (FERC) issued Order 888 to clear the way for instituting competition in the electric power industry in the hopes that it would bring more efficient and lower cost electricity to consumers. In introducing wholesale electricity markets, the Order set a rule that would eliminate any undue discrimination in access to the monopoly owned transmission systems.

The Order required all public utilities that operate a transmission grid:

- to file an open access transmission tariff that is pro forma and non-discriminatory
- to undergo functional unbundling and offer affiliates and independent transmission system users the same separate rates for wholesale generation, transmission, and ancillary services
- to develop and maintain a same-time information system called Open Access Same-Time Information Systems (OASIS) that is available to any transmission customer.

The above, along with stranded cost recovery rules, established the foundation necessary to provide a fair transition to a competitive wholesale electricity market.

Since the issuance of the Order, the landscape of the electricity industry has undergone a complete transformation. As the non-discriminatory tariffs and information about the transmission system became available, the energy trade in wholesale markets grew to provide more of the nation’s acquisition of generation resources. Recently there has been a large increase in the number of power marketers and independent generation facility developers entering the market. These new changes are followed by the development of retail competition, and by the further divestiture of generation plants by vertically integrated utilities.
Utilities are also evolving; there are a significant number of mergers among electric utilities and among electric utilities and gas companies.

The sweeping restructuring activity has led to a more intensive and different usage of the transmission system than anticipated in the past. This has placed new stresses on regional transmission systems. The traditional management of the transmission grid is, however, inadequate for alleviating these stresses and meeting the evolving needs of competitive electricity markets. This calls for a regional solution.

In 1999, FERC issued Order 2000 to address the lack of proper transmission system support. It is recognized that independent regionally operated transmission grids can satisfy the demands of the electricity market and further enhance the competition. According to FERC,\textsuperscript{1}

Appropriate regional transmission institutions could (1) improve efficiencies in transmission grid management (2) improve grid reliability (3) remove remaining opportunities for discriminatory transmission practice (4) improve market performance and (5) facilitate lighter handed regulation.

Those appropriate regional transmission institutions are the Regional Transmission Organizations (RTOs).

Through RTOs, flexible transmission ratemaking, including performance-based regulation, may be provided for pricing many transmission services. RTOs is an evolving institution that meets the market needs in terms of structure, operation, market support and geographic scope. The minimum characteristics and functions of RTOs are defined as the following:

- **Minimum Characteristics:**
  1. independence
  2. scope and regional configuration
  3. operational authority
  4. short-term reliability

- **Minimum Functions:**
  1. tariff administration and design
  2. congestion management
  3. parallel path flow resolution
  4. ancillary services provision
  5. total transmission capability and available transmission capability computation through OASIS
  6. market monitoring
  7. transmission system planning and expansion

\textsuperscript{1}See Regional Transmission Organizations, Order No. 2000, December 20, 1999 Federal Energy Regulatory Commission.
8. interregional coordination

All transmission-owning entities, including non-public utility entities, are expected soon to be placing their transmission facilities under the control of appropriate RTOs.

With the introduction of competition to the electricity industry, the integrity and reliability of the electricity infrastructure can be maintained only by the zealous effort of all participants. Ultimately, competition will lead to a better use of the existing assets and institutions, new market mechanisms, technical innovation, and less rate distortion.

A.1 Reliability and the electric power industry deregulation

Due to the lack of an adequate means of storing electricity, an instantaneous matching of supply and demand is required invariably for providing reliable services since any imbalance may result in a partial or complete breakdown of the system. This translates to, in the short-term, operating a secure transport grid of electricity from generation sources to load centers, and in the long-term, planning an adequate amount of generation and transmission to meet the projected load.

Before the restructuring of the electric power industry, a vertically integrated utility, which is responsible for providing reliable electricity services, has forecasted loads for the next 5 to 10 years and planned the amount and the mix of generation accordingly. This is often referred as the “obligation to serve”. In exchange for utility companies undertaking such a tall task, they are allowed to operate as natural monopolies. The price offered by these utilities is based on a complete recovery of their investments with strict regulation of their operating costs. Thus, at any given time the functioning generators are dispatched starting from the least expensive. Under this industry practice the nature of the operation and planning of the underlying transmission grid is passive, simply connecting the generation sources to load centers.

With the introduction of competition, however, there is no single entity responsible for the operation and planning of the entire electric power system. Rather than being driven by the obligation to serve, participants in the industry respond to factors affecting their respective profits. This seems to be a good mechanism for reinforcing the supply of electricity as independent generation facility owners and developers choose their generation level or make investment decisions in response to their perception of the electricity price. In New England where a shortage of electricity is anticipated, i.e. higher electricity prices, approximately 30,000MW of generation has been proposed or is actually under construction by independent power producers since the industry restructuring process began.

It appears, however, that this reinforcing market mechanism has an adverse effect on maintaining overall reliability due to inadequate support from the transmission system. As generation and transmission become separate on account of the increased effort for divestiture by vertically integrated utilities, the close coordination of generation and transmission in operation and planning is no longer practical. This has caused the derailing of transmission and transmission related facilities planning from generation planning. Despite the
increase in the amount of new generation, the amount of new transmission capacity planned over the next ten years is significantly lower compared to the additions that had been planned five years ago. Plus, most of the planned transmission is supporting local distribution systems.  

The existing transmission grid is not designed to support the new usage of the system based on evolving market need. This is apparent in the number of times that transmission loading relief procedures (TLRs) are invoked. The TLRs are designed to relieve the overloading of transmission lines or other transmission equipment beyond their operation ratings since such overloading may result in damaging the respective lines or equipment, which can in turn become a catalyst for partial or full scale system outages. In 1998, over 300 TLRs have already been implemented. In 1999, more than 400 TLRs were called in the first ten months alone. These numbers are expected to grow in upcoming years unless there is a significant change in the transmission system. This is staggering considering the over 8,000MW of power curtailment in the summer of 1999 due to the implementation of TLR procedures, which rely mainly on curtailing transactions contributing to the overload.

Thus, it is evident that a passive architecture is no longer sustainable without degrading the reliability of the entire system. In the near future the problem is expected to become worse; widely different operation and planning of generation are expected since there are several emerging electricity markets whose fundamental principles are quite distinct from one another.

The existing institutions and practices have served reliably in the past but have to be modified in order to maintain the integrity of the overall transmission system. It is important that this process be thought out starting from the fundamentals. Short-term reliability is the responsibility of the system operators. These operators control transmission grids that are regional in nature. A careful consideration should be given in defining the size of the geographic area that each operator controls.

A few attempts have been made to improve this short-term reliability through the establishing of a new institution called, an Independent System Operator (ISO). Currently there are five ISOs functioning. They are the California ISO, the Pennsylvania-New Jersey-Maryland (PJM) ISO, ISO New England, the New York ISO and the Midwest ISO. Of these five, only the Midwest ISO, which is composed of non-contiguous areas, does not operate a single control area. They differ in the form of energy markets that they support as well as in the ways transmission service is provided and priced.

According to the North America Electricity Reliability Council (NERC) Task Force

a large ISO would (1) be able to identify and address reliability issues most effectively (2) internalize much of the loop flow caused by the growing number of transactions (3) facilitate transmission access across a larger portion of the network consequently improving market effi-

---

3 There are also proposals for another type of institution called TransCo.
4 Three ISOs (New England, PJM, and New York) operate centralized markets, and the California ISO (California) supports a separate market called the Power Exchange. The Midwest ISO does not have an ISO-related centralized market for its region.
ciencies and promoting greater competition and eliminate “pancaking” of transmission rates thus allowing a greater range of economic energy trades across the network.⁵

In spite of the effort to induce necessary modifications to the existing institution and practices, there still is a significant lack of effort in terms of maintaining long-term reliability.

Appendix B

DC load flow problem

This appendix presents the DC load flow problem, which provides an approximate description of the steady state operating conditions of a given electric power network.

Typically, the steady state operation condition of a network is described using four system variables at each bus $i$, namely the voltage magnitude, $V_i$, the phase angle, $\delta_i$, the real power injection, $Q_i^r$, and the imaginary power injection, $Q_i^i$. Two of the system variables are specified by the network users while the other two are determined as a result of the interactions among the specified variables within the network. For modeling a generator at bus $g_i$, typically the voltage magnitude, $V_{g_i}$, and the real power injection $Q_{g_i}$ are assumed to be known ahead of time. For modeling a load at bus $d_j$, the real power injection, $Q_j^r$, and the imaginary power injection, $Q_j^i$ are assumed to be given. The system operator designates one bus (usually a generator bus) in the network, $g_i = 0$ as the slack bus where the voltage magnitude, $V_0$, and phase angle, $\delta_0$, are fixed. This designation is a mathematical convenience because the real power inputs to all buses cannot be known ahead of time since the network losses are unknown ahead of time. The slack bus is assumed to compensate for any mismatches in generation and demand in the modeling of the network operation.\(^1\) To summarize,

- **slack generator**: a power source of infinite power output capability. The controlled variables are the voltage magnitude, $V_0 = 1$ (p.u.), and the phase angle $\delta_0 = 0$ (°).

- **loads**: power sinks. The controlled variables are the real and the reactive power injection, $Q_{d_j}^r$ and $Q_{d_j}^i$, respectively, where $d_j = 1, \ldots, N_D$.

- **generators**: power sources. The controlled variables are the real power injection, $Q_{g_i}^r$, and voltage magnitude, $V_{g_i}$, where $g_i = 1, \ldots, N_G$.

\(^1\)In the real system, any mismatch in generation and demand is compensated for by using automated generation control (AGC)

226
For modeling a transmission line $ij$, the $\pi$-equivalent model shown in Figure B-1 is often used [45]. Using the $\pi$-equivalent representation, two system matrices are computed for characterizing the topological properties of the transmission network, namely the reduced incidence matrix $A$ and the admittance matrix $Y$. $A$ is useful for describing purely topological properties of the system; how buses are connected by transmission lines in the system. Suppose $N_L$ is the number of transmission lines in the system. Then $A$ is the matrix whose dimensions are $N_L \times (N_D + N_G)$ and whose entries, $a_{ij}$, are given by:

$$a_{ij} = \begin{cases} 
1 & \text{if there is a transmission line connecting buses } i \text{ and } j, \text{ and } i < j \\
-1 & \text{if there is a transmission line connecting buses } i \text{ and } j, \text{ and } i > j \\
0 & \text{otherwise}
\end{cases} \quad (B.1)$$

$Y$ combines the topological properties of the system with the characteristic properties of transmission lines. The dimensions of $Y$ are given by $(N_D + N_G) \times (N_D + N_G)$, and the entries of $Y$, $y_{ij}$, are given by:

$$y_{ii} = \sum_{j \neq i} Y_{ij}$$
$$y_{ij} = -Y_{ij} \quad (B.2)$$

where $Y_{ij}$ is the admittance of transmission line $ij$ as shown in Figure B-1. $Y$ provides a convenient notation of relating various system variables. For example, the network bus voltages, $V$, and currents, $I$, are related
simply by:

\[
I = \begin{bmatrix}
I_1 & I_2 & \cdots & I_{N_D + N_G} \\
\vdots & \vdots & & \vdots \\
I_{(N_D + N_G)1} & I_{(N_D + N_G)2} & \cdots & I_{(N_D + N_G)(N_D + N_G)}
\end{bmatrix}
= \begin{bmatrix}
y_{11} & y_{12} & \cdots & y_{1(N_D + N_G)} \\
y_{21} & y_{22} & \cdots & y_{2(N_D + N_G)} \\
\vdots & \vdots & & \vdots \\
y_{(N_D + N_G)1} & y_{(N_D + N_G)2} & \cdots & y_{(N_D + N_G)(N_D + N_G)}
\end{bmatrix}
\begin{bmatrix}
V_1 \\
V_2 \\
\vdots \\
V_{(N_D + N_G)}
\end{bmatrix}
= YV.
\]

(B.3)

\[\text{B.1 Load Flow Problem}\]

From the definition of complex power at bus \( i \), \( \hat{S}_i \),

\[
\hat{S}_i = \hat{V}_i \hat{I}_i^* \\
= \hat{V}_i \sum_{j=0}^{(N_D + N_G)} \hat{Y}_{ij}^* \hat{V}_j^* \\
i = 0, 1, \cdots, (N_D + N_G)
\]

(B.4)

where \((\cdot)^*\) denotes the complex conjugate operator. Let

\[
\hat{V}_i = V_i e^{j\delta_i} \\
\delta_{ij} = \delta_i - \delta_j \\
\hat{Y}_{ij} = G_{ij} + jB_{ij}
\]

where \( G_{ij} \) and \( B_{ij} \) are the conductance and the susceptance of transmission line \( ij \) respectively. Then Eq. (B.4) becomes

\[
\hat{S}_i = \sum_{j=0}^{(N_D + N_G)} V_i V_j e^{j\delta_{ij}} (G_{ij} - jB_{ij}) \\
= \sum_{j=0}^{\text{gen}} V_i V_j (\cos \delta_{ij} + j \sin \delta_{ij}) (G_{ij} - jB_{ij}) \\
i = 0, 1, \cdots, (N_D + N_G)
\]

(B.5)

Separating the real and imaginary part of \( \hat{S}_i \) into \( Q_i^r \) and \( Q_i^l \)

\[
Q_i^r = \sum_{j=0}^{N_D + N_G} V_i V_j (G_{ij} \cos \delta_{ij} + B_{ij} \sin \delta_{ij}) \\
Q_i^l = \sum_{j=0}^{N_D + N_G} V_i V_j (G_{ij} \sin \delta_{ij} - B_{ij} \cos \delta_{ij})
\]

(B.6)
Thus, the load flow equations are of the form:

\[
Q^R = Q^R(V, \delta) \\
Q^I = Q^I(V, \delta). \tag{B.7}
\]

Given the controllable system variables, the objective (often referred to as the full blown load flow problem) is to solve Eq. (B.7) for the uncontrollable system variables. This is a nonlinear problem and the solution cannot be found analytically in most cases. The Gauss iterative method and Newton-Raphson method are often used to find numerical solutions to the problem.

### B.1.1 Simplifications to The Load Flow Problem

Some simplifications to Eq. (B.6) can be made to reduce the number of iteration steps that must be taken. One such simplification comes from considering only the real power part of the load flow problem. It turns out that the real power at any bus \( i \) is not affected much by voltage magnitude deviations at the buses during the iteration steps [45]. For this reason, the voltage magnitude at bus \( i \) can be assumed to be fixed during the iteration steps when solving only the real power part of the load flow problem. The reduced problem is given by:

\[
Q_i^R = \sum_{j=0}^{N_p+N_q} V_i V_j (G_{ij} \cos \delta_{ij} + B_{ij} \sin \delta_{ij}) \tag{B.8}
\]

where both \( V_i \) and \( V_j \) are now assumed to be given. This simplified load flow problem is called the decoupled \((P - \delta)\) load flow. Since we are only considering the real power injection, we drop the superscript \( R \) for convenience. Further simplification can be made to the decoupled \((P - \delta)\) load flow by assuming:

- the voltage magnitude at bus \( i \) is fixed at 1 (p.u.); \( V_i = 1 \) p.u., or some other value corresponding to nominal operation
- typically the resistance \( R_{ij} \), of transmission line \( ij \) is much smaller than the reactance \( X_{ij} \) of the same line; \( X_{ij} \gg R_{ij} \)
- the network is made up of short lines; \(|\delta_i - \delta_j| \ll 1\) so that \( \sin \delta_{ij} \) can approximated by \( \delta_i - \delta_j \)

The linearized form of Eq. (B.8) under the above mentioned assumptions is then

\[
Q_i = \sum_{j=1}^{N_p+N_q} \frac{1}{X_{ij}} \delta_{ij} \tag{B.9}
\]

This most simplified form of the load flow problem is called the DC load flow.

Using the DC load flow approximation the real power flow through transmission line \( l \), \( F_l \) is given by

\[
F_l = \sum_{g_i} H_{lg_i} Q_{g_i} - \sum_{d_j} H_{ld_j} Q_{d_j} \tag{B.10}
\]
Here we define the power transfer distribution factor (PTDF) of line $l$ with respect to bus $i$, $H_{ij}$, as the element of the matrix, $H$.

$$H = \text{diag}\left(\frac{1}{X_{ij}}\right) A Y^{-1}$$  \hspace{1cm} (B.11)
Bibliography


[53] Lecinq, B., Peak-load pricing for transmission in a deregulated electric utility industry, Master’s Thesis, Massachusetts Institute of Technology, 1996.


